SIAM J. MATRIX ANAL. APPL.
Vol. 41, No. 3, pp. 1208–1236
© 2020 Society for Industrial and Applied Mathematics

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

UNIFORM BOUNDS FOR INVARIANT SUBSPACE PERTURBATIONS*

ANIL DAMLE† AND YUEKAI SUN‡

Abstract. For a fixed symmetric matrix A and symmetric perturbation E we develop purely deterministic bounds on how invariant subspaces of A and $A + E$ can differ when measured by a suitable "rowwise" metric rather than via traditional measures of subspace distance. Understanding perturbations of invariant subspaces with respect to such metrics is becoming increasingly important across a wide variety of applications and therefore necessitates new theoretical developments. Under minimal assumptions we develop new bounds on subspace perturbations under the two-to-infinity matrix norm and show in what settings these rowwise differences in the invariant subspaces can be significantly smaller than the analogous two or Frobenius norm differences. We also demonstrate that the constitutive pieces of our bounds are necessary absent additional assumptions and, therefore, our results provide a natural starting point for further analysis of specific problems. Lastly, we briefly discuss extensions of our bounds to scenarios where A and/or E are nonnormal matrices.

Key words. invariant subspaces, perturbation theory, uniform bounds

AMS subject classifications. 15A18, 15A42, 65F99, 68Q25

DOI. 10.1137/19M1262760

1. Introduction.

Given a matrix A, it is natural to try and "understand" how properties of A change after it has been perturbed in some way. Understanding can take many forms, as can the way we choose to perturb A. In this work we are primarily concerned with additive perturbations and understanding how spectral properties of the matrix change—i.e., given a perturbation E, how do certain invariant subspaces of $A + E$ relate to those of A? This is a long standing question and one that has received extensive attention over the years. This includes the well-known Davis–Kahan theorem [10], work by Wedin [29], and more general and extensive perturbation theory; see, e.g., [25] for an overview.

Nevertheless, many problems of growing interest in mathematics, statistics, and computer science require new variants of such theory. Most notably, this manifests as modifications to the metrics we use to assess the similarity of invariant subspaces of A and $A + E$. Concretely, whereas traditional theory is often interested in classical notions of subspace distance measured by spectral or Frobenius norms, we will be interested in rowwise[1] (or closely related) measures of error. In section 2 we will formally outline the specifics of these metrics, contrast them to traditionally considered metrics, and provide additional preliminary material relevant to our work.

The impetus for these new types of bounds is often, though not exclusively, problems arising in statistics and computer science such as matrix completion [5, 17], principal component analysis [4, 20], robust factor analysis [12], spectral clustering

*Received by the editors May 21, 2019; accepted for publication (in revised form) by D. Kressner June 15, 2020; published electronically August 24, 2020.
https://doi.org/10.1137/19M1262760
Funding: The work of the first author was partially supported by the National Science Foundation grant DMS-1830274. The work of the second author was partially supported by the National Science Foundation grant DMS-1830247.
†Computer Science, Cornell University, Ithaca, NY, 14853 (damle@cornell.edu).
‡Statistics, University of Michigan, Ann Arbor, Ann Arbor, MI, 48109 (yuekai@umich.edu).
[1]As elaborated on later, we must somewhat carefully define what it means for two subspaces to be close rowwise.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

[2, 3, 9, 19, 22, 28], and more. In these settings A will often represent some model, and a given instance of the model $\widehat{A}$ can be thought of as a (random) perturbation to this baseline, i.e., $\widehat{A} = A + E$. Many models A have highly structured and meaningful invariant subspaces whose properties form the basis for a wide variety of algorithmic development and analysis of the underlying problem. Therefore, given $\widehat{A}$ we would like to understand if that structure can still be reliably leveraged. For many of these applications traditional measures of distance do not not necessarily provide adequate control over changes to the invariant subspaces.

A simple, concrete illustration of the types of bounds we will develop is encapsulated in the following scenario. Given a rank-k symmetric matrix A, an $n \times k$ matrix V with orthonormal columns representing the subspace associated with the nonzero eigenvalues, and a symmetric perturbation E, when is the dominant invariant subspace of $A + E$ closer to V rowwise than may be expected based on the smallest nonzero eigenvalue of A and the spectral norm of E?

More specifically, if we let $k = 1$ this question reduces to understanding when

$$\min_{s=\pm 1} \|\widehat{v} - sv\|_\infty \ll \min_{s=\pm 1} \|\widehat{v} - sv\|_2,$$

where v is the eigenvector of A associated the eigenvalue $\lambda \neq 0$ and $\widehat{v}$ is the eigenvector of $A+E$ associated with the eigenvalue largest in magnitude (assuming $\|E\|_2$ is small enough for this to be the appropriate pairing). Our main results restricted to this setting answer such a question by providing explicit bounds on $\min_{s=\pm 1} \|\widehat{v} - sv\|_\infty$. In particular we show that

$$\min_{s=\pm 1} \|\widehat{v} - sv\|_\infty \leq 8\|v\|_\infty \frac{\|Ev\|_2^2}{\lambda^2} + 2\zeta \frac{\|Ev\|_\infty}{\lambda} + 4\zeta \frac{\|Ev\|_2}{\lambda^2} \max_i \|E_{i,:}\|_2,$$

where $\zeta = \|I - vv^T\|_\infty$. Notably, this upper bound can be substantially smaller than a traditional Davis–Kahan bound on $\|vv^T - \widehat{v}\widehat{v}^T\|_2$ that, up to constants, behaves like $\|Ev\|_2/\lambda$.

As illustrated by the above bound and the more general results we present later, in situations where E and V have relatively uniform row norms (i.e., they are incoherent [5]) we may expect significantly better bounds than what is captured by traditional subspace perturbation theory. We will formalize these results for symmetric matrices in section 3 where we also provide proofs and investigate the behavior of our bounds. To complement our theoretical developments, section 4 provides several numerical examples illustrating our bounds in appropriate scenarios.

Given the potential usefulness of such bounds and the extent of relevant applications, this area has received significant attention over the past several years [2, 6, 11, 13]. Our main contributions are summarized as follows.

1. We develop new deterministic rowwise perturbation bounds for orthonormal bases of invariant subspaces of symmetric matrices. Prior work often entangles deterministic bounds with assumptions and/or analysis tailored to specific random settings. Nevertheless, our deterministic bounds are easily amenable to further analysis in the probabilistic settings as illustrated in section 3.3.3. Furthermore, in section 5 we provide some basic extensions of our results to situations where A and/or E are nonnormal matrices.
2. We show that our bounds are sharp by constructing adversarial perturbations that saturate the bounds.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

3. Our perturbation bounds apply under more general conditions than preceding results, and we argue that our assumptions are minimal in certain respects by considering specific examples.

While some aspects of our bounds are broadly in alignment with prior work, as noted above others are new, rely on less restrictive assumptions, and are more directly interpretable. We will draw specific comparisons to existing results parallel to the development of our bounds in section 3.

2. Preliminaries.

2.1. Matrices. Let $A \in \mathbf{R}^{n\times n}$ be a symmetric (not necessarily positive definite) matrix. We arrange its eigenvalues in descending order,

$$\lambda_1 \geq \cdots \geq \lambda_n,$$

and denote its eigendecomposition as

$$A = V_1\Lambda_1 V_1^T + V_2\Lambda_2 V_2^T, \tag{2.1}$$

where $\Lambda_1 = \mathsf{diag}(\lambda_1, \ldots, \lambda_r)$ and $\Lambda_2 = \mathsf{diag}(\lambda_{r+1}, \ldots, \lambda_n)$ and $V_1 \in \mathbf{R}^{n\times r}$ and $V_2 \in \mathbf{R}^{n\times(n-r)}$ are matrices with orthonormal columns whose columns are the associated eigenvectors.[2] To make this splitting well defined we assume that $\lambda_r > \lambda_{r+1}$. In addition, for our later results it is important that Λ_2 explicitly include any zero eigenvalues of A as they must be incorporated into our measure of how close Λ_1 and Λ_2 are. Because we are interested in algebraic orderings of the eigenvalues we use the term *dominant* to refer to the algebraically largest eigenvalues (in contrast to the largest in magnitude). Lastly, for a matrix A we let $\Lambda(A)$ denote the set of eigenvalues of A.

Remark 1. Note that the restriction to the r algebraically largest eigenvalues is not essential. Our results will only depend on spectral separation (or closely related quantities) and therefore may be easily adapted to any isolated collection of r eigenvalues. However, such an adaptation introduces notational overhead without adding anything fundamentally new. Similarly, with appropriate adaptation these results are applicable to magnitude based ordering of the eigenvalues—though the ordering itself may be more sensitive to perturbations than the associated subspaces. Therefore, to streamline the exposition we present everything for the r algebraically largest eigenvalues.

Now, let $\widehat{A} = A + E$ represent a perturbation of A by a symmetric matrix E, and let $\widehat{V}_1 \in \mathbf{R}^{n\times r}$ be a matrix with orthonormal columns whose range is the r-dimensional invariant subspace of $\widehat{A}$ associated with the algebraically largest eigenvalues. For the moment we will assume $\hat{\lambda}_r > \hat{\lambda}_{r+1}$ so this notation is well defined; later assumptions we make will ensure this property. The primary contributions of this paper are centered around relating V_1 and $\widehat{V}_1$.

Throughout this work we will be interested in projections of matrices onto the invariant subspaces associated with A (represented by V_1 and V_2). Therefore, for any matrix $B \in \mathbf{R}^{n\times n}$ we define $B_{i,j}$ as $V_i^T B V_j$ with $i, j \in \{1, 2\}$. Lastly, for any matrix $B \in \mathbf{R}^{n\times n}$ we define the Sylvester operator $S_B : \mathbf{R}^{n\times r} \to \mathbf{R}^{n\times r}$ as

$$S_B : Z \to ZB_{1,1} - B_{2,2}Z.$$

Note that we have embedded V_1 and V_2 directly into the definition of this Sylvester operator for convenience.

[2] In the case of repeated eigenvalues any orthonormal basis for the associated eigenspace suffices.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

2.2. Norms. Throughout this paper, we let $\|\cdot\|_1$, $\|\cdot\|_\infty$, and $\|\cdot\|_2$ denote the standard ℓ_p vector norms and their associated induced matrix norms. Similarly, we let $\|\cdot\|_F$ denote the Frobenius norm. In this work we will also be concerned with the two-to-infinity induced matrix norm. Specifically, we denote this norm by $\|\cdot\|_{2,\infty}$ and note that for an $n \times k$ matrix B it can be defined as the maximum ℓ_2 norm of rows of B, i.e.,

$$\|B\|_{2,\infty} = \max_{i=1,\ldots,n} \|B_{i,:}\|_2 .$$

We outline a few easily verified properties of $\|\cdot\|_{2,\infty}$ that will be useful later:

UNITARY INVARIANCE FROM THE RIGHT: For any orthogonal $Z \in \mathbf{R}^{k\times k}$

$$\|BZ\|_{2,\infty} = \|B\|_{2,\infty}.$$

In other words, the norm is invariant under orthogonal transforms from the right (though, notably, not from the left). This property follows immediately from the unitary invariance of the two-norm.

INVARIANCE TO SIGNED PERMUTATIONS FROM THE LEFT: For any signed permutation $\Pi \in \mathbf{R}^{n\times n}$

$$\|\Pi B\|_{2,\infty} = \|B\|_{2,\infty}.$$

This result follows from the invariance of $\|\cdot\|_\infty$ to signed permutations.

SUBMULTIPLICATIVE RELATIONS: The relevant submultiplicative relationships are

$$\|B_1 B_2\|_{2,\infty} \le \|B_1\|_{2,\infty} \|B_2\|_2 \quad \text{and} \quad \|B_1 B_2\|_{2,\infty} \le \|B_1\|_\infty \|B_2\|_{2,\infty}.$$

These inequalities follow from the definition of induced matrix norms and consistency of $\|\cdot\|_2$ and $\|\cdot\|_\infty$.

2.3. Subspace distances. Given our fairly loose assumptions on the eigenvalues of A, we cannot always talk about convergence to individual eigenvectors. Instead, we consider the distances between invariant subspaces. We will often (implicitly) associate matrices with orthonormal columns and subspaces and refer to the range of a matrix W as $\mathsf{ran}W$. Given two $n \times k$ matrices with orthonormal columns W and $\widetilde{W}$ the distance between the subspaces $\mathsf{ran}W$ and $\mathsf{ran}\widetilde{W}$ is

$$\mathsf{dist}(W, \widetilde{W}) \equiv \|WW^T - \widetilde{W}\widetilde{W}^T\|_2.$$

Equivalent definitions include (see, e.g., [14, section 2.6]) the following:

COMPLEMENTARY SUBSPACES: Let $W_2 \in \mathbf{R}^{n\times n-k}$ be an orthonormal basis for the orthogonal complement of the subspace spanned by W; then

$$\mathsf{dist}(W, \widetilde{W}) \equiv \|W_2^T \widetilde{W}\|_2.$$

SINE-Θ DISTANCE: Let $\Theta(W, \widetilde{W})$ be a diagonal matrix containing the principle angles between W and $\widetilde{W}$; then

$$\mathsf{dist}(W, \widetilde{W}) \equiv \|\sin\Theta(W, \widetilde{W})\|_2.$$

The bounds we develop in section 3 will focus on a slightly different measure between W and $\widetilde{W}$. Specifically, we will be concerned with the *rowwise* error metric

$$\min\left\{\|\widetilde{W}U - W\|_{2,\infty} : U \in \mathbf{O}^k\right\}, \tag{2.2}$$

where the minimization over orthogonal matrices ensures the metric is appropriate for subspaces (as opposed to relying on a specific choice of basis). Further motivation for this metric is encapsulated by the following proposition.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

PROPOSITION 2.1. *Given two orthonormal bases W and $\widetilde{W}$, for any $w \in \mathsf{ran}W$ there exists a $\widetilde{w} \in \mathsf{ran}\widetilde{W}$ such that*

$$\|w - \widetilde{w}\|_\infty \leq \min\{\|\widetilde{W}U - W\|_{2,\infty} : U \in \mathbf{O}^k\}.$$

In contrast, the analogy to Proposition 2.1 for traditional subspace distances would provide a bound on $\|w - \widetilde{w}\|_2$.

Conceptually, the metric (2.2) is closely related to the so-called orthogonal Procrustes problem

$$\min\{\|\widetilde{W}U - W\|_F : U \in \mathbf{O}^k\}, \tag{2.3}$$

which is wellstudied and has a known solution easily computable via the SVD of $\widetilde{W}^T W$ (see, e.g., [14, section 12.4]). In fact, the orthogonal Procrustes problem can also be related to subspace distance since (see, e.g., [4] and [14, section 12.4])

$$\|\sin\Theta(W, \widetilde{W})\|_2 \leq \min\{\|\widetilde{W}U - W\|_F : U \in \mathbf{O}^k\} \leq \sqrt{2k}\|\sin\Theta(W, \widetilde{W})\|_2.$$

Notably, the entrywise definitions of $\|\cdot\|_{2,\infty}$ and $\|\cdot\|_F$ immediately show it is plausible that (2.2) can be considerably (by a factor of $1/\sqrt{n}$) smaller than (2.3). In short, using $\|\cdot\|_{2,\infty}$ allows us to understand how well the error is distributed over the rows.

2.4. Applications of two-to-infinity distances. From an application perspective there are many reasons why we may be interested in distances measured via $\|\cdot\|_{2,\infty}$. For example, spectral algorithms for clustering interpret invariant subspaces rowwise to build low-dimensional embeddings of nodes in a graph or points in a point cloud [27] (specifically, rows of V_1 correspond to an r-dimensional embedding of the graph nodes or data points). Therefore, concrete analysis of the performance of spectral algorithms on model problems or justification of observed performance on real-world data benefits from "pointwise" control over perturbations to the embedding.

Concretely, related work has been used to show spectral methods are information theoretically optimal for the stochastic block model with two blocks in the regime where node degree grows logarithmically with graph size [2]. (See [1] for a more general perspective on recent advances for this problem and [18] for how random matrix theory and concentration results play a role in community detection.) Algorithmically, it is significantly more difficult to analyze situations with more than two blocks, but we believe our results paired with specific formulations of spectral algorithms [9] may allow for more general analysis. Other problems that benefit from such bounds include so-called synchronization problems [21, 23], recovery of unknown mixture components via spectral methods [28], and analysis of algorithmic performance on more general graph models such as random dot product graphs [3].

2.5. Separation of matrices. An important concept for our work is the separation of matrices in various norms. Specifically, for any two matrices $B \in \mathbf{R}^{\ell\times\ell}$ and $C \in \mathbf{R}^{m\times m}$ and norm $\|\cdot\|_*$ on $\mathbf{R}^{m\times\ell}$, define[3]

$$\mathsf{sep}_*(B, C) = \inf\{\|ZB - CZ\|_* : \|Z\|_* = 1\}.$$

[3]This specific form of sep differs slightly notationally, though not mathematically, from the standard way it is written for the two or Frobenius norm. Since we will ultimately be dealing with norms where $\|B\|_* \neq \|B^T\|_*$ this definition is required for consistency.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Perhaps the most pervasive use of sep is in traditional invariant subspace perturbation theory. In fact, for normal matrices B and C

$$\mathsf{sep}_F(B,C) = \min_{\lambda\in\Lambda(B),\mu\in\Lambda(C)} |\lambda-\mu|,$$

thereby recovering the commonly used notion of an eigengap (see, e.g., [25]).

Importantly, and in alignment with our algebraic ordering of eigenvalues above, sep is shift invariant in any norm, i.e.,

$$\mathsf{sep}_*(B+\xi I, C+\xi I) = \mathsf{sep}_*(B,C)$$

for any $\xi \in \mathbf{R}$. Furthermore, in any unitarily invariant norm sep is relatively insensitive to perturbations of small spectral norm. In other words (see Proposition 2.1 of [16])

$$\mathsf{sep}_2(B+E_B, C+E_C) \geq \mathsf{sep}_2(B,C) - \|E_B\|_2 - \|E_C\|_2$$

and

$$\mathsf{sep}_F(B+E_B, C+E_C) \geq \mathsf{sep}_F(B,C) - \|E_B\|_2 - \|E_C\|_2$$

for $E_B \in \mathbf{R}^{\ell\times\ell}$ and $E_C \in \mathbf{R}^{m\times m}$. Lastly, given diagonal matrices it is possible to control sep in a variety of norms necessary for our work. These bounds are captured collectively in Lemma 2.2 (the results about sep_2 and sep_F are well known).

LEMMA 2.2. *Let $D_1 \in \mathbf{R}^{\ell\times\ell}$ and $D_2 \in \mathbf{R}^{m\times m}$ be diagonal matrices, and assume that $\lambda_{\min}(D_1) \geq \lambda_{\max}(D_2)$; then*

$$\mathsf{sep}_2(D_1,D_2) = \mathsf{sep}_F(D_1,D_2) = \mathsf{sep}_{2,\infty}(D_1,D_2) = \lambda_{\min}(D_1) - \lambda_{\max}(D_2). \tag{2.4}$$

Proof. We defer the proof to Appendix A.1. ∎

In addition to the above "canonical" definition of separation, some of our bounds requires that we introduce a slight variant of sep. In particular, we will occasionally consider the separation measured only over matrices in a linear subspace. More specifically, let $W \in \mathbf{R}^{n\times k}$ be an orthonormal basis for a k-dimensional linear subspace, and define

$$\mathsf{sep}_{*,W}(B,C) = \inf\{\|ZB - CZ\|_* : Z \in \mathsf{ran}W, \|Z\|_* = 1\}.$$

It is immediate that $\mathsf{sep}_{*,W}(B,C) \geq \mathsf{sep}_*(B,C)$ for any W, and therefore, as will become evident, consideration of this restricted version of sep can only improve our bounds. For us, the key use of this restricted separation quantity will be when $C = WD_2W^T$ for some diagonal matrix D_2.[4]

In anticipation of its use later, we prove some results about this restricted version of sep analogous to our earlier statements. First, we generalize the notion of sep being shift invariant in Lemma 2.3.

LEMMA 2.3. *Consider $B \in \mathbf{R}^{\ell\times\ell}$ and $C \in \mathbf{R}^{m\times m}$, and let $W \in \mathbf{R}^{n\times m}$ with $n \geq m$ be a matrix with orthonormal columns. Then,*

$$\mathsf{sep}_{*,W}\left(B+\xi I, WCW^T + \xi WW^T\right) = \mathsf{sep}_{*,W}\left(B, WCW^T\right)$$

for any $\xi \in \mathbf{R}$.

[4] One consequence of this restriction is that it will allow us to eliminate any artificial requirement that Λ_1 be separated from zero when A is not low-rank.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Proof. The proof follows immediately from the fact that for any $Z \in \mathsf{ran}W$ $\xi WW^T Z = \xi Z$. ∎

Perhaps more importantly, and as illustrated in Lemma 2.4, in any unitarily invariant norm we can relate this restricted version of sep directly to $\mathsf{sep}(B, C)$.

LEMMA 2.4. *Consider $B \in \mathbf{R}^{\ell\times\ell}$ and $C \in \mathbf{R}^{m\times m}$, and let $W \in \mathbf{R}^{n\times m}$ with $n \geq m$ be a matrix with orthonormal columns. Then, for any unitarily invariant norm $\|\cdot\|_*$*

$$\mathsf{sep}_{*,W}\left(B, WCW^T\right) = \mathsf{sep}_*(B, C)$$

for any $\xi \in \mathbf{R}$.

Proof. We first rewrite the infimum over $\mathsf{ran}W$ in terms of coefficients X of Z in the orthonormal basis W as

$$\mathsf{sep}_{*,W}\left(B, WCW^T\right) = \inf\{\|WXB - WCW^TWX\|_* : \|X\|_* = 1\},$$

where we have used the fact that $\|\cdot\|_*$ is unitarily invariant to say that $\|WX\|_* = \|X\|_*$. Using $W^TW = I$ and the unitary invariance of $\|\cdot\|_*$ once more concludes the proof. ∎

We now use these results to control a restricted version of $\mathsf{sep}_{(2,\infty)}$ (something that will be of particular importance to us) in terms of more traditional and directly interpretable quantities. These results are encapsulated in Lemma 2.5, and we stress that they are worse case bounds that may be far from achieved in practice or provably loose in specific cases.[5] Nevertheless, the fact that the $(2,\infty)$-norm is not unitarily invariant from the left significantly changes the landscape of possible outcomes.

LEMMA 2.5. *Consider $B \in \mathbf{R}^{\ell\times\ell}$ and $C \in \mathbf{R}^{m\times m}$, and let $W \in \mathbf{R}^{n\times m}$ with $n \geq m$ be a matrix with orthonormal columns. Then,*

$$\mathsf{sep}_{(2,\infty),W}\left(B, WCW^T\right) \geq \max\left\{\frac{1}{\sqrt{n}}, \beta_W\right\} \mathsf{sep}_F(B, C),$$

where

$$\beta_W = \inf\left\{\frac{\|WX\|_{2,\infty}}{\|W\|_{2,\infty}} : \|X\|_F = 1\right\}.$$

Proof. We prove two lower bounds on $\mathsf{sep}_{(2,\infty),W}(B, WCW^T)$ that always hold and then maximize over them. First, observe that for any $Z \in \mathsf{ran}W$ there exists an X such that $Z = WX$; hence

$$\begin{aligned}\frac{\|ZB - WCW^TZ\|_{2,\infty}}{\|Z\|_{2,\infty}} &= \frac{\|WXB - WCX\|_{2,\infty}}{\|WX\|_{2,\infty}} \\ &\geq \frac{\|W(XB - CX)\|_F}{\sqrt{n}\|X\|_F} \\ &\geq \frac{\mathsf{sep}_F(B, C)}{\sqrt{n}},\end{aligned}$$

where we have used that $\frac{1}{\sqrt{n}}\|A\|_F \leq \|A\|_{2,\infty} \leq \|A\|_F$ for any $A \in \mathbf{R}^{n\times m}$.

[5] Concrete examples are when $C = 0$ or $W = I$ and B and C are diagonal (see Lemma 2.2) in which case $\mathsf{sep}_F = \mathsf{sep}_2 = \mathsf{sep}_{2,\infty}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Once again using that there exists an X such that $Z = WX$ we see that

$$\begin{aligned}\frac{\|ZB - WCW^TZ\|_{2,\infty}}{\|Z\|_{2,\infty}} &= \frac{\|WXB - WCX\|_{2,\infty}}{\|WX\|_{2,\infty}} \\ &\geq \frac{\|W(XB - CX)\|_{2,\infty}}{\|W\|_{2,\infty}\|X\|_F} \\ &\geq \frac{\|WT\|_{2,\infty}}{\|W\|_{2,\infty}},\end{aligned}$$

where $T = (XB - CX)/\|X\|_F$. Since $\|T\|_F \geq \mathsf{sep}_F(B, C)$ we take an infimum of the lower bound over X to conclude the proof. ∎

Lastly, in Lemma 2.6 we provide a direct bound on $\mathsf{sep}_{2,\infty}$ in terms of traditional matrix norms. In some situations, this may provide the most direct control over $\mathsf{sep}_{2,\infty}$ while in others it may be vacuous and one must resort to Lemma 2.5 instead.

LEMMA 2.6. *Consider $B \in \mathbf{R}^{\ell\times\ell}$ and $C \in \mathbf{R}^{m\times m}$, and let $W \in \mathbf{R}^{n\times m}$ with $n \geq m$ be a matrix with orthonormal columns. Then,*

$$\mathsf{sep}_{(2,\infty),W}\left(B, WCW^T\right) \geq \sigma_{\min}(B) - \|WCW^T\|_\infty,$$

where $\sigma_{\min}(B)$ is the minimal singular value of B.

Proof. Staring from $\mathsf{sep}_{(2,\infty),W}(B, WCW^T) \geq \mathsf{sep}_{(2,\infty)}(B, WCW^T)$ we have that

$$\begin{aligned}\frac{\|ZB - WCW^TZ\|_{2,\infty}}{\|Z\|_{2,\infty}} &\geq \frac{\|ZB\|_{2,\infty} - \|WCW^TZ\|_{2,\infty}}{\|Z\|_{2,\infty}} \\ &\geq \frac{\sigma_{\min}\|Z\|_{2,\infty} - \|WCW^T\|_\infty\|Z\|_{2,\infty}}{\|Z\|_{2,\infty}} \\ &\geq \sigma_{\min}(B) - \|WCW^T\|_\infty.\end{aligned}$$ ∎

3. Main result. Given the notation and concepts from section 2 we may now proceed to present our core results bounding $\|\cdot\|_{2,\infty}$ changes in invariant subspaces of symmetric matrices A under symmetric perturbation; the proofs appear in section 3.2. Notably, our main result includes bounds for a specific $U \in \mathbf{O}^r$, not just the minimum over all orthogonal matrices.

THEOREM 3.1. *Let $A \in \mathbf{R}^{n\times n}$ be symmetric with eigendecomposition*

$$A = V_1\Lambda_1V_1^T + V_2\Lambda_2V_2^T$$

following the conventions of (2.1) *and*

$$\mathsf{gap} = \min\left\{\mathsf{sep}_2(\Lambda_1, \Lambda_2), \mathsf{sep}_{(2,\infty),V_2}\left(\Lambda_1, V_2\Lambda_2V_2^T\right)\right\}.$$

If $\|E\|_2 \leq \frac{\mathsf{gap}}{5}$, then

$$\begin{aligned}\|\widehat{V}_1\widetilde{U} - V_1\|_{2,\infty} \leq 8\|V_1\|_{2,\infty}&\left(\frac{\|E_{2,1}\|_2}{\mathsf{sep}_2(\Lambda_1, \Lambda_2)}\right)^2 \\ &+ 2\frac{\|V_2E_{2,1}\|_{2,\infty}}{\mathsf{gap}} + 4\frac{\|V_2E_{2,2}V_2^T\|_{2,\infty}\|E_{2,1}\|_2}{\mathsf{gap}\times\mathsf{sep}_2(\Lambda_1, \Lambda_2)},\end{aligned} \tag{3.1}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

where $\widehat{V}_1$ is any matrix with orthonormal columns whose range is the dominant r-dimensional invariant subspace of $\widehat{A}$, and $\widetilde{U}$ solves the orthogonal Procrustes problem

$$\min\{\|\widehat{V}_1 U - V_1\|_F : U \in \mathbf{O}^r\}.$$

COROLLARY 3.2. *Following the notation of Theorem* 3.1 *and under the same assumptions, we have that*

$$\min\{\|\widehat{V}_1 U - V_1\|_{2,\infty} : U \in \mathbf{O}^r\} \leq 8\|V_1\|_{2,\infty}\left(\frac{\|E_{2,1}\|_2}{\mathsf{sep}_2(\Lambda_1,\Lambda_2)}\right)^2 + 2\frac{\|V_2 E_{2,1}\|_{2,\infty}}{\mathsf{gap}} + 4\frac{\|V_2 E_{2,2} V_2^T\|_{2,\infty}\|E_{2,1}\|_2}{\mathsf{gap}\times\mathsf{sep}_2(\Lambda_1,\Lambda_2)}, \tag{3.2}$$

where $\widehat{V}_1$ is any matrix with orthonormal columns whose range is the dominant r-dimensional invariant subspace of $\widehat{A}$.

First, we briefly remark on the assumptions and implications of Theorem 3.1. The condition $\|E\|_2 \leq \frac{\mathsf{gap}}{5}$ is standard in the literature; it ensures the two parts of $\Lambda(\widehat{A})$ corresponding to the r-largest eigenvalues and the $n-r$ smallest eigenvalues of A are disjoint.[6] The first term on the right-hand side is also expected, it looks like a traditional Davis–Kahan bound reduced by of the incoherence of V_1 and an additional factor of $\|E\|_2$. The second term captures how (in)coherent $V_2 V_2^T E V_1$ is, a term we often expect to be well controlled. Lastly, the third term is controlled by the incoherence of E itself (relative to its spectral norm).[7]

As we will see later, it is often the case that either the second or third term dominates the upper bound. In fact, in section 3.3 we provide an illustrative example showing that both terms are necessary as part of our bound and are effectively tight. Furthermore, many random models for E have the property that $\|V_2 E_{2,2} V_2^T\|_{2,\infty}$ and $\|E_{2,1}\|_2$ are on the same order, so in these settings using Theorem 3.1 may only marginally improve on the classical Davis–Kahan bound. However, the third term can be more sharply controlled when E is drawn from suitable random models by modifying the proof—Theorem 3.6 explicitly shows how such an improvement is constructed. Lastly, in section 3.3 we will argue that the presence of $\mathsf{sep}_{(2,\infty),V_2}(\Lambda_1, V_2\Lambda_2 V_2^T)$ is essential, though Lemma 2.5 provides some control over it via more interpretable quantities.

Remark 2. Of particular note, when A is rank r and, therefore, $\Lambda_2 = 0$ Theorem 3.1 simplifies significantly since

$$\mathsf{gap} = \mathsf{sep}_2(\Lambda_1,\Lambda_2) = \mathsf{sep}_{(2,\infty),V_2}(\Lambda_1, V_2\Lambda_2 V_2^T) = \min_{i=1,\ldots,r}|\lambda_i|.$$

Remark 3. Equation (3.1) of Theorem 3.1 is a particularly useful result since there are circumstances where it is possible to estimate $\widetilde{U}$ given only $\widehat{V}_1$ and some structural assumptions about V_1. Algorithms based around this paradigm have been developed for spectral clustering [9] and localization of basis functions in Kohn–Sham density functional theory [7, 8].

Prior to embarking on a proof of the main result, we present a corollary of independent interest. Corollary 3.3 simplifies our result in the case where the infinity

[6] Technically, the constant in the denominator just needs to be bigger than 4.

[7] Note that if we define $\mu = \sqrt{n}\|V_1\|_{2,\infty}$ it is possible to further simplify the bound by observing that $\|V_2 V_2^T E\|_{2,\infty} \leq \|V_2 V_2^T\|_\infty \|E\|_{2,\infty} \leq (1+\mu^2)\|E\|_{2,\infty}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

norm of E is sufficiently bounded relative to the spectral gap and the incoherence of V_1. This assumption allows us to remove the third term in the upper bound of Theorem 3.1.

COROLLARY 3.3. *Let $A \in \mathbf{R}^{n\times n}$ be symmetric with eigendecomposition*

$$A = V_1\Lambda_1 V_1^T + V_2\Lambda_2 V_2^T$$

following the conventions of (2.1)*,*

$$\mathsf{gap} = \min\{\mathsf{sep}_2(\Lambda_1,\Lambda_2), \mathsf{sep}_{(2,\infty),V_2}(\Lambda_1, V_2\Lambda_2 V_2^T)\},$$

and $\mu = \sqrt{n}\|V_1\|_{2,\infty}$. If $\|E\|_2 \le \frac{\mathsf{gap}}{5}$ and $\|E\|_\infty \le \mathsf{gap}/(4+4\mu^2)$, then

$$\|\widehat{V}_1\widetilde{U} - V_1\|_{2,\infty} \le 8\|V_1\|_{2,\infty}\left(\frac{\|E_{2,1}\|_2}{\mathsf{sep}_2(\Lambda_1,\Lambda_2)}\right)^2 + 4\frac{\|V_2E_{2,1}\|_{2,\infty}}{\mathsf{gap}},$$

where $\widehat{V}_1$ is any matrix with orthonormal columns whose range is the dominant r-dimensional invariant subspace of $\widehat{A}$ and $\widetilde{U}$ solves the orthogonal Procrustes problem

$$\min\{\|\widehat{V}_1U - V_1\|_F : U \in \mathbf{O}^r\}.$$

3.1. Related work. The most closely related results to our own are the two-to-infinity bounds in [6], though other results exist for single eigenvectors [11] and for similar, though distinct, measures of subspace perturbations [2]. The results in [6] concern orthonormal bases of the singular subspaces of (possibly nonsymmetric) matrices. However, when specialized to orthonormal bases of the invariant subspaces of symmetric matrices our results lead to sharper bounds. Specifically, in [6] the authors establish a general decomposition of $\widehat{V}_1\widetilde{U} - V_1$, where $\widetilde{U}$ solves

$$\min\{\|\widehat{V}_1U - V_1\|_F : U \in \mathbf{O}^r\},$$

and deduce general bounds on $\|\widehat{V}_1\widetilde{U} - V_1\|_{2,\infty}$ through repeated use of the triangle inequality.

More concretely, the first bound in Theorem 3.7 of [6] is most similar to our main result, albeit proved in a significantly different manner. For symmetric positive definite matrices, combining Theorems 3.7 and 6.9 in [6] (and recasting them using our notation) yields

$$\|\widehat{V}_1\widetilde{U} - V_1\|_{2,\infty} \le 4\|V_1\|_{2,\infty}\left(\frac{\|E\|_2}{\mathsf{sep}_2(\Lambda_1,\Lambda_2)}\right)^2 + 2\frac{\|V_2E_{2,1}\|_{2,\infty}}{\lambda_r} + 4\frac{\|V_2E_{2,2}V_2^T\|_{2,\infty}\|E\|_2}{\lambda_r\mathsf{sep}_2(\Lambda_1,\Lambda_2)} + 4\frac{\|V_2\Lambda_2V_2^T\|_{2,\infty}\|E\|_2}{\lambda_r\mathsf{sep}_2(\Lambda_1,\Lambda_2)}. \tag{3.3}$$

In the case where A has rank r (3.3) is comparable to our main result.[8] However, if A is not low-rank our result implies tighter upper bounds. In this case, $\|V_2\Lambda_2V_2^T\|_{2,\infty}$ is nonzero, and the right-hand side of the bound from [6] in (3.3) is dominated by a term depending on $\|E\|_2$—the same behavior immediately implied by Davis–Kahan. If, for example, $E \in \mathbf{R}^{n\times n}$ is a matrix of iid $\mathcal{N}(0, 1/n^2)$ random scalars, their bound

[8] While we are able to use the potentially smaller quantity $\|E_{2,1}\|_2$ in place of $\|E\|_2$, for many random models on E these two quantities behave essentially the same.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

implies $\|\widehat{V}_1\widetilde{U} - V_1\|_{2,\infty}$ vanishes at the rate $\tilde{\mathcal{O}}\left(1/\sqrt{n}\right)$. On the other hand, our bound shows that $\|\widehat{V}_1\widetilde{U} - V_1\|_{2,\infty}$ vanishes at the faster rate $\tilde{\mathcal{O}}(1/n)$—an observation illustrated in section 4.

In [2] the authors develop similar results to Theorem 3.1 as corollaries to their main results. Specifically, their final expressions in Theorem 2.1 and Corollary 2.1 of [2] are rowwise perturbations bounds on orthonormal bases of invariant subspaces. However, this is not the main focus of their work, and these results are generally looser than our bounds. Furthermore, in contrast to our results their bounds are not purely deterministic; they rely on probabilistic assumptions on the error matrix E and additional assumptions on A itself.

3.2. Proof of the main result. At a high level, our proof has two parts. In the first part, we develop a specific characterization of $\widehat{V}_1$ parametrized by a matrix $\widehat{X}$ that is a root of a quadratic matrix equation. In the second part, we show that $\|\widehat{V}_1 - V_1\|_{2,\infty}$ is small under our stated assumptions. Throughout the proof we extensively leverage notation from section 2 to refer to projections of arbitrary matrices B with respect to the representations V_1 and V_2 of invariant subspaces associated with A—recall that for any $B \in \mathbf{R}^{n\times n}$ $B_{i,j} = V_i^T B V_j$.

Part 1*:* Our starting point is the bound

$$\min\{\|\widehat{V}_1 U - V_1\|_{2,\infty} : U \in \mathbf{O}^r\} \leq \|\widehat{V}_1\widetilde{U} - V_1\|_{2,\infty},$$

where $\widetilde{U}$ is the solution of the orthogonal Procrustes problem:

$$\min\{\|\widehat{V}_1 U - V_1\|_F : U \in \mathbf{O}^r\}. \tag{3.4}$$

Notably, the solution to this problem is well known and computable given $\widehat{V}_1$ and V_1, which will prove useful in our numerical experiments. More pertinent to our needs at the moment, $\widehat{V}_1\widetilde{U}$ is the closest matrix with orthonormal columns to V_1 in Frobenius norm whose range is the dominant r-dimensional invariant subspace of $\widehat{A}$.

We start by constructing a matrix with orthonormal columns $\widehat{V}_1$ whose range is the dominant r-dimensional invariant subspace of $\widehat{A}$ and a matrix $\widehat{V}_2$ characterizing the orthogonal complement of $\widehat{V}_1$. We will pick $\widehat{V}_1$ such that the solution to (3.4) is the identity, thereby simplifying the remainder of the proof. Nevertheless, any bound for this specific choice of $\widehat{V}_1$ simultaneously holds for any orthonormal basis of $\mathsf{ran}\widehat{V}_1$ since the discrepancy may be formally resolved by solving the orthogonal Procrustes problem. Specifically, consider

$$\widehat{V}_1 = (V_1 + V_2\widehat{X})(I_r + \widehat{X}^T\widehat{X})^{-\frac{1}{2}} \tag{3.5}$$

and

$$\widehat{V}_2 = (V_2 - V_1\widehat{X}^T)(I_{(n-r)} + \widehat{X}\widehat{X}^T)^{-\frac{1}{2}} \tag{3.6}$$

for some $\widehat{X} \in \mathbf{R}^{(n-r)\times r}$.

Remark 4. A clean construction of this characterization is to start with the general formula for an arbitrary invariant subspace $\widehat{V}_1 = V_1 H + V_2 X$ for some $H \in \mathbf{R}^{r\times r}$ and $X \in \mathbf{R}^{(n-r)\times r}$. Requiring that $\widehat{V}_1^T\widehat{V}_1 = I$ ensures that H is nonsingular as long as $\|X\|_2 < 1$, which is guaranteed by our assumptions. Letting $\widehat{X} = XH^{-1}$ the condition that $\widehat{V}_1$ has orthonormal columns shows that

$$H^2 + H\widehat{X}^T\widehat{X}H = I.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Multiplying on the left and right by H^{-1} we conclude that $H^{-2} = I + \widehat{X}^T\widehat{X}$ and arrive at (3.5).

It is not hard to check that $\widehat{V}_1$ and $\widehat{V}_2$ have orthonormal columns and their ranges are complementary subspaces of $\mathbf{R}^n$. Thus $\mathsf{ran}\widehat{V}_1$ is an invariant subspace of $\widehat{A}$ if and only if

$$0 = \widehat{V}_2^T\widehat{A}\widehat{V}_1 = -\widehat{A}_{2,1} + \widehat{X}\widehat{A}_{1,1} - \widehat{A}_{2,2}\widehat{X} + \widehat{X}\widehat{A}_{1,2}\widehat{X}. \tag{3.7}$$

In other words, $\widehat{X}$ is a root of the map $F : \mathbf{R}^{(n-r)\times r} \to \mathbf{R}^{(n-r)\times r}$ defined as

$$F : X \to -\widehat{A}_{2,1} + X\widehat{A}_{1,1} - \widehat{A}_{2,2}X + X\widehat{A}_{1,2}X.$$

We find a root of F by appealing to a Newton-type method (for root-finding). Starting at $X_0 = 0$, we construct the sequence

$$X_{t+1} \leftarrow X_t - S_{\widehat{A}}^{-1}(F(X_t)). \tag{3.8}$$

To characterize the limit of (X_t) we appeal to the Newton–Kantorovich theorem (Theorem B.1). We remark that this construction is similar, but not identical to, that in [24, section 3].

LEMMA 3.4. *As long as* $\|E\|_2 \le \frac{\mathsf{sep}_2(\Lambda_1,\Lambda_2)}{4}$, (X_t) *converges to* $\widehat{X}$ *such that* $\widehat{X}$ *satisfies* (3.7) *and* $\|\widehat{X}\|_2 \le \frac{4\|E_{2,1}\|_2}{\mathsf{sep}_2(\Lambda_1,\Lambda_2)}$.

Proof. We defer the proof to Appendix B.1. ∎

Since $\widehat{X}$ satisfies (3.7), $\mathsf{ran}\widehat{V}_1$ is an invariant subspace of $\widehat{A}$. It remains to show that $\mathsf{ran}\widehat{V}_1$ is the dominant r-dimensional invariant subspace of $\widehat{A}$. We block diagonalize $\widehat{A}$ to obtain

$$\begin{bmatrix}\widehat{V}_1 \mid \widehat{V}_2\end{bmatrix}^T \widehat{A} \begin{bmatrix}\widehat{V}_1 \mid \widehat{V}_2\end{bmatrix} = \begin{bmatrix}\widehat{V}_1^T\widehat{A}\widehat{V}_1 & 0\\ 0 & \widehat{V}_2^T\widehat{A}\widehat{V}_2\end{bmatrix}. \tag{3.9}$$

The first diagonal block is

$$\begin{aligned}\widehat{V}_1^T\widehat{A}\widehat{V}_1 &= (I_r + \widehat{X}^T\widehat{X})^{-\frac{1}{2}}(V_1 + V_2\widehat{X})^T\widehat{A}(V_1 + V_2\widehat{X})(I_r + \widehat{X}^T\widehat{X})^{-\frac{1}{2}}\\ &= (I_r + \widehat{X}^T\widehat{X})^{-\frac{1}{2}}(\widehat{A}_{1,1} + \widehat{A}_{1,2}\widehat{X} + \widehat{X}^T\widehat{A}_{2,1} + \widehat{X}^T\widehat{A}_{2,2}\widehat{X})(I_r + \widehat{X}^T\widehat{X})^{-\frac{1}{2}}.\end{aligned}$$

Recalling $\widehat{X}$ satisfies (3.7), we have

$$\widehat{X}^T\widehat{A}_{2,1} + \widehat{X}^T\widehat{A}_{2,2}\widehat{X} = \widehat{X}^T\widehat{X}\widehat{A}_{1,1} + \widehat{X}^T\widehat{X}\widehat{A}_{1,2}\widehat{X}.$$

Plugging this expression into the right side of the preceding display, we obtain

$$\begin{aligned}\widehat{V}_1^T\widehat{A}\widehat{V}_1 &= (I + \widehat{X}^T\widehat{X})^{-\frac{1}{2}}(\widehat{A}_{1,1} + \widehat{A}_{1,2}\widehat{X} + \widehat{X}^T\widehat{X}\widehat{A}_{1,1} + \widehat{X}^T\widehat{X}\widehat{A}_{1,2}\widehat{X})(I + \widehat{X}^T\widehat{X})^{-\frac{1}{2}}\\ &= (I + \widehat{X}^T\widehat{X})^{-\frac{1}{2}}(I + \widehat{X}^T\widehat{X})(\widehat{A}_{1,1} + \widehat{A}_{1,2}\widehat{X})(I + \widehat{X}^T\widehat{X})^{-\frac{1}{2}}\\ &= (I + \widehat{X}^T\widehat{X})^{\frac{1}{2}}(\widehat{A}_{1,1} + \widehat{A}_{1,2}\widehat{X})(I + \widehat{X}^T\widehat{X})^{-\frac{1}{2}}.\end{aligned}$$

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

In other words, the first diagonal block is similar to $\widehat{A}_{1,1} + \widehat{A}_{1,2}\widehat{X}$. This implies that

$$\begin{aligned}
\Lambda(\widehat{V}_1^T \widehat{A}\widehat{V}_1) &= \Lambda(\widehat{A}_{1,1} + \widehat{A}_{1,2}\widehat{X}) \\
&= \Lambda(\Lambda_1 + E_{1,1} + E_{1,2}\widehat{X}) \\
&\subset \Lambda(\Lambda_1) + (\|E_{1,1}\|_2 + \|E_{1,2}\|_2\|\widehat{X}\|_2)[-1,1] \quad \text{(Bauer–Fike theorem)} \\
&\subset \Lambda(\Lambda_1) + \left(\|E\|_2 + \tfrac{4\|E\|_2^2}{\mathsf{sep}_2(\Lambda_1,\Lambda_2)}\right)[-1,1] \qquad \left(\|\widehat{X}\|_2 \le \tfrac{4\|E\|_2}{\mathsf{sep}_2(\Lambda_1,\Lambda_2)}\right) \\
&\subset \Lambda(\Lambda_1) + 2\|E\|_2[-1,1] \qquad \left(\tfrac{\|E\|_2}{\mathsf{sep}_2(\Lambda_2,\Lambda_2)} \le \tfrac{1}{4}\right),
\end{aligned}$$

where, as before, $\Lambda(A)$ is defined to be the set of eigenvalues of the matrix A.

Similarly, it is possible to show that the second diagonal block is similar to $\widehat{A}_{2,2} - \widehat{X}\widehat{A}_{1,2}$ and

$$\Lambda(\widehat{V}_2^T \widehat{A}\widehat{V}_2) \subset \Lambda(\Lambda_2) + 2\|E\|_2[-1,1].$$

Recalling $\|E\|_2 \le \frac{\mathsf{sep}_2(\Lambda_1,\Lambda_2)}{5}$, we have

$$\begin{aligned}
\min\{\lambda_1 : \lambda_1 \in \Lambda(\widehat{V}_1^T \widehat{A}\widehat{V}_1)\} &\ge \lambda_r - 2\|E\|_2 \\
&> \lambda_{r+1} + 2\|E\|_2 \\
&> \max\{\lambda_2 : \lambda_2 \in \sigma(\widehat{V}_2^T \widehat{A}\widehat{V}_2)\},
\end{aligned}$$

which implies $\mathsf{ran}(\widehat{V}_1)$ is the dominant r-dimensional invariant subspace of $\widehat{A}$ as claimed.

Finally, it is well known that the optimal point $\widetilde{U}$ of the orthogonal Procrustes problem (3.4) is the unitary factor in the polar decomposition of $\widehat{V}_1^T V_1$. Since

$$\begin{aligned}
\widehat{V}_1^T V_1 &= (I + \widehat{X}^T\widehat{X})^{-\frac{1}{2}}(V_1 + V_2\widehat{X})^T V_1 \\
&= (I + \widehat{X}^T\widehat{X})^{-\frac{1}{2}}(V_1^T V_1 + \widehat{X}V_2^T V_1) \\
&= (I + \widehat{X}^T\widehat{X})^{-\frac{1}{2}}
\end{aligned}$$

is symmetric and positive definite, that unitary factor is the identity. Therefore, as desired, $\widehat{V}_1$ is the closest matrix to V_1 in Frobenius distance among all matrices of the form $\widehat{V}_1 U$, where $U \in \mathbf{O}^r$. Note that this is exactly the set of matrices with orthonormal columns whose range is the dominant r-dimensional invariant subspace of $\widehat{A}$.

Part 2*:* For the remainder of the proof $\widehat{V}_1$ is as defined in (3.5) and we proceed to explicitly bound

$$\|\widehat{V}_1 - V_1\|_{2,\infty}.$$

We start by decomposing the error $\widehat{V}_1 - V_1$ into its components in $\mathsf{ran}V_1$ and $(\mathsf{ran}V_1)^\perp$ (recall that $\mathsf{ran}V_2 = (\mathsf{ran}V_1)^\perp$). Specifically, from (3.5) it follows that

$$\widehat{V}_1 - V_1 = V_1((I_r + \widehat{X}^T\widehat{X})^{-\frac{1}{2}} - I_r) + V_2\widehat{X}(I + \widehat{X}^T\widehat{X})^{-\frac{1}{2}}. \tag{3.10}$$

We now proceed to address each part of this decomposition of the error separately.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

The $(2,\infty)$-norm of the first term on the right side of (3.10) is at most

$$\|V_1((I_r + \widehat{X}^T\widehat{X})^{-\frac{1}{2}} - I_r)\|_{2,\infty} \leq \|V_1\|_{2,\infty}\|(I_r + \widehat{X}^T\widehat{X})^{-\frac{1}{2}} - I_r\|_2.$$

Since

$$\|(I_r + \widehat{X}^T\widehat{X})^{-\frac{1}{2}} - I_r\|_2 = |1 - (1 + \|\widehat{X}\|_2^2)^{-1/2}|,$$

we can use the fact that for any $x > 0$

$$\begin{aligned} |1 - (1+x)^{-1/2}| &= \left|\frac{\sqrt{1+x} - 1}{\sqrt{1+x}}\right| \\ &\leq \left|\frac{x}{\sqrt{1+x}+1}\right| \\ &\leq \frac{1}{2}x \end{aligned}$$

to conclude that

$$\|(I_r + \widehat{X}^T\widehat{X})^{-\frac{1}{2}} - I_r\|_2 \leq \frac{1}{2}\|\widehat{X}\|_2^2.$$

Furthermore, by Lemma 3.4 $\|\widehat{X}\|_2 \leq \frac{4\|E_{2,1}\|_2}{\mathsf{sep}_2(\Lambda_1,\Lambda_2)}$, and therefore

$$\|V_1((I_r + \widehat{X}^T\widehat{X})^{-\frac{1}{2}} - I_r)\|_{2,\infty} \leq 8\|V_1\|_{2,\infty}\left(\frac{\|E_{2,1}\|_2}{\mathsf{sep}_2(\Lambda_1,\Lambda_2)}\right)^2.$$

At this point we have control over the first term in (3.10). In addition, since the $(2,\infty)$-norm of the second term on the right side of (3.10) is at most

$$\|V_2\widehat{X}(I_r + \widehat{X}^T\widehat{X})^{-\frac{1}{2}}\|_{2,\infty} \leq \|V_2\widehat{X}\|_{2,\infty}\|(I_r + \widehat{X}^T\widehat{X})^{-\frac{1}{2}}\|_2 \leq \|V_2\widehat{X}\|_{2,\infty}$$

it follows from the triangle inequality that

$$\|\widehat{V}_1 - V_1\|_{2,\infty} \leq 8\|V_1\|_{2,\infty}\left(\frac{\|E_{2,1}\|_2}{\mathsf{sep}_2(\Lambda_1,\Lambda_2)}\right)^2 + \|V_2\widehat{X}\|_{2,\infty}. \tag{3.11}$$

For the remainder of the proof we focus on bounding $\|V_2\widehat{X}\|_{2,\infty}$. At first glance, we are tempted to appeal to the compatibility of $\|\cdot\|_{2,\infty}$ and $\|\cdot\|_2$ to obtain

$$\|V_2\widehat{X}\|_{2,\infty} \leq \|V_2\|_{2,\infty}\|\widehat{X}\|_2 \leq \|V_2\|_{2,\infty}\frac{2\|E_{2,1}\|_2}{\mathsf{sep}_2(\Lambda_1,\Lambda_2)}.$$

Unfortunately, this bound is generally inadequate because $\|V_2\|_{2,\infty}$ may be much larger than $\|V_1\|_{2,\infty}$.[9] Instead, we must study $\|V_2\widehat{X}\|_{2,\infty}$ directly. To start, observe that $V_2\widehat{X}$ satisfies

$$0 = -V_2\widehat{A}_{2,1} + V_2\widehat{X}\widehat{A}_{1,1} - V_2\widehat{A}_{2,2}V_2^TV_2\widehat{X} + V_2\widehat{X}\widehat{A}_{1,2}V_2^TV_2\widehat{X}. \tag{3.12}$$

In other words, $V_2\widehat{X}$ is a root of the map $G : \mathbf{R}^{n\times r} \to \mathbf{R}^{n\times r}$ defined as

$$G : Y \to -V_2\widehat{A}_{2,1} + Y\widehat{A}_{1,1} - V_2\widehat{A}_{2,2}V_2^TY + Y\widehat{A}_{1,2}V_2^TY.$$

[9] In fact, we have that $\|\begin{bmatrix} V_1 & V_2\end{bmatrix}\|_{2,\infty} = 1$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Letting $\widehat{Y} = V_2\widehat{X}$, we rearrange (3.12) to obtain

$$\begin{aligned}\widehat{Y}\widehat{A}_{1,1} - V_2\Lambda_2 V_2^T\widehat{Y} &= -V_2\widehat{A}_{2,1} + \widehat{Y}\widehat{A}_{1,2}V_2^T\widehat{Y} + V_2E_{2,2}V_2^T\widehat{Y} \\ &= -V_2E_{2,1} + \widehat{Y}E_{1,2}V_2^T\widehat{Y} + V_2E_{2,2}V_2^T\widehat{Y},\end{aligned}$$

where we have used that $A_{1,2} = A_{2,1}^T = 0$.

We take norms to see that

$$\begin{aligned}&\|\widehat{Y}\widehat{A}_{1,1} - V_2\Lambda_2V_2^T\widehat{Y}\|_{2,\infty} \\ &\quad\le \|V_2E_{2,1}\|_{2,\infty} + \|\widehat{Y}E_{1,2}V_2^T\widehat{Y}\|_{2,\infty} + \|V_2E_{2,2}V_2^T\widehat{Y}\|_{2,\infty} \\ &\quad\le \|V_2E_{2,1}\|_{2,\infty} + \|\widehat{Y}\|_{2,\infty}\|\widehat{Y}^TV_2E_{1,2}^T\|_2 + \|V_2E_{2,2}V_2^T\widehat{Y}\|_{2,\infty} \\ &\quad\le \|V_2E_{2,1}\|_{2,\infty} + \|\widehat{Y}\|_{2,\infty}\|V_2E_{1,2}^T\|_2\|\widehat{Y}\|_2 + \|V_2E_{2,2}V_2^T\widehat{Y}\|_{2,\infty}.\end{aligned}$$

It now follows from Lemma 3.4 that

$$\begin{aligned}&\|\widehat{Y}\widehat{A}_{1,1} - V_2\Lambda_2V_2^T\widehat{Y}\|_{2,\infty} \\ &\quad\le \|V_2E_{2,1}\|_{2,\infty} + \|\widehat{Y}\|_{2,\infty}\frac{2\|E_{2,1}\|_2^2}{\mathsf{gap}} + \|V_2E_{2,2}V_2^T\widehat{Y}\|_{2,\infty}.\end{aligned}$$

Next, observe that the left side is at least

$$\begin{aligned}\|\widehat{Y}\widehat{A}_{1,1} - V_2\Lambda_2V_2^T\widehat{Y}\|_{2,\infty} &\ge \|\widehat{Y}\Lambda_1 - V_2\Lambda_2V_2^T\widehat{Y}\|_{2,\infty} - \|\widehat{Y}\|_{2,\infty}\|E_{1,1}\|_2 \\ &\ge \mathsf{sep}_{(2,\infty),V_2}(\Lambda_1, V_2\Lambda_2V_2^T)\|\widehat{Y}\|_{2,\infty} - \|\widehat{Y}\|_{2,\infty}\|E_{1,1}\|_2 \\ &\ge \mathsf{sep}_{(2,\infty),V_2}(\Lambda_1, V_2\Lambda_2V_2^T)\|\widehat{Y}\|_{2,\infty} - \|\widehat{Y}\|_{2,\infty}\|E\|_2 \\ &\ge \frac{3}{4}\mathsf{gap}\|\widehat{Y}\|_{2,\infty},\end{aligned}$$

and, therefore,

$$\left(\frac{3}{4}\mathsf{gap} - \frac{2\|E_{2,1}\|_2^2}{\mathsf{gap}}\right)\|\widehat{Y}\|_{2,\infty} \le \|V_2E_{2,1}\|_{2,\infty} + \|V_2V_2^TE\widehat{Y}\|_{2,\infty}.$$

Since $2\|E_{2,1}\|_2/\mathsf{gap} \le 1$ and $\|E_{2,1}\|_2 \le \mathsf{gap}/4$ by assumption (using $\|E_{2,1}\|_2 \le \|E\|_2$) we have that

$$\|\widehat{Y}\|_{2,\infty} \le \frac{2\|V_2E_{2,1}\|_{2,\infty}}{\mathsf{gap}} + \frac{2\|V_2V_2^TE\widehat{Y}\|_{2,\infty}}{\mathsf{gap}}. \tag{3.13}$$

Prior to concluding the proof, we summarize our results up to this point in Lemma 3.5. We partly pause to highlight a natural launching point for problem specific analysis, particularly in settings where it is possible to control $\|V_2V_2^TE\widehat{Y}\|_{2,\infty}$ in a tighter manner than suggested by our worst-case bounds that follow.

LEMMA 3.5. *Let $A \in \mathbf{R}^{n\times n}$ be symmetric with an eigendecomposition $A = V_1\Lambda_1V_1^T + V_2\Lambda_2V_2^T$ following the conventions of* (2.1) *and*

$$\mathsf{gap} = \min\left\{\mathsf{sep}_2(\Lambda_1,\Lambda_2), \mathsf{sep}_{(2,\infty),V_2}\left(\Lambda_1, V_2\Lambda_2V_2^T\right)\right\}.$$

If $\|E\|_2 \le \frac{\mathsf{gap}}{5}$, then

$$\|\widehat{V}_1\widetilde{U} - V_1\|_{2,\infty} \le 8\|V_1\|_{2,\infty}\left(\frac{\|E_{2,1}\|_2}{\mathsf{sep}_2(\Lambda_1,\Lambda_2)}\right)^2 + 2\frac{\|V_2E_{2,1}\|_{2,\infty} + \|V_2E_{2,2}V_2^T\widehat{Y}\|_{2,\infty}}{\mathsf{gap}},$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

where $\widehat{V}_1$ is any matrix with orthonormal columns whose range is the dominant r-dimensional invariant subspace of $\widehat{A}$, $\widetilde{U}$ solves the orthogonal Procrustes problem

$$\min\left\{\|\widehat{V}_1 U - V_1\|_F : U \in \mathbf{O}^r\right\},$$

and $\widehat{Y} = V_2\widehat{X}$, where $\widehat{X}$ is the root of

$$F : X \to -\widehat{A}_{2,1} + X\widehat{A}_{1,1} - \widehat{A}_{2,2}X + X\widehat{A}_{1,2}X$$

found by the iteration (3.8) *starting at $X_0 = 0$ and thereby satisfying Lemma* 3.4.

Moving forward, Lemma 3.5 immediately implies that

$$\|\widehat{Y}\|_{2,\infty} \le 2\frac{\|V_2 E_{2,1}\|_{2,\infty}}{\mathsf{gap}} + 4\frac{\|V_2 E_{2,2} V_2^T\|_{2,\infty}\|E_{2,1}\|_2}{\mathsf{gap} \times \mathsf{sep}_2(\Lambda_1, \Lambda_2)},$$

where we have used the submultiplicative relationships

$$\|V_2 E_{2,2} V_2^T \widehat{Y}\|_{2,\infty} \le \|V_2 E_{2,2} V_2^T\|_{2,\infty}\|\widehat{Y}\|_2$$

in conjunction with Lemma 3.4 to bound $\|\widehat{Y}\|_2$. This concludes the proof of (3.1) in Theorem 3.1, and Corollary 3.2 follows immediately.

We now briefly retrace our steps to prove Corollary 3.3. In particular, returning to Lemma 3.5 we can instead conclude that

$$\|\widehat{Y}\|_{2,\infty}\left(1 - \frac{2(1+\mu^2)}{\mathsf{gap}}\frac{\|E\widehat{Y}\|_{2,\infty}}{\|\widehat{Y}\|_{2,\infty}}\right) \le \frac{2\|V_2 E_{2,1}\|_{2,\infty}}{\mathsf{gap}},$$

where we have used the observation that

$$\begin{aligned}\|V_2 V_2^T\|_\infty &= \|I_n - V_1 V_1^T\|_\infty \\ &\le 1 + \max\left\{\textstyle\sum_{j=1}^n \left|v_i^T v_j\right| : i \in [n]\right\} \\ &\le 1 + n\|V_1\|_{2,\infty}^2 \\ &\le 1 + \mu^2\end{aligned}$$

and that $\widehat{Y} \in \mathsf{ran}V_2$. Now, since $\|E\widehat{Y}\|_{2,\infty}/\|\widehat{Y}\|_{2,\infty} \le \|E\|_\infty$ if we further assume that $\|E\|_\infty \le \mathsf{gap}/(4 + 4\mu^2)$ we get that

$$\left(1 - \frac{2(1+\mu^2)}{\mathsf{gap}}\frac{\|E\widehat{Y}\|_{2,\infty}}{\|\widehat{Y}\|_{2,\infty}}\right) \ge \frac{1}{2}.$$

Therefore,

$$\|\widehat{Y}\|_{2,\infty} \le 4\frac{\|V_2 E_{2,1}\|_{2,\infty}}{\mathsf{gap}},$$

which concludes the proof of Corollary 3.3.

3.3. Observations and implications. We now discuss several aspects of our bounds in greater detail. In particular, we first construct specific examples that show any of the 3 terms in the bound of Theorem 3.1 may tightly control the error and therefore are all necessary. We then argue why $\mathsf{sep}_{(2,\infty),V_2}(\Lambda_1, V_2\Lambda_2 V_2^T)$ should be directly included in our bounds by showing that in the worst case it may be considerably smaller than $\mathsf{sep}_F(\Lambda_1, \Lambda_2)$. Lastly, we discuss the use of our bound in certain probabilistic scenarios motivated by applications and highlight how our bounds can facilitate further analysis of those situations.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

3.3.1. When the upper bound is tight. The first term of our upper bound represents the projection of the error onto V_1 while the latter two terms arise from the projection onto V_2. Therefore, we focus on the latter piece to understand if both the terms are necessary and examine our potentially loose use of the triangle inequality and submultiplicative bounds in the proof. To accomplish this, we construct a specific example and examine the behavior of our bound.

We bounded the projection of the error onto $\mathsf{ran} V_2$ as

$$\|V_2 V_2^T(\widehat{V}_1 - V_1)\|_{2,\infty} \leq \frac{2\|V_2 E_{2,1}\|_{2,\infty}}{\mathsf{gap}} + 4\frac{\|V_2 E_{2,2} V_2^T\|_{2,\infty}\|E_{2,1}\|_2}{\mathsf{gap} \times \mathsf{sep}_2(\Lambda_1, \Lambda_2)}. \tag{3.14}$$

While the first term on the right-hand side of (3.14) is a natural part of our bound given the quadratic form (3.12), the second term arose from the submultiplicative bound

$$\|V_2 E_{2,2} V_2^T \widehat{Y}\|_{2,\infty} \leq \|V_2 E_{2,2} V_2^T\|_{2,\infty}\|\widehat{Y}\|_2 \leq \frac{2}{\mathsf{sep}_2(\Lambda_1, \Lambda_2)}\|V_2 E_{2,2} V_2^T\|_{2,\infty}\|E_{2,1}\|_2.$$

Nevertheless, both terms are necessary—there are perturbations that saturate each part of the bound.

To show this, we build an example that demonstrates two clear regimes—one where the first term of (3.14) controls the error tightly and one where the second term does. We accomplish this by picking E such that for the resulting $\widehat{Y}$

$$\|E\widehat{Y}\|_{2,\infty} \approx \|V_2 E_{2,2} V_2^T\|_{2,\infty}\frac{\|E\|_2}{\mathsf{gap}}.$$

To keep things simple, we consider the $r = 1$ case with $\lambda_1 = 1$ and $\lambda_2, \ldots, \lambda_n = 0$, which implies $\mathsf{gap} = 1$. We then let $V_1 = \mathbf{1}/\sqrt{n}$ and observe that if $E_{1,2} = 0$, and $E_{1,1} = 0$, then $\widehat{Y}$ satisfies

$$\left(I - V_2 E_{2,2} V_2^T\right)\widehat{Y} = V_2 E_{2,1}.$$

The core insight in our construction is that we can now choose $V_2 E_{2,2} V_2^T$ and $V_2 E_{2,1}$ carefully to accomplish our goal. This is because we can essentially determine $\widehat{Y}$ (in fact, to first order it looks like $V_2 E_{2,1}$ if the norm of E is sufficiently small).

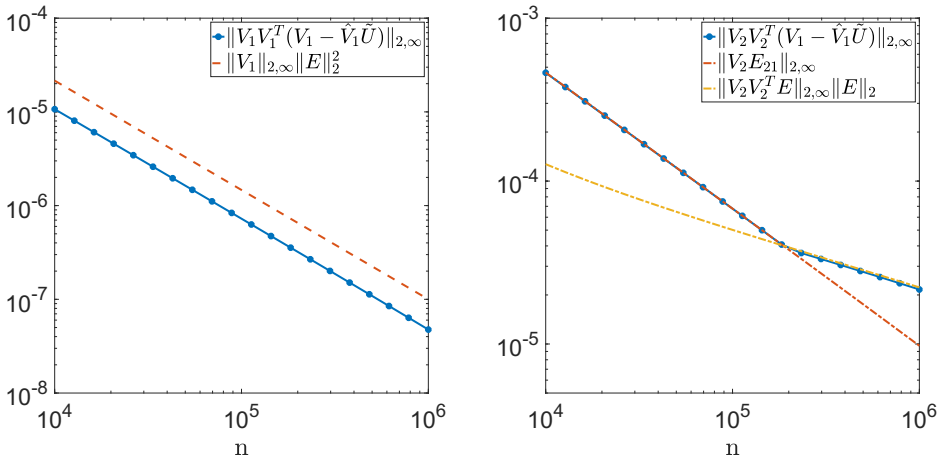
Now, let

$$V_2 E_{2,2} V_2^T = V_2 V_2^T\left(e_1 \mathbf{1}_{\pm}^T/\sqrt{n} + \mathbf{1}_{\pm} e_1^T/\sqrt{n}\right) V_2 V_2^T.$$

In this case there exists a y with $y_1 = 1$ and $y_2, \ldots, y_n = \mathcal{O}(1/\sqrt{n})$ such that $(I - V_2 E_{2,2} V_2^T)y = \mathcal{O}(1/\sqrt{n})$ entrywise. Setting $V_2 E_{2,1}$ to be proportional to $V_2 V_2^T(I - V_2 E_{2,2} V_2^T)y$ lets us deterministically construct an E where $\widehat{Y}$ essentially saturates the submultiplicative bound.[10] The preceding construction yields a purely deterministic example illustrating in Figure 1(b) that either part of (3.14) can be dominant. Similarly, Figure 1(a) shows, as expected, that our bound on the projection of the error onto V_1 tightly captures the asymptotic behavior.

[10] Practically one can make E symmetric by setting $E_{1,2}$ appropriately without destroying the example and scale E by $n^{-1/3}$ so that we expect convergence in n. Details are available in the online materials referred to in the numerical experiments section. Choices of scaling constants in individual parts of E control where the crossover point occurs between the two bounds.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.



(a) Projection onto V_1 (b) Projection onto V_2

FIG. 1. *Asymptotic behavior of* $\min_{\tilde{U}=\pm 1}\|\widehat{V}_1 - V_1\tilde{U}\|_{2,\infty}$ *split into the component of the error in* $\mathsf{ran}V_1$ *and* $\mathsf{ran}V_2$. *This example shows that either part of the upper bound in Theorem* 3.1 *associated with the error projected onto* $\mathsf{ran}V_2$ *can tightly control the rate of decay. Similarly, our control over the projection of the error onto* $\mathsf{ran}V_1$ *matches the observed rate for this example. Note that, as indicated by the legend, we have not included constants (they would appear to be slightly loose in this case) and used the upper bound* $\|V_2E_{2,2}V_2^T\|_{2,\infty} \le \|V_2V_2^TE\|_{2,\infty}$. *Therefore the dotted lines technically represent our upper bounds to within small constants.*

3.3.2. Inclusion of norm specific separation. In the proof of Theorem 3.1 $\mathsf{sep}_{(2,\infty),V_2}(\Lambda_1, V_2\Lambda_2V_2^T)$ arises somewhat naturally. Nevertheless, ideally one would be able to generically relate it tightly to traditional notions of an eigengap. Unfortunately, the lower bound provided in Lemma 2.5 is essentially tight. To show this we explicitly construct an example that achieves (to within a small constant) the lower bound $\mathsf{sep}_{(2,\infty),V_2}(\Lambda_1, V_2\Lambda_2V_2^T) \ge \mathsf{sep}_F(\Lambda_1,\Lambda_2)/\sqrt{n}$.

Assume n is even, and let $\mathbf{1}$ be the vector of all ones and $(\mathbf{1}_\pm)_i = -1$ if $i > n/2$ and 1 otherwise. Now, define $v_1 = [0 \;\; \mathbf{1}^T]^T/\sqrt{n}$ and $v_2 = [0 \;\; \mathbf{1}_\pm^T]^T/\sqrt{n}$, and consider the $(n+1)\times(n+1)$ matrix

$$A = 2 * v_1v_1^T + \begin{bmatrix} e_1 & v_2 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} e_1 & v_2 \end{bmatrix}^T.$$

In our framework this corresponds to setting $\Lambda_1 = 2$, $\Lambda_2 = \mathsf{diag}(1,-1,0,\ldots,0) \in \mathbf{R}^{n\times n}$ and letting V_2 to be any matrix with orthonormal columns spanning the orthogonal complement of v_1 such that

$$\begin{bmatrix} e_1 & v_2 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} e_1 & v_2 \end{bmatrix}^T = V_2\Lambda_2V_2^T.$$

In this case, by picking the vector $q = e_1 + 2v_2$ (which is in the range of V_2 and satisfies $\|q\|_{2,\infty} = 1$ if $n \ge 4$) we see that

$$\begin{aligned} \|q\Lambda_1 - V_2\Lambda_2V_2^Tq\|_{2,\infty} &= \|2q - 2e_1 - v_2\|_{2,\infty} \\ &= \|3v_2\|_{2,\infty} \\ &= \frac{3}{\sqrt{n}}. \end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Therefore, $\mathsf{sep}_{(2,\infty),V_2}(\Lambda_1, V_2\Lambda_2 V_2^T) \leq 3/\sqrt{n}$, and since $\mathsf{sep}_F(\Lambda_1, \Lambda_2) = 1$ this shows that Lemma 2.5 is essentially tight. Nevertheless, Lemmas 2.5 and 2.6 also show situations where $\mathsf{sep}_{(2,\infty),V_2}$ and sep_F are more closely related. Ultimately, the range of possible relationships between $\mathsf{sep}_{(2,\infty),V_2}$ and sep_F motivates the inclusion of $\mathsf{sep}_{(2,\infty),V_2}$ directly in any worst-case deterministic bound.

3.3.3. Probabilistic settings. While the two preceding sections illuminate why various terms in our constructed bounds are necessary, one may expect that in random settings these terms can be controlled more effectively and the expected behavior may be far from the worst case. In particular, if we consider $E = \sigma Z$ where Z is symmetric and $Z_{i,j}$ are $\mathsf{iid}\ \mathcal{N}(0,1)$ (up to the symmetry constraint) random variables we have by standard properties of Gaussian random matrices (see, e.g., [26, section 4.4])

$$\mathbb{P}(\|E\|_2 > 3\sigma\sqrt{n}) \lesssim e^{-\frac{n}{2}}.$$

Furthermore, under an assumption that V_1 is incoherent (recall that $\mu = \sqrt{n}\|V_1\|_{2,\infty}$) and E is independent from V_1 we may assert (again via standard properties of Gaussian random matrices and a union bound [26, section 4.4]) that

$$\|E_{2,1}\|_{2,\infty} \leq (1+\mu^2)\|EV_1\|_{2,\infty} \lesssim \sigma\sqrt{\log n}$$

with high probability. Ideally, these results would directly imply that in such a setting

$$\min\{\|\widehat{V}_1 U - V_1\|_{2,\infty} : U \in \mathbf{O}^r\} \lesssim \sigma\sqrt{\log n} \tag{3.15}$$

with high probability. Unfortunately, this does not follow directly from Theorem 3.1, as for certain values of σ the error bound is dominated by the term $\|E\|_{2,\infty}\|E\|_2 \lesssim \sigma^2 n$.

Figure 1 shows that this is not an artifact of our analysis; it is possible to construct examples that saturate the error bound. However, these examples are adversarial. In particular, the independence among the entries of Z permits more direct control of $\|E\widehat{Y}\|_{2,\infty}$, and the authors of [2] appeal to a leave-one-out technique to achieve such direct control. Based on our analysis, we conjecture that their columnwise independence condition may be relaxed to independence among the $E_{i,j}$'s ($i,j \in \{1,2\}$). While we defer a thorough analysis of this problem in the probabilistic setting for future work, Theorem 3.6 illustrates how additional mild assumptions on E can be used to improve our bounds. While we have stated Theorem 3.6 for the Procrustes solution, the result for the minimum over $U \in \mathbf{O}^r$ (analogous to Corollary 3.2) follows immediately.

THEOREM 3.6. *Let $A \in \mathbf{R}^{n\times n}$ be symmetric with eigendecomposition*

$$A = V_1\Lambda_1 V_1^T + V_2\Lambda_2 V_2^T$$

following the conventions of (2.1)*,*

$$\mathsf{gap} = \min\{\mathsf{sep}_2(\Lambda_1,\Lambda_2), \mathsf{sep}_{(2,\infty),V_2}(\Lambda_1, V_2\Lambda_2 V_2^T)\}$$

be independent of n, and $\|V_1\|_{2,\infty} \lesssim 1/\sqrt{n}$. If $E = \sigma Z$ where Z is symmetric, $Z_{i,j}$ are $\mathsf{iid}\ \mathcal{N}(0,1)$ *random variables (up to the symmetry constraint), and $\sigma \lesssim 1/\sqrt{n}$, then with probability $1 - o(1)$*

$$\min\{\|\widehat{V}_1\widetilde{U} - V_1\|_{2,\infty} : U \in \mathbf{O}^r\} \lesssim \sigma^2\sqrt{n} + \sigma\sqrt{\log n} + (\sigma\sqrt{n})^3,$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

where $\widehat{V}_1$ is any matrix with orthonormal columns whose range is the dominant r-dimensional invariant subspace of $\widehat{A}$, and $\widetilde{U}$ solves the orthogonal Procrustes problem

$$\min\{\|\widehat{V}_1 U - V_1\|_F : U \in \mathbf{O}^r\}.$$

Proof. The key idea behind this proof is to consider a second perturbation $\widetilde{E}$ drawn from the same distribution as E conditioned on $\widetilde{E}_{1,2} = E_{1,2}$ and $\widetilde{E}_{1,1} = E_{1,1}$ — nevertheless, $E_{2,2}$ and $\widetilde{E}_{2,2}$ are still independent. Given E and $\widetilde{E}$, appropriate control of $\|E\|_2$ and $\|\widetilde{E}\|_2$ [26, section 4.4] with probability $1 - o(1)$ allows us to invoke Lemma C.1 to get that

$$\begin{aligned}\|\widehat{V}_1\widetilde{U} - V_1\|_{2,\infty} \lesssim \frac{1}{\sqrt{n}}\|E_{2,1}\|_2^2 + \|V_2 E_{2,1}\|_{2,\infty} + \|V_2 E_{2,2} V_2^T \widetilde{Y}\|_{2,\infty} \\ + \|E_{2,2}\|_2 \|V_2 E_{2,2} V_2^T\|_{2,\infty}\|E_{2,1}\|_2 + \|V_2 E_{2,2} V_2^T\|_{2,\infty}\|E_{2,1}\|_2^3,\end{aligned}$$

where $\widetilde{Y} \in \mathbf{R}^{n\times r}$ is in the range of V_2 and satisfies

$$0 = -V_2\widetilde{A}_{2,1} + \widetilde{Y}\widetilde{A}_{1,1} - V_2\widetilde{A}_{2,2}V_2^T\widetilde{Y} + \widetilde{Y}\widetilde{A}_{1,2}V_2^T\widetilde{Y}.$$

This implies that $\widetilde{Y}$ is independent of $E_{2,2}$ and satisfies $\|\widetilde{Y}\|_2 \lesssim \|E_{2,1}\|_2$. The stated bound follows immediately from the following bounds on E that all hold with probability $1 - o(1)$ [26, section 4.4]:

$$\begin{aligned}\|E_{2,1}\|_2 &\lesssim \sigma\sqrt{n} \\ \|V_2 E_{2,1}\|_{2,\infty} &\lesssim \sigma\sqrt{\log n} \\ \|V_2 E_{2,2} V_2^T \widetilde{Y}\|_{2,\infty} &\lesssim \sigma^2\sqrt{n\log n} \\ \|E_{2,2}\|_2 &\lesssim \sigma\sqrt{n} \\ \|V_2 E_{2,2} V_2^T\|_{2,\infty} &\lesssim \sigma\sqrt{n},\end{aligned}$$

where because $\|V_1\|_{2,\infty} \lesssim 1/\sqrt{n}$ we have that $\|V_2 V_2^T\|_\infty \lesssim 1$. ∎

Remark 5. If we consider μ to be constant in n (synonymous with $\|V_1\|_{2,\infty} \lesssim 1/\sqrt{n}$), the rate of convergence of $\min\{\|\widehat{V}_1 U - V_1\|_{2,\infty} : U \in \mathbf{O}^r\}$ implied by Theorem 3.6 is faster than that given by direct application of Theorem 3.1. In particular, we get that $\min\{\|\widehat{V}_1 U - V_1\|_{2,\infty} : U \in \mathbf{O}^r\} \to 0$ as $(\sigma\sqrt{n})^3$ vs $(\sigma\sqrt{n})^2$. While this is an improvement, we believe more intricate probabilistic techniques will be necessary to prove an upper bound that matches our conjectured rate of $\sigma\sqrt{\log n}$ (see (3.15)) in this setting.

4. Numerical simulations. We now provide numerical simulations to illustrate the effectiveness of our bounds and elaborate on a key difference between them and prior work. We consider two settings, one where A is low-rank and one where A is not low-rank and $\|V_2\Lambda_2 V_2^T\|_{2,\infty}$ is constant with respect to n. In all these experiments, and as before, we let $\mathbf{1}$ be the vector of all ones and $(\mathbf{1}_\pm)_i = -1$ be 1 in the first half of the entires and -1 in the second half. We also let $E = \sigma Z$ where Z is a symmetric matrix whose entries are $\mathsf{iid}\ \mathcal{N}(0,1)$. Code to generate these plots (and Figure 1) is available at `https://github.com/asdamle/rowwise-perturbation`.

4.1. A is low-rank. Assume n is even, let

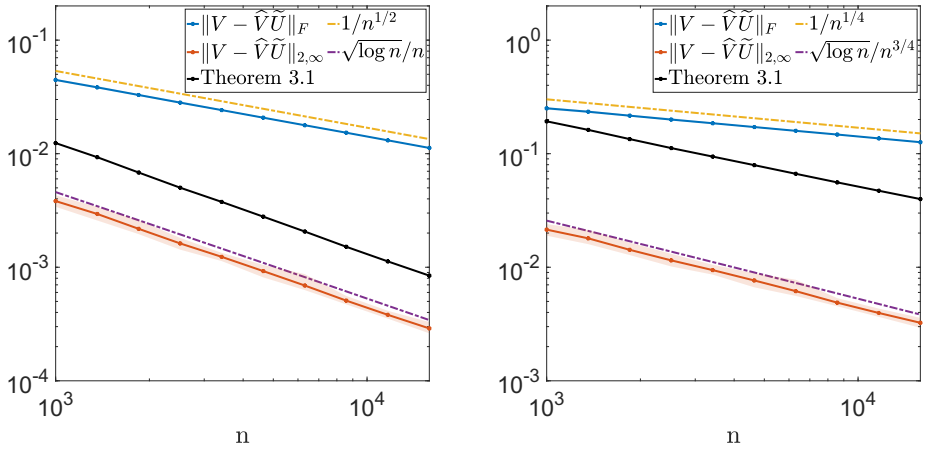
$$V_1 = \frac{1}{\sqrt{n}}\begin{bmatrix}\mathbf{1} & \mathbf{1}_\pm\end{bmatrix},$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

and consider $A = V_1 V_1^T$. In this setting, $\mathsf{gap} = 1$, and if $\sigma = 1/n$ Theorem 3.1 shows that $\min\{\|\widehat{V}_1 U - V_1\|_{2,\infty} : U \in \mathbf{O}^r\} \lesssim_P \frac{\sqrt{\log n}}{n}$, where the term controlling the rate with respect to n is $\|V_2 E_{2,1}\|_{2,\infty}$. Furthermore, Theorem 3.1 shows that $\|\widehat{V}_1 \widetilde{U} - V_1\|_{2,\infty} \lesssim_P \frac{\sqrt{\log n}}{n}$. Therefore, we conduct an experiment where for increasing values of n we construct $\widehat{A}$ for several instances of E, compute $\widehat{V}_1$ and the solution to the orthogonal Procrustes problem, and measure $\|\widehat{V}_1 \widetilde{U} - V_1\|_{2,\infty}$. Figure 2(a) clearly shows the expected behavior for both the traditional subspace distance and our bound on $\|\cdot\|_{2,\infty}$. We also explicitly include our upper bound from Theorem 3.1, showing that it tightly describes the behavior to within constants. Lastly, Figure 2(a) shows a clear distinction between our upper bound and simply using Davis–Kahan to upper bound $\|\widehat{V}_1 \widetilde{U} - V_1\|_{2,\infty}$ (viable since $\|\cdot\|_{2,\infty} \le \|\cdot\|_2$)—it is not just a matter of constants; in this case the rate of decay with respect to n is fundamentally different for the two bounds.

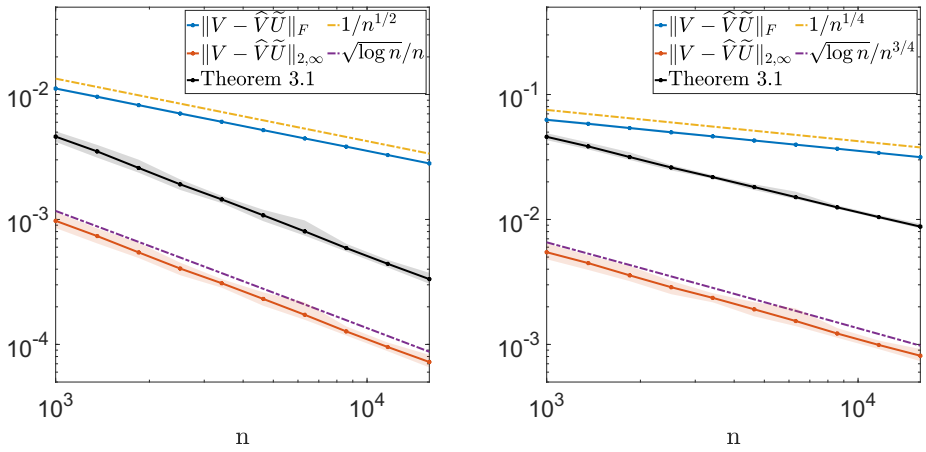
Perhaps more interestingly, we also consider the case where $E = (1/n^{3/4})Z$. In this case, simply applying concentration bounds to Theorem 3.1 predicts $\|\widehat{V}_1 \widetilde{U} - V_1\|_{2,\infty} \lesssim_P \frac{\sqrt{\log n}}{\sqrt{n}}$. However, as expected in this setting, the bound used to control $\|E\widehat{Y}\|_{2,\infty}$ is loose, and Figure 2(b) shows that the error acts as if E and $\widehat{Y}$ were independent (though they are decidedly not) yielding an observed convergence rate of $\frac{\sqrt{\log n}}{n^{3/4}}$. Nevertheless, as before we explicitly compute and plot the upper bound from Theorem 3.1 for comparison. Notably, in this setting the improved bound of Theorem 3.6 is sufficient to correctly capture the observed asymptotic behavior.



(a) $E = (1/n)Z$ (b) $E = (1/n^{3/4})Z$

FIG. 2. *Asymptotic behavior of* $\|\widehat{V}_1 \widetilde{U} - V_1\|_{2,\infty}$ *where* A *is low-rank and* $E = \sigma Z$ *for varying* σ *(in this case* $\mathbb{E}[\|E\|_2] \sim \sigma\sqrt{n}$*), along with the upper bound given by Theorem* 3.1 *for each instance of the problem. For each* n *the experiment was repeated* 30 *times, and we report the mean of the error and computed upper bounds. The shaded regions represent the area between the* 0.05 *and* 0.95 *quantiles of the respective lines. We see that the bounds provided by Theorems* 3.1 *and* 3.6 *are sharp (modulo a multiplicative constant). In the plot on the left, the slope of red line (*$\|\widehat{V}_1 \widetilde{U} - V_1\|_{2,\infty}$*) matches the slope of the black line (bound provided by Theorem* 3.1*). In the plot on the right, the slope of the red line matches the slope of the purple line (rate provided by Theorem* 3.6*). For reference we provide the more slowly converging quantity* $\|\widehat{V}_1 \widetilde{U} - V_1\|_F$*, which is within constants of* $\mathsf{dist}(\widehat{V}_1, V_1)$ *and behaves as predicted by classical theory such as the Davis–Kahan theorem* [10].

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.



(a) $E = (1/n)Z$ (b) $E = (1/n^{3/4})Z$

FIG. 3. *Asymptotic behavior of* $\|\widehat{V}_1\widetilde{U} - V_1\|_{2,\infty}$ *where* A *is not low-rank and* $E = \sigma Z$ *for varying* σ *(in this case* $\mathbb{E}[\|E\|_2] \sim \sigma\sqrt{n}$*), along with the upper bound given by Theorem* 3.1 *for each instance of the problem. For each* n *the experiment was repeated* 30 *times, and we report the mean of the error and computed upper bounds. The shaded regions represent the area between the* 0.05 *and* 0.95 *quantiles of the respective lines. We see that the bounds provided by Theorems* 3.1 *and* 3.6 *are sharp (modulo a multiplicative constant). In the plot on the left, the slope of red line (*$\|\widehat{V}_1\widetilde{U} - V_1\|_{2,\infty}$*) matches the slope of the black line (bound provided by Theorem* 3.1*). In the plot on the right, the slope of the red line matches the slope of the purple line (rate provided by Theorem* 3.6*). For reference we provide the more slowly converging quantity* $\|\widehat{V}_1\widetilde{U} - V_1\|_F$*, which is within constants of* $\mathsf{dist}(\widehat{V}_1, V_1)$ *and behaves as predicted by classical theory such as the Davis–Kahan theorem* [10].

4.2. A is not low-rank. Next, we consider the case where A itself is no longer low-rank. For even n let

$$V_1 = \frac{1}{\sqrt{n}}\begin{bmatrix}\mathbf{1} & \mathbf{1}_{\pm}\end{bmatrix},$$

$v_2 = e_1 - e_2 \in \mathbf{R}^{n\times 1}$, and

$$A = 4V_1V_1^T + v_2v_2^T.$$

Notably, A is no longer low-rank, and the component of A orthogonal to V_1 is coherent, of significant relative magnitude, and does not decay with n. Nevertheless, our results immediately imply that the asymptotic behavior of $\|\widehat{V}_1\widetilde{U} - V_1\|_{2,\infty}$ should match that of the low-rank case.[11] In contrast, this behavior is not accurately predicted by the upper bounds from [6] reproduced in (3.3). Using the same experimental set up as before, Figure 3(a) and 3(b) clearly illustrate the asymptotic behavior we expect—mirroring that of Figure 2(a) and 2(b), respectively. We also compute and plot our upper bound from Theorem 3.1 for reference, observing that it is once again descriptive. Inclusion of the Davis–Kahan bound shows, once again, that there is a clear distinction between $\|\widehat{V}_1\widetilde{U} - V_1\|_{2,\infty}$ and $\|\widehat{V}_1\widetilde{U} - V_1\|_F$.

5. Extensions of our bounds for nonnormal matrices. We have constructed our bounds for symmetric matrices A subject to arbitrary additive symmetric perturbations E. Nevertheless, they may be extended in several directions for nonnormal A and/or E by considering more general invariant subspaces arising in the

[11] Here $\mathsf{sep}_{(2,\infty)}(\Lambda_1, V_2\Lambda_2V_2^T) \geq 2$ as a consequence of Lemma 2.6.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Schur forms of A and $A+E$. We briefly articulate how such extensions are readily obtained following the same proof strategy used for Theorem 3.1.

5.1. Schur form subspaces. Our results can be directly extended to the Schur form for nonnormal matrices; notably, we also no longer require any assumptions on E. We now let

$$A = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} T_{1,1} & T_{1,2} \\ 0 & T_{2,2} \end{bmatrix} \begin{bmatrix} U_1 & U_2 \end{bmatrix}^*, \tag{5.1}$$

where $U_1 \in \mathbb{C}^{n\times r}$ and $U_2 \in \mathbb{C}^{n-r\times r}$ have orthonormal columns and $T_{1,1}$ and $T_{2,2}$ are upper triangular.

Adding one additional assumption about the norm of $T_{1,2}$ our results extend via Theorem 5.1 to the Schur form. For simplicity we have been loose with our assumptions on $\|E\|_2$ and $\|T_{1,2}\|_2$, and small improvements to the necessary constants are possible. Importantly, rather than $T_{1,2}$ showing up in the upper bounds it shows up in the assumptions—thereby controlling the matrices for which this result is valid.

THEOREM 5.1. *Let $A \in \mathbf{R}^{n\times n}$ have the Schur form*

$$A = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} T_{1,1} & T_{1,2} \\ 0 & T_{2,2} \end{bmatrix} \begin{bmatrix} U_1 & U_2 \end{bmatrix}^*$$

following (5.1)*, and let* $\mathsf{gap} = \min\{\mathsf{sep}_2(T_{1,1}, T_{2,2}), \mathsf{sep}_{(2,\infty),U_2}(T_{1,1}, U_2 T_{2,2} U_2^*)\}$*. If* $\|E\|_2 \le \frac{\mathsf{gap}}{10}$ *and* $\|T_{1,2}\|_2 \le \frac{\mathsf{gap}}{10}$*, then there exists a matrix* $\widehat{Y} \in \mathbb{C}^{n-r\times r}$ *such that*

$$\widehat{U}_1 = (U_1 + U_2\widehat{Y})(I + \widehat{Y}^*\widehat{Y})^{-1/2}$$

forms an invariant subspace for $\widehat{A}$ *satisfying*

$$\min\{\|\widehat{U}_1 Q - U_1\|_{2,\infty} : Q \in \mathbf{O}^r\} \le 8\|U_1\|_{2,\infty}\left(\frac{\|E\|_2}{\mathsf{sep}_2(T_{1,1}, T_{2,2})}\right)^2 + 2\frac{\|U_2U_2^*EU_1\|_{2,\infty}}{\mathsf{gap}} + 4\frac{\|U_2U_2^*EU_2U_2^*\|_{2,\infty}\|E_{2,1}\|_2}{\mathsf{gap}\times\mathsf{sep}_2(T_{1,1}, T_{2,2})}.$$

Proof. Following the proof of the main result, in this setting the upper right block of (3.9) is no longer zero but $\widehat{V}_1^T\widehat{A}\widehat{V}_2$. However, our additional assumptions ensure that $\|\widehat{A}_{1,2}\|_2 \le \mathsf{gap}/5$. Therefore, the result follows from the same argument as Lemma 3.4 and Theorem 3.1 where we simply use $\widehat{A}_{1,2}$ in place of $E_{1,2}$ and $T_{1,1}$ and $T_{2,2}$ in lieu of Λ_1 and Λ_2. ∎

Remark 6. While we have formulated Theorem 5.1 for general A and E, in the case where A is normal the upper bound simplifies significantly (independent of any structural assumptions on E). In particular, $T_{1,1}$ and $T_{2,2}$ become diagonal and $T_{1,2} = 0$. This implies that when we consider invariant subspaces of $\widehat{A}$ for symmetric A we can constructed qualitatively similar bounds regardless of whether or not E is symmetric.

5.2. Singular vectors and subspaces. More generally, and perhaps of more interest for nonnormal matrices, analogous questions about subspace perturbations can be posed for singular subspaces. While omitted here, we believe the proof strategy employed in this work can be extended to develop similar bounds for pairs of singular subspaces. This assertion is based on the quadratic forms given in [24] for singular subspaces of $A+E$, though we leave such developments for future work.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

6. Conclusions. Throughout this manuscript we have developed bounds on $\|\widehat{V}_1 - V_1\widetilde{U}\|_{2,\infty}$ that are characterized by easily interpretable quantities (such as $\|V_1\|_{2,\infty}$) and rely on minimal assumptions. By additionally demonstrating that various aspects of our bounds are "essential" when allowing for arbitrary symmetric A and E we clearly show where the limits are for this problem absent additional assumptions. Nevertheless, this effort also provides a natural launching point for further analysis, as it points to the key assumptions that have to (or may) be made to further understand the behavior of $\|\widehat{V}_1 - V_1\widetilde{U}\|_{2,\infty}$ in specific settings. One concrete example of this is the random setting explored in section 4, where more refined control of $\|E\widehat{Y}\|_{2,\infty}$ is possible. Lastly, there are several ways in which our bounds show commonly made assumptions in prior work (such as incoherence of A or certain assumptions on $V_2\Lambda_2 V_2^T$) are unnecessary. The consequence of this is that our bounds are sharper in certain situations. Collectively, we believe that these qualities make our bounds useful and interpretable across a broad range of applications.

Appendix A. Proofs on properties of separation.

A.1. Proof of Lemma 2.2. First, we observe that because sep is shift invariant it suffices to prove the result for nonnegative D_1 and D_2. Therefore, we assume D_1 and D_2 have nonnegative entries for the remainder of this proof. We now prove lower bounds for all three variants of sep.

2-NORM. For any Z we let $U_Z\Sigma_Z V_Z^T$ denote its reduced SVD and note that since $\|Z\|_2 = 1$ we have that $\sigma_1 = 1$. Now we observe that

$$\begin{aligned}\|ZD_1 - D_2Z\|_2 &\geq \|ZD_1\|_2 - \|D_2Z\|_2 \\ &\geq \lambda_{\min}(D_1) - \|D_2\|_2 \\ &\geq \lambda_{\min}(D_1) - \lambda_{\max}(D_2),\end{aligned}$$

where we have used that

$$\begin{aligned}\|XD_1\|_2 &= \|\Sigma_Z V_Z^T D_1\|_2 \\ &\geq \|\Sigma_Z V_Z^T D_1 V_Z e_1\|_2 \\ &\geq \|\sigma_1 e_1^T V_Z^T D_1 V_Z e_1\|_2 \\ &\geq \lambda_{\min}(D_1).\end{aligned}$$

FROBENIUS NORM. For any Z with unit Frobenius norm observe that

$$\begin{aligned}\|ZD_1 - D_2Z\|_F &\geq \|ZD_1\|_F - \|D_2Z\|_F \\ &\geq \lambda_{\min}(D_1) - \|D_2Z\|_2 \\ &\geq \lambda_{\min}(D_1) - \lambda_{\max}(D_2).\end{aligned}$$

$(2,\infty)$-NORM. For any Z with $\|Z\|_{2,\infty} = 1$ there exists an index k such that $\|e_k^T Z\|_2 = 1$ where $e_k \in \mathbf{R}^m$ is a canonical basis vector. Now, observe that

$$\begin{aligned}\|ZD_1 - D_2Z\|_{2,\infty} &\geq \|ZD_1\|_{2,\infty} - \|D_2Z\|_{2,\infty} \\ &\geq \|e_k^T ZD_1\|_2 - \|D_2Z\|_{2,\infty} \\ &\geq \lambda_{\min}(D_1) - \lambda_{\max}(D_2),\end{aligned}$$

where the last inequality follows because D_2 represents a row scaling of Z.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Finally, let j denote the column in which $\lambda_{\min}(D_1)$ arises, and let i denote the column in which $\lambda_{\max}(D_2)$ arises. Now, observe that

$$e_i e_j^T D_1 - D_2 e_i e_j^T = (\lambda_{\min}(D_1) - \lambda_{\max}(D_1))\, e_i e_j^T,$$

where $e_i \in \mathbf{R}^m$ and $e_j \in \mathbf{R}^\ell$ are canonical basis vectors. Since

$$\|e_i e_j^T\|_2 = \|e_i e_j^T\|_F = \|e_i e_j^T\|_{2,\infty} = 1$$

we achieve the aforementioned lower bounds in all cases and thereby conclude the proof.

Appendix B. The Newton–Kantorovich theorem. This is the version of the the Newton-Kantorovich theorem we appeal to. It appears in [15, pages 536].

THEOREM B.1. *Let X, Y be Banach spaces and $F : X \to Y$ be twice-continuously (Fréchet) differentiable in a neighborhood of U of $x \in X$. Assume there is a linear map $J : X \to Y$ such that S_A^{-1} is bounded and satisfies*

1. $\|J^{-1}(F(x))\| \le \eta$,
2. $\|J^{-1} \circ \partial F(x) - I\| \le \delta$,
3. $\|J^{-1} \circ \partial^2 F(y)\| \le K$ *for all* $y \in U$.

If $\delta < 1$ and $h := \frac{\eta K}{(1-\delta)^2} < \frac{1}{2}$, then the sequence (x_t) defined recursively as

$$x_0 \leftarrow x$$
$$x_{t+1} \leftarrow x_t - J^{-1}(F(x_t))$$

converges to $\bar{x} \in X$ such that $F(\bar{x}) = 0$ and

$$\|\bar{x} - x\| \le \frac{2\eta}{(1-\delta)(1+\sqrt{1-2h})}.$$

B.1. Proof of Lemma 3.4. We start by evaluating the derivatives of F:

$$\partial F(0) : X \to X\widehat{A}_{1,1} - \widehat{A}_{2,2}X,$$
$$\partial^2 F(X) : X_1, X_2 \to X_1 \widehat{A}_{1,2} X_2.$$

Recognizing $\|S_{\widehat{A}}^{-1}\|_2 = \frac{1}{\mathsf{sep}_2(\widehat{A}_{1,1}, \widehat{A}_{2,2})}$ and $F(0) = E_{2,1}$, we have

$$\|S_{\widehat{A}}^{-1}(F(0))\|_2 \le \frac{\|E_{2,1}\|_2}{\mathsf{sep}_2(\widehat{A}_{1,1}, \widehat{A}_{2,2})},$$

so the first condition of Theorem B.1 is satisfied by $\eta = \frac{\|E_{2,1}\|_2}{\mathsf{sep}_2(\widehat{A}_{1,1}, \widehat{A}_{2,2})}$. We recognize $S_{\widehat{A}} = \partial F(0)$, so the second condition of Theorem B.1 is satisfied by $\delta = 0$. Finally, we have

$$\|S_{\widehat{A}}^{-1}(\partial^2 F(0)(X_1, X_2))\|_2 \le \frac{\|X_1 \widehat{A}_{1,2} X_2\|_2}{\mathsf{sep}_2(\widehat{A}_{1,1}, \widehat{A}_{2,2})} \le \frac{\|X_1\|_2 \|E_{1,2}\|_2 \|X_2\|_2}{\mathsf{sep}_2(\widehat{A}_{1,1}, \widehat{A}_{2,2})},$$

so the third condition of Theorem B.1 is satisfied by $K = \frac{\|E_{1,2}\|_2}{\mathsf{sep}_2(\widehat{A}_{1,1}, \widehat{A}_{2,2})}$. From Proposition 2.1 of [16] we then have that

$$\begin{aligned}\mathsf{sep}_2(\widehat{A}_{1,1}, \widehat{A}_{2,2}) &\ge \mathsf{sep}_2(\Lambda_1 + E_{1,1}, \Lambda_2 + E_{2,2}) \\ &= \mathsf{sep}_2(\Lambda_1, \Lambda_2) - \|E_{1,1}\|_2 - \|E_{2,2}\|_2 \\ &\ge \mathsf{sep}_2(\Lambda_1, \Lambda_2) - 2\|E\|_2.\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

We combine this bound on $\mathsf{sep}_2(\widehat{A}_{1,1}, \widehat{A}_{2,2})$ with the condition $\|E\|_2 \leq \frac{\mathsf{sep}_2(\Lambda_1,\Lambda_2)}{4}$ to obtain

$$\eta = \frac{\|E_{2,1}\|_2}{\mathsf{sep}_2(\widehat{A}_{1,1}, \widehat{A}_{2,2})} \leq \frac{\|E_{2,1}\|_2}{\mathsf{sep}_2(\Lambda_1,\Lambda_2) - 2\|E\|_2} \leq \frac{2\|E_{2,1}\|_2}{\mathsf{sep}_2(\Lambda_1,\Lambda_2)},$$
$$h = \frac{\eta K}{(1-\delta)^2} = \frac{\|E_{2,1}\|_2\|E_{1,2}\|_2}{\mathsf{sep}_2(\widehat{A}_{1,1}, \widehat{A}_{2,2})^2} \leq \frac{\|E\|_2^2}{(\mathsf{sep}_2(\Lambda_1,\Lambda_2) - 2\|E\|_2)^2} \leq \frac{1}{4} < \frac{1}{2},$$

so the Newton–Kantorovich theorem implies F has a root $\widehat{X}$ such that

$$\|\widehat{X}\|_2 \leq \frac{2\eta}{(1-\delta)(1+\sqrt{1-2h})} < 2\eta \leq \frac{4\|E_{2,1}\|_2}{\mathsf{sep}_2(\Lambda_1,\Lambda_2)}$$

as claimed.

Appendix C. Multiple perturbations. Here, we characterize how changes in the invariant subspaces may be controlled when we have multiple perturbations of the matrix A. Our primary use of this result is in developing Theorem 3.6. Specifically, when E is drawn from a random model satisfying certain independence assumptions we can use multiple draws of E to develop tighter bounds on the changes in V_1 generated by any single instance of E.

LEMMA C.1. *Let $A \in \mathbf{R}^{n\times n}$ be symmetric with eigendecomposition*

$$A = V_1\Lambda_1 V_1^T + V_2\Lambda_2 V_2^T$$

following our conventions in (2.1) *and*

$$\mathsf{gap} = \min\{\mathsf{sep}_2(\Lambda_1,\Lambda_2), \mathsf{sep}_{(2,\infty),V_2}(\Lambda_1, V_2\Lambda_2 V_2^T)\}.$$

Consider two symmetric perturbations E and $\widetilde{E}$ satisfying $E_{1,2} = \widetilde{E}_{1,2}$, $E_{1,1} = \widetilde{E}_{1,1}$, $\|E\|_2 \leq \frac{\mathsf{gap}}{5}$, and $\|\widetilde{E}\|_2 \leq \frac{\mathsf{gap}}{5}$. In this setting

$$\begin{aligned}\|\widehat{V}_1\widetilde{U} - V_1\|_{2,\infty} \leq 8\|V_1\|_{2,\infty}&\left(\frac{\|E_{2,1}\|_2}{\mathsf{sep}_2(\Lambda_1,\Lambda_2)}\right)^2 \\ &+ 2\frac{\|V_2E_{2,1}\|_{2,\infty}}{\mathsf{gap}} + 4\frac{\|V_2E_{2,2}V_2^T\widetilde{Y}\|_{2,\infty}}{\mathsf{gap}} \\ &+ 5\frac{(\|E_{2,2}\|_2 + \|\widetilde{E}_{2,2}\|_2)\|V_2E_{2,2}V_2^T\|_{2,\infty}\|E_{2,1}\|_2}{\mathsf{gap}\times\mathsf{sep}_2(\Lambda_1,\Lambda_2)^2} \\ &+ 10\frac{\|V_2E_{2,2}V_2^T\|_{2,\infty}\|E_{2,1}\|_2^3}{\mathsf{gap}\times\mathsf{sep}_2(\Lambda_1,\Lambda_2)^3},\end{aligned}$$

where $\widehat{V}_1$ is any matrix with orthonormal columns whose range is the dominant r-dimensional invariant subspace of $A+E$ and $\widetilde{U}$ solves the orthogonal Procrustes problem

$$\min\{\|\widehat{V}_1U - V_1\|_F : U \in \mathbf{O}^r\}.$$

In addition, $\widetilde{Y} \in \mathbf{R}^{n\times r}$ is in the range of V_2, is a root of

$$G : Y \to -V_2\widetilde{A}_{2,1} + Y\widetilde{A}_{1,1} - V_2\widetilde{A}_{2,2}V_2^TY + Y\widetilde{A}_{1,2}V_2^TY$$

where $\widetilde{A} = A + \widetilde{E}$, and satisfies $\|\widetilde{Y}\|_2 \leq \frac{4\|E_{2,1}\|_2}{\mathsf{sep}_2(\Lambda_1,\Lambda_2)}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Proof. First, the assumption $\|E\|_2 \leq \frac{\mathsf{gap}}{5}$ allows us to invoke Lemma 3.5 and deduce the first two terms in the bound directly. To construct the last three terms we bound

$$4\frac{\|V_2 E_{2,2} V_2^T \widehat{Y}\|_{2,\infty}}{\mathsf{gap}}$$

in a more nuanced manner than the naïve submultiplicative bound used in section 3.

Letting $\widetilde{A} = A + \widetilde{E}$ we may invoke Lemma 3.4 to assert that there exists $\widetilde{X} \in \mathbf{R}^{(n-r)\times r}$ such that

$$\widetilde{V}_1 = \left(V_1 + V_2\widetilde{X}\right)\left(I_r + \widetilde{X}^T\widetilde{X}\right)^{-\frac{1}{2}}$$

is an orthonormal basis for the dominant r-dimensional invariant subspace of $\widetilde{A}$ and $\|\widetilde{X}\|_2 \leq \frac{4\|E_{2,1}\|_2}{\mathsf{sep}_2(\Lambda_1,\Lambda_2)}$. As before, $\widetilde{X}$ is a root of

$$F : X \to -\widetilde{A}_{2,1} + X\widetilde{A}_{1,1} - \widetilde{A}_{2,2}X + X\widetilde{A}_{1,2}X,$$

and therefore $\widetilde{Y} = V_2\widetilde{X}$ is a root of

$$G : Y \to -V_2\widetilde{A}_{2,1} + Y\widetilde{A}_{1,1} - V_2\widetilde{A}_{2,2}V_2^T Y + Y\widetilde{A}_{1,2}V_2^T Y.$$

We now proceed by observing that

$$\begin{aligned}4\frac{\|V_2 E_{2,2} V_2^T \widehat{Y}\|_{2,\infty}}{\mathsf{gap}} &\leq 4\frac{\|V_2 E_{2,2} V_2^T \widetilde{Y}\|_{2,\infty}}{\mathsf{gap}} + 4\frac{\|V_2 E_{2,2} V_2^T (\widehat{Y} - \widetilde{Y})\|_{2,\infty}}{\mathsf{gap}} \\ &\leq 4\frac{\|V_2 E_{2,2} V_2^T \widetilde{Y}\|_{2,\infty}}{\mathsf{gap}} + 4\frac{\|V_2 E_{2,2} V_2^T\|_{2,\infty}\|\widehat{Y} - \widetilde{Y}\|_2}{\mathsf{gap}}.\end{aligned}$$

Control over $\|\widehat{Y} - \widetilde{Y}\|_2$ is achieved by observing that $\widetilde{Y}$ and $\widehat{Y}$ satisfy

$$0 = -V_2\widehat{A}_{2,1} + \widehat{Y}\widehat{A}_{1,1} - V_2\widehat{A}_{2,2}V_2^T\widehat{Y} + \widehat{Y}\widehat{A}_{1,2}V_2^T\widehat{Y}$$

and

$$0 = -V_2\widetilde{A}_{2,1} + \widetilde{Y}\widetilde{A}_{1,1} - V_2\widetilde{A}_{2,2}V_2^T\widetilde{Y} + \widetilde{Y}\widetilde{A}_{1,2}V_2^T\widetilde{Y}.$$

In particular, since $\widetilde{A}_{i,j} = \widehat{A}_{i,j}$ for all i,j pairs except $i = j = 2$, subtracting the equations yields

$$0 = (\widehat{Y} - \widetilde{Y})\widehat{A}_{1,1} - V_2\widehat{A}_{2,2}V_2^T\widehat{Y} + V_2\widetilde{A}_{2,2}V_2^T\widetilde{Y} + \widehat{Y}\widehat{A}_{1,2}V_2^T\widehat{Y} - \widetilde{Y}\widetilde{A}_{1,2}V_2^T\widetilde{Y}$$

and via further rearrangement

$$\begin{aligned}(\widehat{Y} - \widetilde{Y})\widehat{A}_{1,1} - V_2 A_{2,2}V_2^T(\widehat{Y} - \widetilde{Y}) = V_2 E_{2,2}V_2^T\widehat{Y} - V_2\widetilde{E}_{2,2}V_2^T\widetilde{Y} \\ - \widehat{Y}\widehat{A}_{1,2}V_2^T\widehat{Y} + \widetilde{Y}\widetilde{A}_{1,2}V_2^T\widetilde{Y}.\end{aligned}$$

Taking norms we observe that the left-hand side is at least

$$\frac{4}{5}\mathsf{sep}_2(\Lambda_1,\Lambda_2)\|(\widehat{Y} - \widetilde{Y})\|_2 \leq \|(\widehat{Y} - \widetilde{Y})\widehat{A}_{1,1} - V_2 A_{2,2}V_2^T(\widehat{Y} - \widetilde{Y})\|_2, \tag{C.1}$$

where we have used that $\|E_{1,1}\|_2 \leq \mathsf{sep}_2(\Lambda_1,\Lambda_2)/5$. We may bound the right-hand side via the triangle inequality and repeated used of Lemma 3.4 as

$$\begin{aligned}&\|V_2 E_{2,2}V_2^T\widehat{Y} - V_2\widetilde{E}_{2,2}V_2^T\widetilde{Y} - \widehat{Y}\widehat{A}_{1,2}V_2^T\widehat{Y} + \widetilde{Y}\widetilde{A}_{1,2}V_2^T\widetilde{Y}\|_2 \\ &\quad\leq 4\frac{(\|E_{2,2}\|_2 + \|\widetilde{E}_{2,2}\|_2)\|E_{2,1}\|}{\mathsf{sep}_2(\Lambda_1,\Lambda_2)} + 8\frac{\|E_{2,1}\|_2^3}{\mathsf{sep}_2(\Lambda_1,\Lambda_2)^2},\end{aligned} \tag{C.2}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

where we have also used that $A_{1,2} = 0$. Combining (C.1) and C.2 yields the desired upper bound. ∎

Acknowledgments. We would like to thank the anonymous referees for their many helpful suggestions that improved this manuscript.

REFERENCES

[1] E. Abbe, *Community detection and stochastic block models: Recent developments*, J. Mach. Learn. Res., 18 (2017), pp. 6446–6531.

[2] E. Abbe, J. Fan, K. Wang, and Y. Zhong, *Entrywise Eigenvector Analysis of Random Matrices with Low Expected Rank*, e-print, arXiv:1709.09565 [math. ST], 2017, https://arxiv.org/abs/1709.09565.

[3] A. Athreya, D. E. Fishkind, M. Tang, C. E. Priebe, Y. Park, J. T. Vogelstein, K. Levin, V. Lyzinski, and Y. Qin, *Statistical inference on random dot product graphs: A survey*, J. Mach. Learn. Res., 18 (2017), pp. 8393–8484.

[4] T. Tony Cai, Z. Ma, and Y. Wu, *Sparse PCA: Optimal rates and adaptive estimation*, Ann. Statist., 41 (2013), pp. 3074–3110.

[5] E. J. Candès and B. Recht, *Exact matrix completion via convex optimization*, Found. Comput. math., 9 (2009), p. 717.

[6] J. Cape, M. Tang, C. E. Priebe, et al., *The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics*, Ann. Statist., 47 (2019), pp. 2405–2439.

[7] A. Damle, L. Lin, and L. Ying, *Compressed representation of Kohn-Sham orbitals via selected columns of the density matrix*, J. Chem. Theory Comput., 11 (2015), pp. 1463–1469.

[8] A. Damle, L. Lin, and L. Ying, *Computing localized representations of the Kohn–Sham subspace via randomization and refinement*, SIAM J. Sci. Comput., 39 (2017), pp. B1178–B1198.

[9] A. Damle, V. Minden, and L. Ying, *Simple, direct and efficient multi-way spectral clustering*, Informa. Inference, 8 (2018), pp. 181–203.

[10] C. Davis and W. Kahan, *The rotation of eigenvectors by a perturbation.* III, SIAM J. Numer. Anal., 7 (1970), pp. 1–46.

[11] J. Eldridge, M. Belkin, and Y. Wang, *Unperturbed: Spectral analysis beyond Davis-Kahan*, in Proceedings of Machine Learning Research, 2018, pp. 321–358.

[12] J. Fan, K. Wang, Y. Zhong, and Z. Zhu, *Robust High Dimensional Factor Models with Applications to Statistical Machine Learning*, e-print, arXiv:1808.03889 [Stat. ME], 2018, https://arxiv.org/abs/1808.03889.

[13] J. Fan, W. Wang, and Y. Zhong, *An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation*, J. Mach. Learn. Res., 18 (2018), pp. 1–42.

[14] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, MD, 1996.

[15] L. V. Kantorovich and G. P. Akilov, 2nd ed., *Functional Analysis*, Pergamon, Press, Elmsford, NY, 1982.

[16] M. Karow and D. Kressner, *On a perturbation bound for invariant subspaces of matrices*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 599–618.

[17] R. H. Keshavan, A. Montanari, and S. Oh, *Matrix completion from noisy entries*, J. Mach. Learn. Res., 11 (2010), pp. 2057–2078.

[18] C. M. Le, E. Levina, and R. Vershynin, *Concentration of Random Graphs and Application to Community Detection*, e-print, arXiv:1801.08724 [math. ST], 2018, https://arxiv.org/abs/1801.08724.

[19] J. Lei and A. Rinaldo, *Consistency of spectral clustering in stochastic block models*, Ann. Statist., 43 (2015), pp. 215–237.

[20] B. Nadler, *Finite sample approximation results for principal component analysis: A matrix perturbation approach*, Ann. Statist., 36 (2008), pp. 2791–2817.

[21] A. Perry, A. S. Wein, A. S. Bandeira, and A. Moitra, *Optimality and Sub-Optimality of PCA for Spiked Random Matrices and Synchronization*, e-print, arXiv:1609.05573 [math. ST], 2016, https://arxiv.org/abs/1609.05573.

[22] K. Rohe, S. Chatterjee, and B. Yu, *Spectral clustering and the high-dimensional stochastic blockmodel*, Ann. Statist., 39 (2011), pp. 1878–1915.

[23] A. Singer, *Angular synchronization by eigenvectors and semidefinite programming*, Appl. comput. harmon. anal., 30 (2011), pp. 20–36.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[24] G. W. Stewart, *Error and perturbation bounds for subspaces associated with certain eigenvalue problems*, SIAM Rev., 15 (1973), pp. 727–764.
[25] G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*, Academic Press, New York, 1990.
[26] R. Vershynin, *High-Dimensional Probability*, Cambridge University Press, Cambridge, UK, 2018.
[27] U. Von Luxburg, *A tutorial on spectral clustering*, Stat. comput., 17 (2007), pp. 395–416.
[28] U. von Luxburg, M. Belkin, and O. Bousquet, *Consistency of spectral clustering*, Ann. Statist., 36 (2008), pp. 555–586.
[29] P. Wedin, *Perturbation bounds in connection with singular value decomposition*, BIT, 12 (1972), pp. 99–111.

Downloaded 11/26/21 to 132.174.252.179 Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.