Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 158-168, March 2020.

A NOTE ON THE ORTHOGONAL PROCRUSTES PROBLEM AND NORM-DEPENDENT OPTIMALITY*

JOSHUA CAPE†

Abstract. This note revisits the classical orthogonal Procrustes problem and investigates the norm-dependent geometric behavior underlying Procrustes alignment for subspaces. It presents generic, deterministic bounds quantifying the performance of a specified Procrustes-based choice of subspace alignment. Numerical examples illustrate the theoretical observations and offer additional, empirical findings which are discussed in detail. This note complements recent advances in statistics involving Procrustean matrix perturbation decompositions and eigenvector estimation.

Key words. Procrustes problems, CS decomposition, Singular subspaces, Canonical angles, Two-to-infinity norm.

AMS subject classifications. 15A21, 15A60, 15B10.

1. Introduction.

Let $U, V \in \mathbb{O}_{m,r}$ where $\mathbb{O}_{m,r}$ denotes the set of orthonormal r-frames in $\mathbb{R}^m$ (i.e., Stiefel matrices) and $\mathbb{O}_{m,m} \equiv \mathbb{O}_m$ denotes the set of $m \times m$ real orthogonal matrices. In this paper we consider the orthogonal Procrustes problem(s)

$$\inf_{W \in \mathbb{O}_r} \|U - VW\|, \tag{1.1}$$

where the norm $\|\cdot\|$ is taken to be any of the following: the Frobenius norm $\|\cdot\|_F$, the spectral norm $\|\cdot\|_2$, or the two-to-infinity norm $\|\cdot\|_{2\to\infty}$, the latter of which is defined for any matrix $A \in \mathbb{R}^{m\times n}$ as

$$\|A\|_{2\to\infty} := \sup_{\|x\|_2 = 1} \|Ax\|_\infty. \tag{1.2}$$

Our motivation for considering the two-to-infinity norm stems from recent matrix analysis applications in statistics. There, the two-to-infinity norm is leveraged together with perturbation analysis to yield novel estimation results for eigenvectors and principal subspaces.

Considerable effort has been devoted to Procrustes problems over the years (see [7] for an overview), beginning with the classical orthogonal Procrustes problem which is well-understood under the Frobenius norm [11]. In particular, when writing the singular value decomposition of $V^\top U \in \mathbb{R}^{r\times r}$ as $V^\top U \equiv W_1 \Sigma W_2$ with $W_1, W_2 \in \mathbb{O}_r$, then $W^\star := W_1 W_2 \in \mathbb{O}_r$ solves equation (1.1) under $\|\cdot\|_F$. Less is generally known about the orthogonal Procrustes problem under different norms.

Taken together, and given the analytic tractability of $W^\star$, we are led to investigate

$$\|U - VW^\star\| \tag{1.3}$$

in both spectral and two-to-infinity norm. We also study the behavior of $\|U - VW\|$ when $W \neq W^\star$.

*Received by the editors on May 29, 2019. Accepted for publication on February 18, 2020. Handling Editor: James G. Nagy.

†Department of Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA (joshua.cape@pitt.edu). This work was partially supported by the D3M program of the Defense Advanced Research Projects Agency administered through Air Force Research Laboratory contract FA8750-17-2-0112, and by the Acheson J. Duncan Fund for the Advancement of Research in Statistics at Johns Hopkins University. The author also gratefully acknowledges support from National Science Foundation grant DMS-1902755.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 158-168, March 2020.

2. Overview and motivation. This paper is intended to serve as a baseline, perturbation-free counterpart to [6], providing both complementary theory and numerics. Here and in [6], we advocate for more widespread consideration of the two-to-infinity subordinate vector norm on matrices and demonstrate how $\|\cdot\|_{2\to\infty}$ provides an easily interpretable, operationally useful, and geometrically significant quantity.

This paper does not assume an *a priori* relationship between the matrices $U, V \in \mathbb{O}_{m,r}$ in equation (1.3). As such, the results discussed here hold in broad generality. By way of contrast, stronger results can be obtained under additional structural assumptions, particularly in the classical matrix perturbation setting $\widetilde{A} := A+E$. In statistics, for example, the papers [3, 5, 6] consider such settings where U (derived from some matrix $\widetilde{A}$) is viewed as a (stochastic) perturbation of V (derived from some matrix A), making it possible to considerably improve (probabilistically) upon certain bounds (e.g., Theorem 4.4).

With an eye towards applications and numerics, we remark that an orthonormal r-frame $U \in \mathbb{O}_{m,r}$ may be viewed as a concatenated matrix of r orthonormal (column) eigenvectors (or singular vectors) corresponding to some underlying $m \times m$ matrix. As such, $\|U\|_{2\to\infty}$ provides a measure of joint eigenvector homogeneity and basis coherence. Indeed, this two-to-infinity norm quantity is relevant for various problems in fields including statistics, computer science, and applied mathematics, arising either implicitly or explicitly in the following settings:

- Matrix completion and recovery in optimization [4, 8];
- Consistency of clustering in random graph inference [9];
- Delocalization of eigenvectors in random matrix theory [10];
- Singular subspace perturbation in high-dimensional statistics [3, 6].

The remainder of this paper is organized as follows. Section 3 establishes notation and discusses elementary properties of the two-to-infinity norm that can be derived in a straightforward fashion from first principles. Section 4 provides a collection of bounds relating specifications of $\|U - VW\|$ (e.g., in two-to-infinity norm; for $W^\star$) to canonical angles, matrix dimensions, entrywise matrix behavior, and two-to-infinity norm quantities. Section 5 provides numerical investigations of the ideas considered in this work and is the focus of this paper. In addition to illustrating theoretical bounds provided earlier, Section 5 further explores the complexities of Procrustes problems when varying the choice of norm and choice of orthogonal matrix W. We observe diverse, subtle behavior occurring already in low-dimensional Euclidean space. Collectively, these contributions do not seem to appear elsewhere in the literature. We conclude with additional discussion in Section 6 and collect proof details in the appendix.

3. Preliminaries. We begin by establishing notation, introducing the two-to-infinity norm with its basic properties, and reviewing the concept of canonical angles.

3.1. Notation. All vectors and matrices in this paper are taken to be real-valued. The symbols $:=$ and $\equiv$ are used to assign definitions and to denote formal equivalence, respectively. For any positive integer m, let $[m] := \{1, 2, \dots, m\}$. For (column) vectors $x, y \in \mathbb{R}^m$ where $x \equiv (x_1, \dots, x_m)^\top$, the standard Euclidean inner product between x and y is denoted by $\langle x, y\rangle$. The vector of all ones is denoted by $\mathbf{e}$. The classical ℓ_p vector norms are denoted by $\|x\|_p := \left(\sum_{i=1}^m |x_i|^p\right)^{1/p}$ for $1 \le p < \infty$ and $\|x\|_\infty := \max_i |x_i|$. In addition to the Frobenius norm, $\|\cdot\|_F$, the spectral norm, $\|\cdot\|_2$, and two-to-infinity norm, $\|\cdot\|_{2\to\infty}$, let $\|A\|_\infty := \max_i \sum_j |a_{ij}|$ denote the maximum absolute row sum of A, i.e., the matrix norm induced by the ℓ_∞ vector norm. We also consider the matrix norm given by $\|A\|_{\max} := \max_{i,j} |a_{ij}|$.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 158-168, March 2020.

3.2. The two-to-infinity norm. We catalog several preliminary facts about the two-to-infinity norm in the form of several propositions. The proofs are straightforward in nature, and we refer to [6] for additional details. Below, in summary:

- Proposition 3.1 says that $\|A\|_{2\to\infty}$ corresponds to the maximum Euclidean row norm of A.
- Proposition 3.2 makes explicit the relationship between $\|\cdot\|_{2\to\infty}$ and $\|\cdot\|_2$ in terms of underlying matrix dimensions.
- Proposition 3.3 records the sub-multiplicative behavior of $\|\cdot\|_{2\to\infty}$ as a subordinate operator norm (see [12] for more general discussion of such norms).
- Proposition 3.4 recalls the partial isometry invariance of $\|\cdot\|_2$ and notes the "restricted" partial isometry invariance of $\|\cdot\|_{2\to\infty}$.

Proposition 3.1. *Let $A \in \mathbb{R}^{m\times n}$ and $A_i \in \mathbb{R}^n$ denote the i-th row of A. Then*

$$\|A\|_{2\to\infty} = \max_{i\in[m]} \|A_i\|_2. \tag{3.4}$$

Proposition 3.2. *For $A \in \mathbb{R}^{m\times n}$,*

$$\|A\|_{2\to\infty} \leq \|A\|_2 \leq \min\{\sqrt{m}\|A\|_{2\to\infty}, \sqrt{n}\|A^\top\|_{2\to\infty}\}. \tag{3.5}$$

Proposition 3.3. *For $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$, and $C \in \mathbb{R}^{l\times m}$,*

$$\|AB\|_{2\to\infty} \leq \|A\|_{2\to\infty}\|B\|_2 \text{ and } \|CA\|_{2\to\infty} \leq \|C\|_\infty\|A\|_{2\to\infty}. \tag{3.6}$$

Proposition 3.4. *For $A \in \mathbb{R}^{r\times s}$, $U \in \mathbb{O}_{m,r}$, and $V \in \mathbb{O}_{n,s}$,*

$$\|A\|_2 = \|UA\|_2 = \|AV^\top\|_2 = \|UAV^\top\|_2, \text{ and } \|A\|_{2\to\infty} = \|AV^\top\|_{2\to\infty}. \tag{3.7}$$

However, $\|UA\|_{2\to\infty}$ need not equal $\|A\|_{2\to\infty}$.

We emphasize that for $U \in \mathbb{O}_{m,r}$, $\|U\|_{2\to\infty} = \|UU^\top\|_{2\to\infty}$ by Proposition 3.4.

This subsection concludes with a simple proposition describing the possible two-to-infinity norm values of orthonormal r-frames. The key observation underlying Proposition 3.5 is that the standard basis vectors and $\mathbf{e}/\sqrt{m} \in \mathbb{R}^m$ represent the "extremal" unit vectors with respect to the two-to-infinity norm.

Proposition 3.5. *For $1 \leq r < m$ let $U \in \mathbb{O}_{m,r}$ and $U_\perp \in \mathbb{O}_{m,m-r}$ be such that $[U|U_\perp] \in \mathbb{O}_m$. Then $\sqrt{\frac{r}{m}} \leq \|U\|_{2\to\infty} \leq 1$, $\sqrt{\frac{m-r}{m}} \leq \|U_\perp\|_{2\to\infty} \leq 1$, and*

$$\|[U|U_\perp]\|_{2\to\infty} = 1 < \tfrac{\sqrt{r}+\sqrt{m-r}}{\sqrt{m}} \leq \|U\|_{2\to\infty} + \|U_\perp\|_{2\to\infty} \leq 2. \tag{3.8}$$

3.3. Canonical angles and $\sin\Theta$ distance. The columns of matrices $U, V \in \mathbb{O}_{m,r}$ form orthonormal bases for r-dimensional subspaces in $\mathbb{R}^m$, respectively. Letting $\{\sigma_i(V^\top U)\}_{i=1}^r$ denote the singular values of $V^\top U$ indexed in non-increasing order, it follows from the classical CS matrix decomposition [1] that the *canonical angles* between the subspaces corresponding to U and V are given by the main diagonal elements of the $r \times r$ diagonal matrix

$$\Theta(U,V) := \mathrm{diag}(\cos^{-1}(\sigma_1(V^\top U)), \cos^{-1}(\sigma_2(V^\top U)), \ldots, \cos^{-1}(\sigma_r(V^\top U))). \tag{3.9}$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 158-168, March 2020.

In turn, it is well-known (e.g., see [1]) that for the matrix of canonical angles,

$$\|\sin\Theta(U,V)\|_2 = \|V_\perp V_\perp^\top UU^\top\|_2 \equiv \|V_\perp^\top U\|_2, \tag{3.10}$$

where the trigonometric sine operation is applied to the main diagonal elements of the matrix $\Theta(U,V)$, and $U_\perp \in \mathbb{O}_{m,m-r}$ is such that $[U|U_\perp] \in \mathbb{O}_m$. In certain instances that follow, it will be expedient to write $\mathfrak{s}_{U,V} := \|\sin\Theta(U,V)\|_2$.

4. Theory. Lemma 4.1 (below) represents a first step in the direction of understanding the relationship between equation (1.3) and the geometry of $\sin\Theta$ distance. A proof can be found in [6].

LEMMA 4.1. *For $U,V \in \mathbb{O}_{m,r}$ and $T \in \mathbb{R}^{r\times r}$, then $\|\sin\Theta(U,V)\|_2 \leq \|U - VT\|_2$. Furthermore, for any $W \in \mathbb{O}_r$,*

$$\tfrac{1}{\sqrt{m}}\|\sin\Theta(U,V)\|_2 \leq \|U - VW\|_{2\to\infty}. \tag{4.11}$$

Lemma 4.2 (below) is fundamentally a statement about the geometry of two orthogonal projections, more precisely about the distance between subspaces. It is not new (e.g., see [2]), though it plays an important role in this paper. A concise proof using our notation is provided in the appendix for convenience.

LEMMA 4.2. *For any $U,V \in \mathbb{O}_{m,r}$ and $W^\star \in \mathbb{O}_r$ as in Section 1,*

$$\|U - VW^\star\|_2 = \sqrt{2 - 2\sqrt{1 - \|\sin\Theta(U,V)\|_2^2}}; \tag{4.12a}$$

$$\|V^\top U - W^\star\|_2 = 1 - \sqrt{1 - \|\sin\Theta(U,V)\|_2^2}. \tag{4.12b}$$

For notational convenience we define the functions $\alpha(\cdot)$ and $\beta(\cdot)$ on $[0,1]$ to be

$$\alpha(a) := \sqrt{2 - 2\sqrt{1-a^2}}, \quad \beta(b) := 1 - \sqrt{1-b^2},$$

noting that $\alpha(\cdot) = \sqrt{2\beta(\cdot)}$.

COROLLARY 4.3. *Let $U,V \in \mathbb{O}_{m,r}$. Proposition 3.2 and equation (4.12a) together yield*

$$\tfrac{1}{\sqrt{m}}\alpha(\mathfrak{s}_{U,V}) \leq \|U - VW^\star\|_{2\to\infty} \leq \alpha(\mathfrak{s}_{U,V}). \tag{4.13}$$

In addition, Proposition 3.3, Proposition 3.4, and equation (3.10) together yield

$$\|\sin\Theta(U,V)\|_2 \geq \left(\frac{\|V_\perp V_\perp^\top UU^\top\|_{2\to\infty}}{\|V_\perp V_\perp^\top\|_{2\to\infty}}\right). \tag{4.14}$$

Equation (4.13) provides simple, general bounds for the two-to-infinity norm formulation of equation (1.3) in terms of $\sin\Theta$ distance by passing between norms. The lower bound here is also a bound for equation (1.1) under $\|\cdot\|_{2\to\infty}$ as can be seen in the proof.

Equation (4.14) relates $\sin\Theta$ distance between subspaces to the two-to-infinity norm of associated orthogonal projections. The lower bound improves upon the trivial bound $\|\sin\Theta(U,V)\|_2 \geq \|V_\perp V_\perp^\top UU^\top\|_{2\to\infty}$ by equation (3.10), since $\|V_\perp V_\perp^\top\|_{2\to\infty} = \|V_\perp\|_{2\to\infty} \leq 1$.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 158-168, March 2020.

Theorem 4.4 (below) relates the quantity $\|U - VW^\star\|_{2\to\infty}$ to both the $\sin\Theta$ distance and to the row structure of V and $V_\perp$. We remark that in various (statistical) settings involving (stochastic) matrix perturbations, the term implicitly bounded by $\mathfrak{s}_{U,V}\|V_\perp\|_{2\to\infty}$ can be analyzed more delicately to yield an improved leading-order term, as pursued in [5, 6].

Theorem 4.4. *Let $U, V \in \mathbb{O}_{m,r}$, and $W^\star \in \mathbb{O}_r$ be as in Section 1. Then for $V_\perp \in \mathbb{O}_{m,m-r}$ such that $[V|V_\perp] \in \mathbb{O}_m$,*

$$\|U - VW^\star\|_{2\to\infty} \leq \mathfrak{s}_{U,V}\|V_\perp\|_{2\to\infty} + \beta(\mathfrak{s}_{U,V})\|V\|_{2\to\infty}. \tag{4.15}$$

Furthermore, when $m = 2$, $r = 1$, and $V \equiv \frac{1}{\sqrt{2}}\mathbf{e}$, the above inequality becomes equality.

Establishing a lower bound for $\|U - VW^\star\|_{2\to\infty}$ in terms of $\mathfrak{s}_{U,V}$ and two-to-infinity norm quantities is a more delicate task. The earlier lower bound in Corollary 4.3 is suboptimal in the presence of additional structure (in V) given the corollary's level of generality. This point is addressed further in Section 5.

We conclude this section with Theorem 4.5 which describes the deviation between the matrices $U - VW$ and $UU^\top - VV^\top$ in a row-wise and entrywise sense.

Theorem 4.5. *For $U, V \in \mathbb{O}_{m,r}$ and $W \in \mathbb{O}_r$,*

$$\|UU^\top - VV^\top\|_{\max} \leq (\|U\|_{2\to\infty} + \|V\|_{2\to\infty})\,\|U - VW\|_{2\to\infty}. \tag{4.16}$$

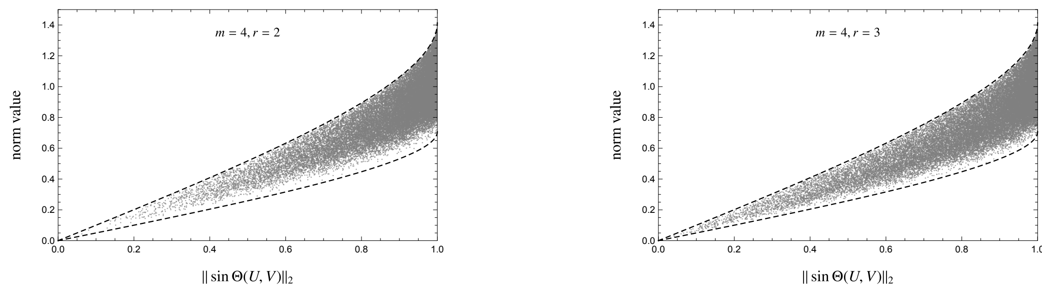
In addition, $\|U\|_{2\to\infty} \leq \|V\|_{2\to\infty} + \|U - VW\|_{2\to\infty}$.

5. Numerics. This section presents numerical simulation results which complement the theoretical observations in Section 4. Specifically, we provide numerical verifications for several aforementioned bounds. Furthermore, we investigate the quality (optimality) of candidate orthogonal matrices W in terms of Procrustes alignment for the less well understood two-to-infinity norm.

5.1. Sampling scheme for $U, V \in \mathbb{O}_{m,r}$. Sampling matrices from $\mathbb{O}_{m,r}$ can be done in a straightforward manner by performing Gram–Schmidt (GS) orthonormalization on collections of independent and identically distributed (i.i.d.) random vectors generated from the m-dimensional multivariate standard normal distribution, $\mathcal{N}(0, I_m)$. More specifically, one can generate U (sim. V) by setting $U := \mathsf{GS}(Z) \in \mathbb{O}_{m,r}$ where $Z := [z_1|z_2|\cdots|z_r] \in \mathbb{R}^{m\times r}$ and the random vector z_i is independently distributed according to $\mathcal{N}(0, I_m)$ for each $i \in [r]$. We explicitly consider small values of m and r in order to obtain good coverage in our simulations and to avoid probabilistic concentration in high dimensions.

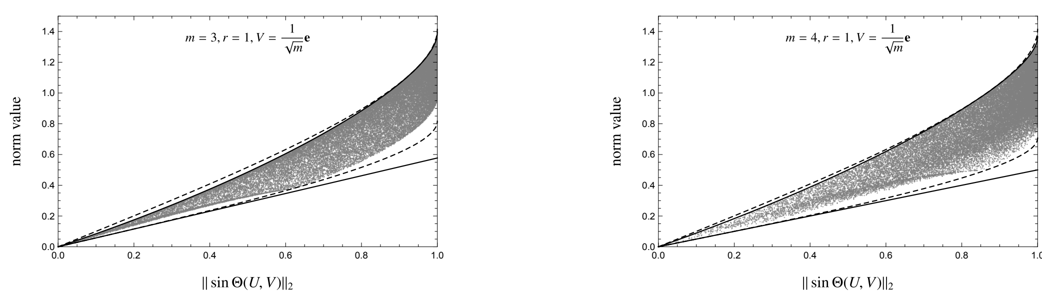
5.2. Simulated $\|U - VW^\star\|_{2\to\infty}$ for i.i.d. U and V. We first consider an unstructured setting in which pairs of $U, V \in \mathbb{O}_{m,r}$ are taken to be i.i.d. as described in the previous paragraph. Figure 1 plots $\|U - VW^\star\|_{2\to\infty}$ as a function of $\|\sin\Theta(U, V)\|_2$ for two regimes of m and r. In each regime, we see that, given the underlying absence of additional structure, the depicted tight bounds are simply those obtained from passing between the two-to-infinity and spectral norms (Corollary 4.3).

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 158-168, March 2020.



(a) (b)

Figure 1: Scatter plots of $\|U - VW^{\star}\|_{2\to\infty}$ as a function of $\|\sin\Theta(U,V)\|_2$ for 40,000 replicates of i.i.d. matrix pairs U, V where $W^{\star} \equiv W^{\star}(U,V)$ as in Section 1. The dashed lines represent theoretical upper and lower bounds given by Corollary 4.3.



(a) (b)

Figure 2: Scatter plots of $\|U - VW^{\star}\|_{2\to\infty}$ as a function of $\|\sin\Theta(U,V)\|_2$ for 40,000 replicates of i.i.d. U and fixed V. The dashed lines represent theoretical upper and lower bounds given by Corollary 4.3. The solid lines represent theoretical upper and lower bounds given by Theorem 4.4 and Lemma 4.1, respectively. Observe the appearance of piecewise continuous lower bound behavior in the scatter plots, with cusps near the $\|\sin\Theta(U,V)\|_2$ values of 0.65 and 0.85, respectively. Here, tight lower bounds with exact endpoints can in principle be computed explicitly via the case-by-case approach employed in the proof of Theorem 4.4.

5.3. Simulated $\|U - VW^{\star}\|_{2\to\infty}$ for random U and fixed V. Next, we again sample $U \in \mathbb{O}_{m,r}$ i.i.d. as described in Section 5.1 but now specify that $V \equiv \frac{1}{\sqrt{m}}\mathbf{e}$.

Figure 2 shows realizations of the quantity $\|U - VW^{\star}\|_{2\to\infty}$ as a function of $\|\sin\Theta(U,V)\|_2$ for the regimes $m \in \{3, 4\}$ and $r = 1$. The choices of V and m here provide low-dimensional examples to contrast the special equality case in Theorem 4.4. In this structured setting, the upper and lower bounds behave

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 158-168, March 2020.

in a more nuanced manner compared to the unstructured setting depicted in Figure 1. Namely, here the spectral norm-implied theoretical upper bound (top dashed line) is improved by the theoretical upper bound (top solid line) which takes into account the structure of V and $V_{\perp}$. The spectral norm-implied theoretical lower bound (bottom dashed line) is no longer tight. Rather, the tight lower bounds in each regime (not depicted) are piecewise continuous functions of $\|\sin\Theta(U,V)\|_2$ and can be derived in a similar fashion as in the case $m=2$, $r=1$ in Theorem 4.4. We also plot $\frac{1}{\sqrt{m}}\|\sin\Theta(U,V)\|_2$ from Theorem 4.4 (bottom solid line) which demonstrates the theoretical gap between two-to-infinity norm optimality and the Frobenius norm optimality of the matrix $W^{\star}$.

5.4. On norm-dependent optimality and optimal orthogonal matrices. In the previous sections, $W^{\star}$ is viewed as a surrogate for the (analytically intractable) two-to-infinity-optimal orthogonal Procrustes solution, namely,

$$\underset{W\in\mathbb{O}_r}{\arg\inf}\ \|U-VW\|_{2\to\infty}.$$

This section addresses the question "*How good is the choice of* $W^{\star}$*?*" by providing an illustrative empirical investigation of this query.

To begin, we recapitulate that given $U,V\in\mathbb{O}_{m,r}$ and the associated Frobenius-optimal orthogonal matrix $W^{\star}$, then for all $W\in\mathbb{O}_r$,

$$\frac{1}{\sqrt{m}}\|\sin\Theta(U,V)\|_2 \le \|U-VW^{\star}\|_{2\to\infty} \le \|U-VW^{\star}\|_F \le \|U-VW\|_F, \tag{5.17}$$

together with $\|U-VW\|_{2\to\infty}\le\|U-VW\|_F$.

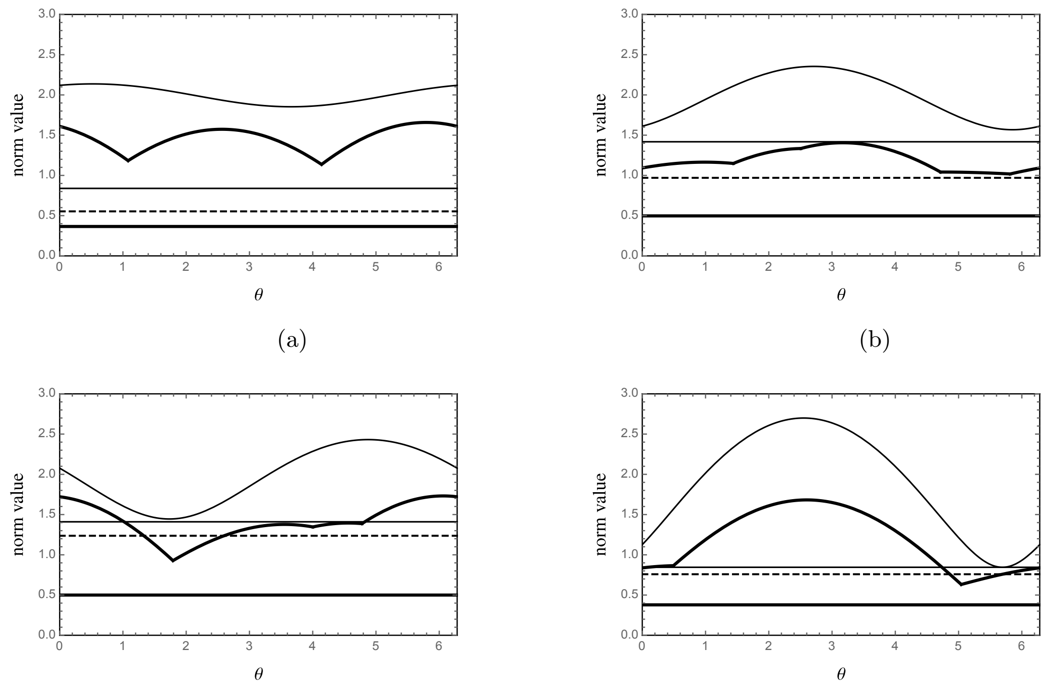
For $\theta\in[0,2\pi)$ let $R_{\theta}^{\mathrm{rot}}$ denote the two-by-two counterclockwise (orthogonal) rotation matrix with angle θ, i.e.,

$$R_{\theta}^{\mathrm{rot}} := \begin{bmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{bmatrix}.$$

For each sub-figure in Figure 3, we independently simulate a single pair of matrices U,V from $\mathbb{O}_{4,2}$ and plot both $\|U-VR_{\theta}^{\mathrm{rot}}\|$ and $\|U-VW^{\star}\|$ in two-to-infinity norm and Frobenius norm, together with a lower bound on the two-to-infinity-optimal Procrustes norm value. We summarize Figure 3 shown below.

- Figure 3a: the rotation $R_{\theta}^{\mathrm{rot}}$ is uniformly (in θ) dominated under both $\|\cdot\|_{2\to\infty}$ and $\|\cdot\|_F$. Moreover, $R_{\theta}^{\mathrm{rot}}$ under $\|\cdot\|_{2\to\infty}$ is always strictly worse than $W^{\star}$ under $\|\cdot\|_F$, namely, $\|U-VW^{\star}\|_F\le\|U-VR_{\theta}^{\mathrm{rot}}\|_{2\to\infty}$.
- Figure 3b: the rotation $R_{\theta}^{\mathrm{rot}}$ uniformly improves upon $\|\cdot\|_F$ in $\|\cdot\|_{2\to\infty}$ but is inferior to $W^{\star}$, namely, $\|U-VW^{\star}\|_{2\to\infty}\le\|U-VR_{\theta}^{\mathrm{rot}}\|_{2\to\infty}\le\|U-VW^{\star}\|_F$.
- Figures 3c and 3d: the behavior of $R_{\theta}^{\mathrm{rot}}$ is more nuanced in that, depending on the value of θ, $\|U-VR_{\theta}^{\mathrm{rot}}\|_{2\to\infty}$ is smaller than or larger than each of $\|U-VW^{\star}\|_{2\to\infty}$ and $\|U-VW^{\star}\|_F$, respectively. In Figure 3c observe the improvement over $W^{\star}$ in that $\|U-VR_{\theta}^{\mathrm{rot}}\|_{2\to\infty}<\|U-VW^{\star}\|_{2\to\infty}$ for θ near $\|U-VR_{\theta}^{\mathrm{rot}}\|_F\approx\|U-VW^{\star}\|_F$. Figure 3d demonstrates a pronounced effect of norm-dependent optimality for orthogonal Procrustes in that, for $\theta\approx 5$, then $\|U-VR_{\theta}^{\mathrm{rot}}\|_{2\to\infty}<\|U-VW^{\star}\|_{2\to\infty}$ while $\|U-VR_{\theta}^{\mathrm{rot}}\|_F>\|U-VW^{\star}\|_F$, whereas $\|U-VR_{\theta}^{\mathrm{rot}}\|_{2\to\infty}=\|U-VW^{\star}\|_{2\to\infty}$ as well as $\|U-VR_{\theta}^{\mathrm{rot}}\|_F=\|U-VW^{\star}\|_F$ for $\theta\approx 5.85$.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 158-168, March 2020.



(a) (b) (c) (d)

Figure 3: For underlying matrices $U, V \in \mathbb{O}_{4,2}$ each plot depicts two curves and three (constant) lines as functions of θ. The upper and bottom solid lines represent $\|U - VW^\star\|_F$ and $\frac{1}{2}\|\sin\Theta(U,V)\|_2$, respectively. The dashed line represents $\|U - VW^\star\|_{2\to\infty}$. The upper and lower solid curves represent $\|U - VR_\theta^{\mathrm{rot}}\|_F$ and $\|U - VR_\theta^{\mathrm{rot}}\|_{2\to\infty}$, respectively.

It is in fact possible to explicitly parametrize all 2×2 real orthogonal matrices and therefore to solve equation (1.1) under the two-to-infinity norm in the present special setting. To this end, for $\theta \in [0, 2\pi)$ let R_θ^{ref} denote the two-by-two (orthogonal) reflection matrix given by

$$R_\theta^{\mathrm{ref}} := \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix}.$$

Table 1 displays the results when numerically solving equation (1.1) for each of the four pairs of matrices U and V underlying Figure 3. In the setting corresponding to Figure 3d, both the Frobenius optimal and two-to-infinity norm optimal orthogonal transformations are in fact rotation matrices. Plots for R_θ^{ref} resemble the range of behavior exhibited in Figure 3 and are therefore not shown.

In each simulation example, $W^\star \in \{R_\theta^{\mathrm{rot}}\}_{\theta\in[0,2\pi)} \cup \{R_\theta^{\mathrm{ref}}\}_{\theta\in[0,2\pi)}$, and $W^\star$ is suboptimal for the two-to-infinity norm orthogonal Procrustes problem. Figure 3d depicts this suboptimality near the true two-to-infinity and Frobenius minimizers, whereas Figure 3c demonstrates similar behavior in a small parameter region for θ near a local minimum in Frobenius norm. In all cases, the location of $W^\star$ is "close" to that of the true minimizing orthogonal transformation (not shown in Figures 3a and 3b since the optima are in R_θ^{ref}). Such "closeness" can be quantified more precisely in matrix analysis applications found in high-dimensional

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 158-168, March 2020.

statistics.[1]

Table 1

	Pair (a)	Pair (b)	Pair (c)	Pair (d)
opt_F	$(0.83, \theta_{\mathrm{ref}} = 2.22)$	$(1.41, \theta_{\mathrm{ref}} = 5.28)$	$(1.41, \theta_{\mathrm{ref}} = 5.00)$	$(0.84, \theta_{\mathrm{rot}} = 5.69)$
$\mathrm{opt}_{2\to\infty}$	$(0.44, \theta_{\mathrm{ref}} = 2.01)$	$(0.94, \theta_{\mathrm{ref}} = 6.26)$	$(0.90, \theta_{\mathrm{ref}} = 4.16)$	$(0.63, \theta_{\mathrm{rot}} = 5.04)$

6. Discussion. This paper investigates the geometry corresponding to $U, V \in \mathbb{O}_{m,r}$ both column-wise (i.e., via $\sin\Theta$ distance) and row-wise (i.e., via $\|U\|_{2\to\infty}$, $\|V\|_{2\to\infty}$). Geometrically, the value $\|V\|_{2\to\infty}$ corresponds to the radius of the minimal Euclidean ball in $\mathbb{R}^r$ centered at the origin which contains all the rows of V. In contrast, the columns of V are all on the unit Euclidean sphere in $\mathbb{R}^m$.

The goal of minimizing $\|U - VW\|$ in terms of W may be viewed as seeking the best-possible alignment of U subject to a ground-truth subspace basis V. In Section 4 we show how additional structural considerations on the ground-truth V influence the two-to-infinity norm behavior of the matrix $U - VW$ and related quantities (see Theorem 4.4 and Theorem 4.5). We emphasize that the two-to-infinity norm, unlike the spectral and Frobenius norms, is an example of a basis-dependent norm (recall Proposition 3.4). In Section 5 we illustrate the interplay between the choice of norm, choice of subspace, and orthogonal transformation when quantifying subspace alignment.

Appendix A. Technical material.

Proof of Lemma 4.2. For any real matrix A, the spectral norm satisfies $\|A\|_2^2 = \|A^\top A\|_2$. This observation facilitates the computation

$$\begin{aligned}
\|U - VW^\star\|_2^2 &= \|(U - VW^\star)^\top (U - VW^\star)\|_2 \\
&= \|U^\top U - U^\top VW^\star - (W^\star)^\top V^\top U + (W^\star)^\top V^\top VW^\star\|_2 \\
&= \|2I - (V^\top U)^\top W^\star - (W^\star)^\top (V^\top U)\|_2 \\
&= \|2I - (W_1\Sigma W_2)^\top W_1 W_2 - (W_1 W_2)^\top (W_1 \Sigma W_2)\|_2 \\
&= \|2I - W_2^\top \Sigma W_2 - W_2^\top \Sigma W_2\|_2 \\
&= 2\|I - \Sigma\|_2 \\
&= 2(1 - \min_i \cos(\theta_i)) \\
&= 2\left(1 - \sqrt{1 - \max_i \sin^2(\theta_i)}\right).
\end{aligned}$$

Taking square roots establishes the first claim. As for the second claim, observe that $\|V^\top U - W^\star\|_2 = \|\Sigma - I\|_2 = 1 - \min_i \cos(\theta_i) = 1 - \sqrt{1 - \max_i \sin^2(\theta_i)}$. ∎

[1] Remark: Within the perturbation framework considered in [5, 6], it is shown that the approximation error incurred by considering $W^\star$ rather than the true-but-intractable two-to-infinity optimal orthogonal transformation is asymptotically negligible in certain settings.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 158-168, March 2020.

Proof of Theorem 4.4. Using the triangle inequality together with Proposition 3.3 and equation (4.12b) yields

$$\begin{aligned}\|U - VW^{\star}\|_{2\to\infty} &= \|U - VV^{\top}U + VV^{\top}U - VW^{\star}\|_{2\to\infty}\\ &\leq \|(V_{\perp}V_{\perp}^{\top})U\|_{2\to\infty} + \|V(V^{\top}U - W^{\star})\|_{2\to\infty}\\ &\leq \|V_{\perp}^{\top}U\|_{2}\|V_{\perp}\|_{2\to\infty} + \|V^{\top}U - W^{\star}\|_{2}\|V\|_{2\to\infty}\\ &= \mathfrak{s}_{U,V}\|V_{\perp}\|_{2\to\infty} + \beta(\mathfrak{s}_{U,V})\|V\|_{2\to\infty}. \qquad \square\end{aligned}$$

Proof of equality special case in Theorem 4.4. Consider the case when $m = 2$ and $r = 1$ with $U \equiv u \in \mathbb{R}^2$ and $V \equiv v := (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^{\top} \in \mathbb{R}^2$. Then $\|U-VW^{\star}\|_{2\to\infty} = \max_i|u_i - \frac{1}{\sqrt{2}}\omega|$ where $\omega = 1$ when $\cos\theta \equiv \langle u, v\rangle > 0$ and $\omega = -1$ when $\langle u, v\rangle < 0$. If $\langle u, v\rangle = 0$, note that $\|U - VW^{\star}\|_{2\to\infty} = \sqrt{2}$.

Here, $\mathfrak{s}_{U,V} \equiv |\sin\theta|$, so squaring both sides of the equality $\cos\theta \equiv \langle u, v\rangle$, expanding the inner product, and using the fact that $\|u\|_2 = 1$ yields $u_1u_2 = \frac{1}{2} - \mathfrak{s}_{U,V}^2$. There are three cases to consider:

- Case 1: If $\mathfrak{s}_{U,V} = \frac{1}{\sqrt{2}}$ then $\max_i|u_i| = 1$ and so $\|U - VW^{\star}\|_{2\to\infty} = \frac{1}{\sqrt{2}}$;
- Case 2: If $0 \leq \mathfrak{s}_{U,V} < \frac{1}{\sqrt{2}}$, then u_1 and u_2 have the same sign;
- Case 3: If $\frac{1}{\sqrt{2}} < \mathfrak{s}_{U,V} \leq 1$, then u_1 and u_2 have different signs.

In Case 2, one has $\frac{1}{2} - \mathfrak{s}_{U,V}^2 = |u_1||u_2| = |u_1|(1 - |u_1|^2)^{1/2}$ and similarly for $|u_2|$ by symmetry. Solving the previous equation yields

$$\begin{aligned}\max_i|u_i| &= \tfrac{1}{\sqrt{2}}\sqrt{1 + 2\mathfrak{s}_{U,V}\sqrt{1-\mathfrak{s}_{U,V}^2}} = \tfrac{1}{\sqrt{2}}\sqrt{\left(\mathfrak{s}_{U,V} + \sqrt{1-\mathfrak{s}_{U,V}^2}\right)^2}\\ &= \tfrac{1}{\sqrt{2}}\left|\mathfrak{s}_{U,V} + \sqrt{1-\mathfrak{s}_{U,V}^2}\right|, \text{ and similarly}\\ \min_i|u_i| &= \tfrac{1}{\sqrt{2}}\left|\mathfrak{s}_{U,V} - \sqrt{1-\mathfrak{s}_{U,V}^2}\right|.\end{aligned}$$

If $u_1, u_2 > 0$ then $\omega = 1$ whereas if $u_1, u_2 < 0$ then $\omega = -1$. So, for $0 \leq \mathfrak{s}_{U,V} < \frac{1}{\sqrt{2}}$,

$$\begin{aligned}\|U - VW^{\star}\|_{2\to\infty} &= \max\left\{\left|\pm\max_i|u_i| \mp \tfrac{1}{\sqrt{2}}\right|, \left|\pm\min_i|u_i| \mp \tfrac{1}{\sqrt{2}}\right|\right\}\\ &= \tfrac{1}{\sqrt{2}}\left(1 + \mathfrak{s}_{U,V} - \sqrt{1-\mathfrak{s}_{U,V}^2}\right).\end{aligned}$$

On the other hand, for Case 3 when u_1 and u_2 differ in sign, then

$$\begin{aligned}\|U - VW^{\star}\|_{2\to\infty} &= \max\left\{\left|\pm\max_i|u_i| \mp \tfrac{1}{\sqrt{2}}\right|, \left|\mp\min_i|u_i| \mp \tfrac{1}{\sqrt{2}}\right|\right\}\\ &= \tfrac{1}{\sqrt{2}}\left(1 + \mathfrak{s}_{U,V} - \sqrt{1-\mathfrak{s}_{U,V}^2}\right).\end{aligned}$$

Hence, for V, m, and r as specified above, over the entire domain $\mathfrak{s}_{U,V} \in [0, 1]$,

$$\begin{aligned}\|U - VW^{\star}\|_{2\to\infty} &= \tfrac{1}{\sqrt{2}}\left(1 + \mathfrak{s}_{U,V} - \sqrt{1-\mathfrak{s}_{U,V}^2}\right)\\ &= \mathfrak{s}_{U,V}\|V_{\perp}\|_{2\to\infty} + \beta(\mathfrak{s}_{U,V})\|V\|_{2\to\infty}. \qquad \square\end{aligned}$$

Proof of Theorem 4.5. For $U, V \in \mathbb{O}_{m,r}$ and $W \in \mathbb{O}_r$, observe that

$$UU^{\top} - VV^{\top} = (U - VW)U^{\top} + V(WU^{\top} - V^{\top}).$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 36, pp. 158-168, March 2020.

The result follows by applying the triangle inequality with respect to $\|\cdot\|_{\max}$ and then jointly invoking Proposition 3.4 together with the observation that $\|AB^\top\|_{\max} \leq \|A\|_{2\to\infty}\|B\|_{2\to\infty}$ for matrices A and B. $\square$

Acknowledgment. The author is grateful to Donniell E. Fishkind and Kamel Lahouel for their detailed feedback. The author also thanks Chi-Kwong Li, Carey E. Priebe, and Minh Tang for productive discussions.

REFERENCES

[1] R. Bhatia. *Matrix Analysis*. Springer-Verlag, New York, 1997.
[2] A. Böttcher and I. Spitkovsky. A gentle guide to the basics of two projections theory. *Linear Algebra and its Applications*, 432:1412–1459, 2010.
[3] T.T. Cai and A. Zhang. Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *Annals of Statistics*, 46:60–89, 2018.
[4] E. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9:717–772, 2009.
[5] J. Cape, M. Tang, and C.E. Priebe. Signal-plus-noise matrix models: eigenvector deviations and fluctuations. *Biometrika*, 106:243–250, 2019.
[6] J. Cape, M. Tang, and C.E. Priebe. The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *Annals of Statistics*, 47:2405–2439, 2019.
[7] J. Gower and G. Dijksterhuis. *Procrustes Problems*. Oxford University Press, New York, 2004.
[8] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory*, 57:1548–1566, 2011.
[9] V. Lyzinski, D. Sussman, M. Tang, A. Athreya, and C.E. Priebe. Perfect clustering for stochastic blockmodel graphs via adjacency spectral embedding. *Electronic Journal of Statistics*, 8:2905–2922, 2014.
[10] M. Rudelson and R. Vershynin. Delocalization of eigenvectors of random matrices with independent entries. *Duke Mathematical Journal*, 164:2507–2538, 2015.
[11] P. Schönemann. A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, 31:1–10, 1966.
[12] G.W. Stewart and J. Sun. *Matrix Perturbation Theory*. Academic Press, Boston, 1990.