LOW-RANK MATRIX APPROXIMATIONS DO NOT NEED A SINGULAR VALUE GAP[*]

PETROS DRINEAS[†] AND ILSE C. F. IPSEN[‡]

Abstract. Low-rank approximations to a real matrix $\mathbf{A}$ can be computed from $\mathbf{ZZ}^T\mathbf{A}$, where $\mathbf{Z}$ is a matrix with orthonormal columns, and the accuracy of the approximation can be estimated from some norm of $\mathbf{A} - \mathbf{ZZ}^T\mathbf{A}$. We show that computing $\mathbf{A} - \mathbf{ZZ}^T\mathbf{A}$ in the two-norm, Frobenius norms, and more generally any Schatten p-norm is a well-posed mathematical problem; and, in contrast to dominant subspace computations, it does not require a singular value gap. We also show that this problem is well-conditioned (insensitive) to additive perturbations in $\mathbf{A}$ and $\mathbf{Z}$, and to dimension-changing or multiplicative perturbations in $\mathbf{A}$—regardless of the accuracy of the approximation. For the special case when $\mathbf{A}$ does indeed have a singular values gap, connections are established between low-rank approximations and subspace angles.

Key words. singular value decomposition, principal angles, additive perturbations, multiplicative perturbations

AMS subject classifications. 15A12, 15A18, 15A42, 65F15, 65F35

DOI. 10.1137/18M1163658

1. Introduction. An emerging problem in theoretical computer science and data science is the low-rank approximation $\mathbf{ZZ}^T\mathbf{A}$ of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ by means of an orthonormal basis $\mathbf{Z} \in \mathbb{R}^{m \times k}$ [9, 28].

The ideal low-rank approximations in the two most popular Schatten p-norms, the two (operator) norm $\|\cdot\|_2$ and the Frobenius norm $\|\cdot\|_F$, consist of left singular vectors $\mathbf{U}_k$ associated with the k dominant singular values $\sigma_1(\mathbf{A}) \geq \cdots \geq \sigma_k(\mathbf{A})$ of $\mathbf{A}$. The approximation errors are minimal and depend on subdominant singular values,

$$\|(\mathbf{I} - \mathbf{U}_k\mathbf{U}_k^T)\mathbf{A}\|_2 = \max_{j \geq k+1} \sigma_j(\mathbf{A}), \qquad \|(\mathbf{I} - \mathbf{U}_k\mathbf{U}_k^T)\mathbf{A}\|_F = \sqrt{\sum_{j \geq k+1} \sigma_j(\mathbf{A})^2}.$$

A popular approach is to compute $\mathbf{Z}$ as an orthonormal basis for a dominant subspace of $\mathbf{A}$, via subspace iteration or Krylov space methods [12, 19].

However, the computation of dominant subspaces range($\mathbf{U}_k$), an important problem in numerical linear algebra [21, 22], is well posed only if the associated singular values are separated from the subdominant singular values by a gap $\sigma_k(\mathbf{A}) - \sigma_{k+1}(\mathbf{A}) > 0$ [18, 23, 25, 26, 27, 29], which exploit perturbation results for invariant subspaces of Hermitian matrices [4, 5]. It is not enough, though, for the gap to exist. It must also be sufficiently large to guarantee that range($\mathbf{U}_k$) is robust (well-conditioned) to tiny perturbations in $\mathbf{A}$, such as roundoff errors. Thus, effort has been put into deriving bounds that not require the existence of the singular value gap $\sigma_k(\mathbf{A}) - \sigma_{k+1}(\mathbf{A}) > 0$.

[†]Department of Computer Science, Purdue University, West Lafayette, IN 47907 (pdrineas@purdue.edu).

[‡]Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205 (ipsen@ncsu.edu).

Contribution. The purpose of our paper, following up on [6], is to establish a clear distinction between the mathematical problems of low-rank approximation on the one hand, and approximation of dominant subspaces on the other. We show that the approximation problem $\||(\mathbf{I} - \mathbf{ZZ}^T)\mathbf{A}\||_p$ is well posed in any Schatten p-norm, and furthermore well-conditioned under perturbations in $\mathbf{A}$ and $\mathbf{Z}$. To the best of our knowledge, these findings (summarized in section 1.4) are novel, of unparalleled clarity, and fully general. Specifically, they

1. make no demands on the accuracy of the approximation $\mathbf{ZZ}^T\mathbf{A}$,
2. hold in all Schatten p-norms,
3. apply to large classes of perturbations: Additive rank-preserving perturbations in the basis $\mathbf{Z}$; and additive, multiplicative, and even dimension-changing perturbations in $\mathbf{A}$.

However, if one so chooses to compute $\mathbf{Z}$ from a dominant subspace of $\mathbf{A}$, one better be assured of the existence of a sufficiently large singular value gap, for otherwise this is a numerically unstable algorithm.

Overview. After reviewing the singular value decomposition (section 1.1), Schatten p-norms (section 1.2), and angles between subspaces (section 1.3), we highlight the main results (section 1.4) and discuss their relevance. This is followed by proofs for low-rank approximations (section 2), and relations to subspace angles (section 3, Appendix A).

1.1. Singular value decomposition (SVD). Let the non-zero matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ have a full singular value decomposition (SVD) $\mathbf{A} = \mathbf{U \Sigma V}^T$, where $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V} \in \mathbb{R}^{n \times n}$ are orthogonal matrices[1], i.e.,

$$\mathbf{UU}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}_m, \qquad \mathbf{VV}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}_n,$$

and $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ is a diagonal matrix with diagonal elements

$$(1.1) \qquad \|\mathbf{A}\|_2 = \sigma_1(\mathbf{A}) \geq \cdots \geq \sigma_r(\mathbf{A}) \geq 0, \qquad r \equiv \min\{m, n\}.$$

For $1 \leq k \leq \operatorname{rank}(\mathbf{A})$, the respective leading k columns of $\mathbf{U}$ and $\mathbf{V}$ are $\mathbf{U}_k \in \mathbb{R}^{m \times k}$ and $\mathbf{V}_k \in \mathbb{R}^{m \times k}$. They are orthonormal, $\mathbf{U}_k^T\mathbf{U}_k = \mathbf{I}_k = \mathbf{V}_k^T\mathbf{V}_k$, and are associated with the k dominant singular values

$$\mathbf{\Sigma}_k \equiv \operatorname{diag}\begin{pmatrix} \sigma_1(\mathbf{A}) & \cdots & \sigma_k(\mathbf{A}) \end{pmatrix} \in \mathbb{R}^{k \times k}.$$

Then

$$(1.2) \qquad \mathbf{A}_k \equiv \mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^T = \mathbf{U}_k\mathbf{U}_k^T\mathbf{A}$$

is a best rank-k approximation of $\mathbf{A}$ in the two norm, and in the Frobenius norm,

$$\|(\mathbf{I} - \mathbf{U}_k\mathbf{U}_k^T)\mathbf{A}\|_{2,F} = \|\mathbf{A} - \mathbf{A}_k\|_{2,F} = \min_{\operatorname{rank}(\mathbf{B})=k} \|\mathbf{A} - \mathbf{B}\|_{2,F}.$$

Projectors. We construct orthogonal projectors to capture the target space, i.e., a dominant subspace of $\mathbf{A}$.

DEFINITION 1.1. *A matrix* $\mathbf{P} \in \mathbb{R}^{m \times m}$ *is an* orthogonal projector, *if it is idempotent and symmetric:*

$$(1.3) \qquad \mathbf{P}^2 = \mathbf{P} = \mathbf{P}^T.$$

---

[1]The superscript T denotes the transpose.

Examples.
- If $\mathbf{Z} \in \mathbb{R}^{m \times k}$ has orthonormal columns with $\mathbf{Z}^T\mathbf{Z} = \mathbf{I}_k$, then $\mathbf{Z}\mathbf{Z}^T$ is an orthogonal projector.
- For $1 \le k \le \mathrm{rank}(\mathbf{A})$, the matrix $\mathbf{U}_k\mathbf{U}_k^T = \mathbf{A}_k\mathbf{A}_k^\dagger$ is the orthogonal projector onto the k-dimensional dominant subspace $\mathrm{range}(\mathbf{U}_k) = \mathrm{range}(\mathbf{A}_k)$. Here the pseudo inverse is $\mathbf{A}_k^\dagger = \mathbf{V}_k\mathbf{\Sigma}_k^{-1}\mathbf{U}_k^T$.

1.2. Schatten p-norms. These are norms defined on the singular values of real and complex matrices and thus special cases of symmetric gauge functions. We briefly review their properties, based on [3, Chapter IV] and [16, sections 3.4–3.5].

DEFINITION 1.2. *For integers $p \ge 1$, the* Schatten p-norms *on $\mathbb{R}^{m \times n}$ are*

$$\||\mathbf{A}\||_p \equiv \sqrt[p]{\sigma_1(\mathbf{A})^p + \cdots + \sigma_r(\mathbf{A})^p}, \qquad r \equiv \min\{m, n\}.$$

Popular Schatten p-norms.

$p = 1$: Nuclear (trace) norm $\|\mathbf{A}\|_* = \sum_{j=1}^r \sigma_j(\mathbf{A}) = \||\mathbf{A}\||_1$.

$p = 2$: Frobenius norm $\|\mathbf{A}\|_F = \sqrt{\sum_{j=1}^r \sigma_j(\mathbf{A})^2} = \||\mathbf{A}\||_2$.

$p = \infty$: Two (operator) norm $\|\mathbf{A}\|_2 = \sigma_1(\mathbf{A}) = \||\mathbf{A}\||_\infty$.

We will make ample use of the following properties.

LEMMA 1.3. *Let $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times \ell}$, and $\mathbf{C} \in \mathbb{R}^{s \times m}$.*
- *Unitary invariance:*
  If $\mathbf{Q}_1 \in \mathbb{R}^{s \times m}$ with $\mathbf{Q}_1^T\mathbf{Q}_1 = \mathbf{I}_m$ and $\mathbf{Q}_2 \in \mathbb{R}^{\ell \times n}$ with $\mathbf{Q}_2^T\mathbf{Q}_2 = \mathbf{I}_n$, then

$$\||\mathbf{Q}_1\mathbf{A}\mathbf{Q}_2^T\||_p = \||\mathbf{A}\||_p.$$

- *Submultiplicativity:* $\||\mathbf{A}\mathbf{B}\||_p \le \||\mathbf{A}\||_p \||\mathbf{B}\||_p$.
- *Strong submultiplicativity (symmetric norm):*

$$\||\mathbf{C}\mathbf{A}\mathbf{B}\||_p \le \sigma_1(\mathbf{C})\,\sigma_1(\mathbf{B})\,\||\mathbf{A}\||_p = \|\mathbf{C}\|_2\,\|\mathbf{B}\|_2\,\||\mathbf{A}\||_p.$$

- *Best rank-k approximation:*

$$\||(\mathbf{I} - \mathbf{U}_k\mathbf{U}_k^T)\mathbf{A}\||_p = \||\mathbf{A} - \mathbf{A}_k\||_p = \min_{\mathrm{rank}(\mathbf{B})=k} \||\mathbf{A} - \mathbf{B}\||_p.$$

1.3. Principal angles between subspaces. We review the definition of angles between subspaces, and the connections between angles and projectors.

DEFINITION 1.4 (section 6.4.3 in [11] and section 2 in [29]). *Let $\mathbf{Z} \in \mathbb{R}^{m \times k}$ and $\widehat{\mathbf{Z}} \in \mathbb{R}^{m \times \ell}$ with $\ell \ge k$ have orthonormal columns so that $\mathbf{Z}^T\mathbf{Z} = \mathbf{I}_k$ and $\widehat{\mathbf{Z}}^T\widehat{\mathbf{Z}} = \mathbf{I}_\ell$. Let the singular values of $\mathbf{Z}^T\widehat{\mathbf{Z}}$ be the diagonal elements of the $k \times k$ diagonal matrix*

$$\cos\mathbf{\Theta}(\mathbf{Z}, \widehat{\mathbf{Z}}) \equiv \mathrm{diag}\begin{pmatrix}\cos\theta_1 & \cdots & \cos\theta_k\end{pmatrix}.$$

Then θ_j, $1 \le j \le k$, are defined as the principal (canonical) angles θ_j between $\mathrm{range}(\mathbf{Z})$ and $\mathrm{range}(\widehat{\mathbf{Z}})$.

To extract such principal angles between subspaces of possibly different dimensions, we make use of projectors.

LEMMA 1.5. *Let $\mathbf{P} \equiv \mathbf{Z}\mathbf{Z}^T$ and $\widehat{\mathbf{P}} \equiv \widehat{\mathbf{Z}}\widehat{\mathbf{Z}}^T$ be orthogonal projectors, where $\mathbf{Z} \in \mathbb{R}^{m \times k}$ and $\widehat{\mathbf{Z}} \in \mathbb{R}^{m \times \ell}$ with $\ell \ge k$ have orthonormal columns. With θ_j being the k principal angles between $\mathrm{range}(\mathbf{Z})$ and $\mathrm{range}(\widehat{\mathbf{Z}})$, define*

$$\sin\mathbf{\Theta}(\mathbf{P}, \widehat{\mathbf{P}}) = \sin\mathbf{\Theta}(\mathbf{Z}, \widehat{\mathbf{Z}}) \equiv \mathrm{diag}\begin{pmatrix}\sin\theta_1 & \cdots & \sin\theta_k\end{pmatrix}.$$

1. *If* $\text{rank}(\widehat{\mathbf{Z}}) = k = \text{rank}(\mathbf{Z})$, *then*

$$\||\sin \boldsymbol{\Theta}(\mathbf{Z}, \widehat{\mathbf{Z}})\||_p = \||(\mathbf{I} - \mathbf{P})\widehat{\mathbf{P}}\||_p = \||(\mathbf{I} - \widehat{\mathbf{P}})\mathbf{P}\||_p \,.$$

   In particular

$$\|(\mathbf{I} - \mathbf{P})\widehat{\mathbf{P}}\|_2 = \|\mathbf{P} - \widehat{\mathbf{P}}\|_2$$

   represents the distance between the subspaces $\text{range}(\mathbf{P})$ *and* $\text{range}(\widehat{\mathbf{P}})$.

2. *If* $\text{rank}(\widehat{\mathbf{Z}}) > k = \text{rank}(\mathbf{Z})$, *then*

$$\||\sin \boldsymbol{\Theta}(\mathbf{Z}, \widehat{\mathbf{Z}})\||_p = \||(\mathbf{I} - \widehat{\mathbf{P}})\mathbf{P}\||_p \ \leq \ \||(\mathbf{I} - \mathbf{P})\widehat{\mathbf{P}}\||_p \,.$$

Proof. The two-norm expressions follow from [11, section 2.5.3] and [27, section 2]. The Schatten p-norm expressions follow from the CS decomposition in [20, Theorem 8.1], [29, section 2], and Appendix A. □

1.4. Highlights of the main results. We present a brief overview of the main results: The well-conditioning of low-rank approximations under additive perturbations in $\mathbf{A}$ and the projector basis $\mathbf{Z}$ (section 1.4.1); the well-conditioning of low-rank approximations under perturbations in $\mathbf{A}$ that change the column dimension (section 1.4.2); and the connection between low-rank approximation errors and angles between subspaces (section 1.4.3).

Thus low-rank approximations of the form $\||\mathbf{A} - \mathbf{Z}\mathbf{Z}^T\mathbf{A}\||_p$ are well posed, well conditioned, and do not need a singular value gap.

1.4.1. Additive perturbations in the projector basis and the matrix. We show that the low-rank approximation error is insensitive to additive rank-preserving perturbations in the projector basis (Theorem 1 and Corollary 1), and to additive perturbations in the matrix (Theorem 2 and Corollary 2).

THEOREM 1 (additive rank-preserving perturbations in the projector basis). *Let* $\mathbf{A} \in \mathbb{R}^{m \times n}$, *and let* $\mathbf{Z} \in \mathbb{R}^{m \times \ell}$ *be a projector basis with orthonormal columns so that* $\mathbf{Z}^T\mathbf{Z} = \mathbf{I}_\ell$. *Denote by* $\widehat{\mathbf{Z}} \in \mathbb{R}^{m \times \ell}$ *a perturbation, and define*

$$\epsilon_Z \equiv \ \|\widehat{\mathbf{Z}}^\dagger\|_2 \, \|\mathbf{Z} - \widehat{\mathbf{Z}}\|_2.$$

1. *If* $\text{rank}(\widehat{\mathbf{Z}}) = \text{rank}(\mathbf{Z})$, *then*

$$\||(\mathbf{I} - \mathbf{Z}\mathbf{Z}^T)\mathbf{A}\||_p - \epsilon_Z \, \||\mathbf{A}\||_p \leq \||(\mathbf{I} - \widehat{\mathbf{Z}}\widehat{\mathbf{Z}}^\dagger)\mathbf{A}\||_p$$
$$\leq \||(\mathbf{I} - \mathbf{Z}\mathbf{Z}^T)\mathbf{A}\||_p + \epsilon_Z \, \||\mathbf{A}\||_p \,.$$

2. *If* $\|\mathbf{Z} - \widehat{\mathbf{Z}}\|_2 \leq 1/2$, *then* $\text{rank}(\widehat{\mathbf{Z}}) = \text{rank}(\mathbf{Z})$ *and* $\epsilon_Z \leq 2\,\|\mathbf{Z} - \widehat{\mathbf{Z}}\|_2$.

Proof. See section 2, and in particular Theorem 2.2. □

REMARK 1. *Theorem 1 shows what happens to the approximation error when the basis changes from* $\mathbf{Z}$ *to* $\widehat{\mathbf{Z}}$. *The absolute change in the approximation error is small if* $\||\mathbf{A}\||_p$ *is small, and if* $\widehat{\mathbf{Z}}$ *is close to* $\mathbf{Z}$ *and is well conditioned with respect to (left) inversion.*

1. *The error for the perturbed basis* $\widehat{\mathbf{Z}}$ *is bounded in terms of* ϵ_Z *amplified by the norm of* $\mathbf{A}$.

   The additive two-norm expression ϵ_Z *represents both, an absolute and a relative perturbation, as*

$$\epsilon_Z \equiv \ \|\widehat{\mathbf{Z}}^\dagger\|_2 \, \|\mathbf{Z} - \widehat{\mathbf{Z}}\|_2 \ = \ \underbrace{\|\widehat{\mathbf{Z}}\|_2 \|\widehat{\mathbf{Z}}^\dagger\|_2}_{\substack{\text{Deviation from} \\ \text{orthonormality}}} \ \underbrace{\frac{\|\widehat{\mathbf{Z}} - \mathbf{Z}\|_2}{\|\widehat{\mathbf{Z}}\|_2}}_{\substack{\text{Relative distance} \\ \text{from exact basis}}}.$$

The first factor is the two-norm condition number $\|\widehat{\mathbf{Z}}\|_2\|\widehat{\mathbf{Z}}^\dagger\|_2$ of the perturbed basis with regard to (left) inversion. The second factor is the relative two-norm distance between the bases.

The assumption forces the perturbed vectors $\widehat{\mathbf{Z}}$ to be linearly independent, but not necessarily orthonormal. Hence the Moore–Penrose inverse replaces the transpose in the orthogonal projector, and the condition number represents the deviation of $\widehat{\mathbf{Z}}$ from orthonormality.

2. *For simplicity, we consider $\|\mathbf{Z} - \widehat{\mathbf{Z}}\|_2 \leq 1/2$ instead of $\|\mathbf{Z} - \widehat{\mathbf{Z}}\|_2 < 1$. Both requirements insure that $\widehat{\mathbf{Z}}$ has linearly independent columns, hence represents a basis.*

   The stronger requirement $\|\mathbf{Z} - \widehat{\mathbf{Z}}\|_2 \leq 1/2$ also guarantees that the perturbed basis $\widehat{\mathbf{Z}}$ is well conditioned and that it is close to the exact basis $\mathbf{Z}$.

The lower bound in Theorem 1 simplifies when the columns of $\mathbf{Z}$ are dominant singular vectors of $\mathbf{A}$. No singular value gap is required below, as we merely pick the leading k columns of $\mathbf{U}$ from some SVD of $\mathbf{A}$, and then perturb them.

COROLLARY 1 (rank-preserving perturbations of dominant singular vectors). *Let $\mathbf{U}_k \in \mathbb{R}^{m \times k}$ in (1.2) be k dominant left singular vectors of $\mathbf{A}$. Denote by $\widehat{\mathbf{U}} \in \mathbb{R}^{m \times k}$ a perturbation with $\mathrm{rank}(\widehat{\mathbf{U}}) = k$ or $\|\mathbf{U}_k - \widehat{\mathbf{U}}\|_2 \leq 1/2$, and define*

$$\epsilon_U \equiv \|\widehat{\mathbf{U}}^\dagger\|_2 \, \|\mathbf{U}_k - \widehat{\mathbf{U}}\|_2.$$

Then

$$\||(\mathbf{I} - \mathbf{U}_k\mathbf{U}_k^T)\mathbf{A}\||_p \leq \||(\mathbf{I} - \widehat{\mathbf{U}}\widehat{\mathbf{U}}^\dagger)\mathbf{A}\||_p \leq \||(\mathbf{I} - \mathbf{U}_k\mathbf{U}_k^T)\mathbf{A}\||_p + \epsilon_U \,\||\mathbf{A}\||_p.$$

Proof. See section 2, and in particular Theorem 2.2. ☐

Next we consider perturbations in the matrix, with a bound that is completely general and holds for any projector $\mathbf{P}$ in any Schatten p-norm.

THEOREM 2 (additive perturbations in the matrix). *Let $\mathbf{A}$ and $\mathbf{A} + \mathbf{E} \in \mathbb{R}^{m \times n}$, and denote by $\mathbf{P} \in \mathbb{R}^{m \times m}$ an orthogonal projector as in (1.3). Then*

$$\left| \, \||(\mathbf{I} - \mathbf{P})(\mathbf{A} + \mathbf{E})\||_p - \||(\mathbf{I} - \mathbf{P})\mathbf{A}\||_p \, \right| \leq \||\mathbf{E}\||_p.$$

Proof. See section 2, and in particular Theorem 2.3. ☐

REMARK 2. *Theorem 2 shows what happens to the approximation error when the matrix changes from $\mathbf{A}$ to $\mathbf{A} + \mathbf{E}$. The change in the approximation error is proportional to the change $\||\mathbf{E}\||_p$ in the matrix. Thus, the low-rank approximation error is well conditioned (in the absolute sense) to additive perturbations in the matrix.*

Theorem 2 also implies the following upper bound for a low-rank approximation from singular vectors of $\mathbf{A} + \mathbf{E}$. Again, no singular value gap is required. We merely pick the leading k columns $\mathbf{U}_k$ obtained from some SVD of $\mathbf{A}$, and the leading k columns $\widehat{\mathbf{U}}_k$ obtained from some SVD of $\mathbf{A} + \mathbf{E}$.

COROLLARY 2 (low-rank approximation from additive perturbations). *Let $\mathbf{U}_k \in \mathbb{R}^{m \times k}$ in (1.2) be the k dominant left singular vectors of $\mathbf{A}$. Denote by $\widehat{\mathbf{U}}_k \in \mathbb{R}^{m \times k}$ the same number of dominant left singular vectors of $\mathbf{A} + \mathbf{E}$. Then*

$$\||(\mathbf{I} - \mathbf{U}_k\mathbf{U}_k^T)\mathbf{A}\||_p \leq \||(\mathbf{I} - \widehat{\mathbf{U}}_k\widehat{\mathbf{U}}_k^T)\mathbf{A}\||_p \leq \||(\mathbf{I} - \mathbf{U}_k\mathbf{U}_k^T)\mathbf{A}\||_p + 2\,\||\mathbf{E}\||_p.$$

Proof. See section 2, and in particular Corollary 2.4.   □

Bounds with an additive dependence on $\mathbf{E}$, like the two-norm bound above, can be derived for other Schatten p-norms as well, and can then be combined with bounds for $\mathbf{E}$ in [1, 2, 10] where $\mathbf{A} + \mathbf{E}$ is obtained from elementwise sampling from $\mathbf{A}$.

In the context of a different error measure, one can show [13, Theorem 4.5] that $\|\mathbf{A}_k - \mathbf{Y}\|_F \leq \frac{1+\sqrt{5}}{2}\|\mathbf{A} - \mathbf{Y}\|_F$ holds for any $\mathbf{Y} \in \mathbb{R}^{m \times n}$ with rank$(\mathbf{Y}) \leq k$.

1.4.2. Perturbations that change the matrix dimension. We consider perturbations that can change the number of columns in $\mathbf{A}$ and include, among others, multiplicative perturbations of the form $\widehat{\mathbf{A}} = \mathbf{A}\mathbf{X}$. However, our bounds are completely general and hold even in the absence of any relation between range$(\mathbf{A})$ and range$(\widehat{\mathbf{A}})$, for the two-norm (Theorem 3), the Frobenius norm (Theorem 4), and general Schatten p-norms (Theorem 5).

Our bounds are inspired by the analysis of randomized low-rank approximations [8], with errors evoking Gram matrix approximations $\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T$. Compared to existing work [7, 15, 28], (1.4) is much more general: It holds for any orthogonal projector $\mathbf{P}$ and is not limited to multiplicative perturbations $\widehat{\mathbf{A}} = \mathbf{A}\mathbf{X}$ where $\mathbf{X}$ samples and rescales columns. The bound (1.7) is identical to [8, Theorem 3].

THEOREM 3 (two-norm). *Let* $\mathbf{A} \in \mathbb{R}^{m \times n}$ *and* $\widehat{\mathbf{A}} \in \mathbb{R}^{m \times c}$. *Denote by* $\mathbf{P} \in \mathbb{R}^{m \times m}$ *an orthogonal projector as in* (1.3). *Then*

$$(1.4) \qquad \left| \|(\mathbf{I} - \mathbf{P})\,\mathbf{A}\|_2^2 - \|(\mathbf{I} - \mathbf{P})\,\widehat{\mathbf{A}}\|_2^2 \right| \;\leq\; \|\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T\|_2,$$

and

$$(1.5) \qquad \|(\mathbf{I} - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^\dagger)\,\mathbf{A}\|_2^2 \;\leq\; \|\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T\|_2.$$

More generally, denote by $\widehat{\mathbf{A}}_k \in \mathbb{R}^{m \times c}$ *a best rank-k approximation of* $\widehat{\mathbf{A}}$. *Then*

$$(1.6) \qquad \|(\mathbf{I} - \widehat{\mathbf{A}}_k\widehat{\mathbf{A}}_k^\dagger)\,\mathbf{A}\|_2^2 \;\leq\; \|\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}_k\widehat{\mathbf{A}}_k^T\|_2,$$

and also

$$(1.7) \qquad \|(\mathbf{I} - \widehat{\mathbf{A}}_k\widehat{\mathbf{A}}_k^\dagger)\,\mathbf{A}\|_2^2 \;\leq\; \|\mathbf{A} - \mathbf{A}_k\|_2^2 + 2\,\|\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T\|_2.$$

Proof. See section 2. Specifically, see Theorem 2.5 for (1.4); Theorem 2.6 for (1.5); and Theorem 2.7 for (1.7). The bound (1.6) simply follows from (1.5).   □

REMARK 3. *Theorem* 3, *as well as Theorems* 4 *and* 5 *to follow, show what happens to the approximation error when the matrix* $\mathbf{A}$ *is replaced by a potentially unrelated matrix of different dimension.*

Since $\mathbf{A}$ *and* $\widehat{\mathbf{A}}$ *do not have the same number of columns, the difference* $\mathbf{A} - \widehat{\mathbf{A}}$ *is not defined and cannot be used to measure the perturbations. Without any knowledge about* $\widehat{\mathbf{A}}$, *other than it has the same number of rows of* $\mathbf{A}$, *the most general approach is to represent the perturbation as the Gram matrix difference* $\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T$.

1. *The change in approximation error is small if* $\widehat{\mathbf{A}}$ *is a good Gram matrix approximation to* $\mathbf{A}$.
2. *The contribution of* range$(\mathbf{A})$ *orthogonal to* range$(\widehat{\mathbf{A}})$ *is bounded by the Gram matrix approximation error. Both errors are small if the column spaces of* $\mathbf{A}$ *and* $\widehat{\mathbf{A}}$ *are close. Note that* $\mathbf{I} - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^\dagger$ *is the orthogonal projector onto* range$(\widehat{\mathbf{A}})^\perp$.

3. *If we project* $\mathbf{A}$ *instead onto a larger space, that is, the space orthogonal to a best rank-k approximation of* $\widehat{\mathbf{A}}$, *then the contribution of* $\mathbf{A}$ *in that space can be bounded by the largest subdominant singular value* $\sigma_{k+1}(\mathbf{A})$ *plus the Gram matrix approximation error.*

The Frobenius norm bound (1.8) below is the first one of its kind in this generality, as it holds for any projector $\mathbf{P}$. The bound (1.11) is similar to *[8, Theorem 2]*, being weaker for smaller k but tighter for larger k.

THEOREM 4 (Frobenius norm). *Let* $\mathbf{A} \in \mathbb{R}^{m \times n}$ *and* $\widehat{\mathbf{A}} \in \mathbb{R}^{m \times c}$. *Denote by* $\mathbf{P} \in \mathbb{R}^{m \times m}$ *an orthogonal projector as in* (1.3) *with* $s \equiv \operatorname{rank}(\mathbf{P})$. *Then*

$$(1.8) \quad \left| \|(\mathbf{I} - \mathbf{P})\, \widehat{\mathbf{A}}\|_F^2 - \|(\mathbf{I} - \mathbf{P})\, \mathbf{A}\|_F^2 \right|$$

$$\leq \min \left\{ \|\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T\|_*, \ \sqrt{m-s}\, \|\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T\|_F \right\},$$

and

$$(1.9) \quad \|(\mathbf{I} - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^\dagger)\, \mathbf{A}\|_F^2 \leq \min \left\{ \|\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T\|_*, \ \sqrt{m-s}\, \|\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T\|_F \right\}.$$

More generally, denote by $\widehat{\mathbf{A}}_k \in \mathbb{R}^{m \times c}$ *a best rank-k approximation of* $\widehat{\mathbf{A}}$. *Then*

$$(1.10) \quad \|(\mathbf{I} - \widehat{\mathbf{A}}_k\widehat{\mathbf{A}}_k^\dagger)\, \mathbf{A}\|_F^2 \leq \min \left\{ \|\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}_k\widehat{\mathbf{A}}_k^T\|_*, \ \sqrt{m-k}\, \|\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}_k\widehat{\mathbf{A}}_k^T\|_F \right\},$$

and

$$(1.11)$$
$$\|(\mathbf{I} - \widehat{\mathbf{A}}_k\widehat{\mathbf{A}}_k^\dagger)\, \mathbf{A}\|_F^2 \leq \|\mathbf{A} - \mathbf{A}_k\|_F^2$$
$$+ 2 \min \left\{ \|\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T\|_*, \ \sqrt{m-k}\, \|\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T\|_F \right\}.$$

Proof. See section 2. Specifically, see Theorem 2.5 for (1.8); Theorem 2.6 for (1.9); and Theorem 2.7 for (1.11). The bound (1.10) simply follows from (1.9). □

REMARK 4. *Theorem* 4 *has the same interpretation as Theorem* 3. *There are two differences, though.*

First, the Gram matrix approximation error in the Frobenius norm is amplified by the factor $\sqrt{m - \operatorname{rank}(\mathbf{P})}$, *which is small if* $\operatorname{rank}(\mathbf{P})$ *is close to* m *and* $\operatorname{range}(\mathbf{P})$ *covers a large portion of* $\mathbb{R}^m$. *In this case the low-rank approximation errors are small.*

Second, Theorem 4 *relates the approximation error in the Frobenius norm to the Gram matrix approximation error in the trace norm, i.e., the Schatten one-norm. This is a novel connection and should motivate further work into understanding the behavior of the trace norm, thereby complementing prior investigations into the two- and Frobenius norms.*

To the best of our knowledge, Theorem 5 is new. It generalizes Theorems 3 and 4, and is the first nontrivial bound to connect low-rank approximations with Gram matrix approximation errors in general Schatten p-norms.

THEOREM 5 (general Schatten p-norms). *Let* $\mathbf{A} \in \mathbb{R}^{m \times n}$ *and* $\widehat{\mathbf{A}} \in \mathbb{R}^{m \times c}$. *Denote by* $\mathbf{P} \in \mathbb{R}^{m \times m}$ *an orthogonal projector as in* (1.3) *with* $s \equiv \operatorname{rank}(\mathbf{P})$. *Then*

$$(1.12) \quad \left| \|\|(\mathbf{I} - \mathbf{P})\,\mathbf{A}\|\|_p^2 - \|\|(\mathbf{I} - \mathbf{P})\,\widehat{\mathbf{A}}\|\|_p^2 \right|$$

$$\leq \min \left\{ \|\|\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T\|\|_{p/2}, \quad \sqrt[p]{m-s}\, \|\|\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T\|\|_p \right\},$$

and

(1.13) $\||(\mathbf{I} - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^\dagger)\,\mathbf{A}\||_{p/2}^2$

$$\leq \min\left\{ \,\||\,\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T\||_{p/2}, \quad \sqrt[p]{m-s}\,\,\||\,\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T\,\||_p \,\right\}.$$

More generally, denote by $\widehat{\mathbf{A}}_k \in \mathbb{R}^{m\times c}$ a best rank-k approximation of $\widehat{\mathbf{A}}$. Then

(1.14) $\||(\mathbf{I} - \widehat{\mathbf{A}}_k\widehat{\mathbf{A}}_k^\dagger)\,\mathbf{A}\||_{p/2}^2$

$$\leq \min\left\{ \,\||\,\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}_k\widehat{\mathbf{A}}_k^T\||_{p/2}, \quad \sqrt[p]{m-k}\,\,\||\,\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}_k\widehat{\mathbf{A}}_k^T\,\||_p \,\right\},$$

and

(1.15)
$$\||(\mathbf{I} - \widehat{\mathbf{A}}_k\widehat{\mathbf{A}}_k^\dagger)\,\mathbf{A}\||_{p/2}^2 \leq \||\mathbf{A} - \mathbf{A}_k\||_p^2$$
$$+ 2\min\left\{ \,\||\,\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T\||_{p/2},\ \sqrt[p]{m-k}\,\,\||\,\mathbf{A}\mathbf{A}^T - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T\,\||_p \,\right\}.$$

Proof. See section 2. Specifically, see Theorem 2.5 for (1.12); Theorem 2.6 for (1.13); and Theorem 2.7 for (1.15). The bound (1.14) simply follows from (1.13). □

1.4.3. Relations between low-rank approximation error and subspace angle. For matrices $\mathbf{A}$ whose dominant singular values are separated by a gap from the subdominant singular values, we bound the low-rank approximation error in terms of the subspace angle (Theorem 6) and discuss the tightness of the bounds (Remark 6).

To guarantee that the k-dimensional dominant subspace of $\mathbf{A}$ is well defined requires the existence of gap after the kth singular value,

$$\|\mathbf{A}\|_2 = \sigma_1(\mathbf{A}) \geq \cdots \geq \sigma_k(\mathbf{A}) > \sigma_{k+1}(\mathbf{A}) \geq \cdots \geq \sigma_r(\mathbf{A}) \geq 0, \qquad r \equiv \min\{m,n\}.$$

THEOREM 6. *Let $\mathbf{P}_k \equiv \mathbf{A}_k\mathbf{A}_k^\dagger$ be the orthogonal projector onto the dominant k-dimensional subspace of $\mathbf{A}$. Denote by $\mathbf{P} \in \mathbb{R}^{m\times m}$ any orthogonal projector as in (1.3) with $k \leq \mathrm{rank}(\mathbf{P}) < m - k$. Then*

$$\sigma_k(\mathbf{A})\,\||\sin\boldsymbol{\Theta}(\mathbf{P},\mathbf{P}_k)\||_p \leq \||(\mathbf{I}-\mathbf{P})\mathbf{A}\||_p \leq \|\mathbf{A}\|_2\,\||\sin\boldsymbol{\Theta}(\mathbf{P},\mathbf{P}_k)\,\||_p +\||\,\mathbf{A} - \mathbf{A}_k\,\||_p.$$

Proof. See section 3, and in particular Theorem 3.1 for the lower bound, and Theorems 3.2 and 3.3 for the upper bounds. □

REMARK 5. *This is a comparison between the approximation error $(\mathbf{I} - \mathbf{P})\mathbf{A}$ on the one hand, and the angle between the approximation $\mathrm{range}(\mathbf{P})$ and the target space $\mathrm{range}(\mathbf{P}_k)$ on the other—without any assumptions on the accuracy of the approximation.*

The singular value gap is required to guarantee that the dominant subspace, represented by $\mathbf{P}_k$, is well defined.

However, the remaining assumptions pose no constraints in the proper context. Practical low-rank approximations target subspaces whose dimension is small compared to that of the host space. To be effective at all, an approximation $\mathrm{range}(\mathbf{P})$ must have a dimension that covers, if not exceeds, that of the target space.

REMARK 6 (tightness of Theorem 6). *The subspace angles in lower and upper bounds are amplified by a dominant singular value; and the upper bound contains an additive term consisting of subdominant singular values.*

- If $\operatorname{rank}(\mathbf{A}) = k$, so that $\mathbf{A} - \mathbf{A}_k = \mathbf{0}$, then the tightness depends on the spread of the nonzero singular values,

$$\sigma_k(\mathbf{A}) \; \||| \sin \mathbf{\Theta}(\mathbf{P}, \mathbf{P}_k) \||\|_p \leq \||| (\mathbf{I} - \mathbf{P}) \, \mathbf{A} \||\|_p \leq \|\mathbf{A}\|_2 \; \||| \sin \mathbf{\Theta}(\mathbf{P}, \mathbf{P}_k) \, \||\|_p \, .$$

- If $\operatorname{rank}(\mathbf{A}) = k$ and $\sigma_1(\mathbf{A}) = \cdots = \sigma_k(\mathbf{A})$, then the bounds are tight, and they are equal to

$$\||| (\mathbf{I} - \mathbf{P}) \, \mathbf{A} \||\|_p = \|\mathbf{A}\|_2 \; \||| \sin \mathbf{\Theta}(\mathbf{P}, \mathbf{P}_k) \, \||\|_p \, .$$

- If $\operatorname{range}(\mathbf{P}) = \operatorname{range}(\mathbf{P}_k)$, so that $\sin \mathbf{\Theta}(\mathbf{P}, \mathbf{P}_k) = \mathbf{0}$, then the upper bound is tight and equal to

$$\||| (\mathbf{I} - \mathbf{P}) \, \mathbf{A} \||\|_p = \||| \mathbf{A} - \mathbf{A}_k \, \||\|_p \, .$$

2. Well-conditioning of low-rank approximations. We investigate the effect of additive perturbations in the projector basis $\mathbf{Z}$ on the orthogonal projector as a whole (section 2.1) and on the approximation error (section 2.2); and the effect of matrix perturbations on the approximation error (section 2.3). At last, we relate the low-rank approximation error to the error in Gram matrix approximation (section 2.4).

2.1. Orthogonal projectors, and perturbations in the projector basis. We show that orthogonal projectors and subspace angles are insensitive to additive, rank-preserving perturbations in the projector basis (Theorem 2.1) if the perturbed projector basis is well conditioned.

THEOREM 2.1. *Let* $\mathbf{Z} \in \mathbb{R}^{m \times s}$ *be a projector basis with orthonormal columns so that* $\mathbf{Z}^T \mathbf{Z} = \mathbf{I}_s$. *Denote by* $\widehat{\mathbf{Z}} \in \mathbb{R}^{m \times s}$ *a perturbation, and set* $\epsilon_Z \equiv \|\widehat{\mathbf{Z}}^\dagger\|_2 \|\mathbf{Z} - \widehat{\mathbf{Z}}\|_2$.
1. *If* $\operatorname{rank}(\widehat{\mathbf{Z}}) = \operatorname{rank}(\mathbf{Z})$, *then the* distance *between* $\operatorname{range}(\mathbf{Z})$ *and* $\operatorname{range}(\widehat{\mathbf{Z}})$ *is*

$$\|\mathbf{Z}\mathbf{Z}^T - \widehat{\mathbf{Z}}\widehat{\mathbf{Z}}^\dagger\|_2 = \| \sin \mathbf{\Theta}(\mathbf{Z}, \widehat{\mathbf{Z}})\|_2 \leq \epsilon_Z.$$

2. *If* $\|\mathbf{Z} - \widehat{\mathbf{Z}}\|_2 \leq 1/2$, *then* $\operatorname{rank}(\widehat{\mathbf{Z}}) = \operatorname{rank}(\mathbf{Z})$ *and* $\epsilon_Z \leq 2 \|\mathbf{Z} - \widehat{\mathbf{Z}}\|_2$.

Proof.
1. The equality follows from Lemma 1.5. The upper bound follows from [24, Theorem 3.1] and [14, Lemma 20.12], but we provide a simpler proof adapted for this context. Set $\mathbf{F} = \widehat{\mathbf{Z}} - \mathbf{Z}$, and substitute

$$\widehat{\mathbf{P}} \equiv \widehat{\mathbf{Z}}\widehat{\mathbf{Z}}^\dagger = (\mathbf{Z} + \widehat{\mathbf{Z}} - \mathbf{Z}) \, \widehat{\mathbf{Z}}^\dagger = (\mathbf{Z} + \mathbf{F}) \, \widehat{\mathbf{Z}}^\dagger$$

into

$$(\mathbf{I} - \mathbf{Z}\mathbf{Z}^T)\widehat{\mathbf{P}} = (\mathbf{I} - \mathbf{Z}\mathbf{Z}^T)(\mathbf{Z} + \mathbf{F}) \, \widehat{\mathbf{Z}}^\dagger = (\mathbf{I} - \mathbf{Z}\mathbf{Z}^T) \, \mathbf{F} \, \widehat{\mathbf{Z}}^\dagger,$$

and apply Lemma 1.5,

$$(2.1) \qquad \| \sin \mathbf{\Theta}(\mathbf{Z}, \widehat{\mathbf{Z}})\|_2 = \|(\mathbf{I} - \mathbf{Z}\mathbf{Z}^T)\widehat{\mathbf{P}}\|_2 \leq \|\widehat{\mathbf{Z}}^\dagger\|_2 \|\mathbf{F}\|_2.$$

2. To show $\operatorname{rank}(\widehat{\mathbf{Z}}) = \operatorname{rank}(\mathbf{Z})$ in the case $\|\mathbf{Z} - \widehat{\mathbf{Z}}\|_2 \leq 1/2$, consider the singular values $\sigma_j(\mathbf{Z}) = 1$ and $\sigma_j(\widehat{\mathbf{Z}})$, $1 \leq j \leq s$. The well-conditioning of singular values [11, Corollary 8.6.2] implies

$$\left|1 - \sigma_j(\widehat{\mathbf{Z}})\right| = \left|\sigma_j(\mathbf{Z}) - \sigma_j(\widehat{\mathbf{Z}})\right| \leq \|\mathbf{F}\|_2 \leq 1/2, \qquad 1 \leq j \leq s.$$

Thus $\min_{1 \leq j \leq s} \sigma_j(\widehat{\mathbf{Z}}) \geq 1/2 > 0$ and $\operatorname{rank}(\widehat{\mathbf{P}}) = \operatorname{rank}(\widehat{\mathbf{Z}}) = s = \operatorname{rank}(\mathbf{P})$. Hence (2.1) holds with

$$(2.2) \qquad \| \sin \mathbf{\Theta}(\mathbf{Z}, \widehat{\mathbf{Z}})\|_2 \leq \|\widehat{\mathbf{Z}}^\dagger\|_2 \|\mathbf{F}\|_2 \leq 2\|\mathbf{F}\|_2. \qquad \square$$

2.2. Approximation errors, and perturbations in the projector basis.
We show that low-rank approximation errors are insensitive to additive, rank-preserving perturbations in the projector basis (Theorem 2.2), provided the perturbed projector basis remains well-conditioned.

THEOREM 2.2. *Let* $\mathbf{A} \in \mathbb{R}^{m \times n}$, *and let* $\mathbf{Z} \in \mathbb{R}^{m \times k}$ *be a projector basis with orthonormal columns so that* $\mathbf{Z}^T \mathbf{Z} = \mathbf{I}_k$. *Denote by* $\widehat{\mathbf{Z}} \in \mathbb{R}^{m \times k}$ *a perturbation, and set* $\epsilon_Z \equiv \|\widehat{\mathbf{Z}}^\dagger\|_2 \|\mathbf{Z} - \widehat{\mathbf{Z}}\|_2$.

1. *If* $\mathrm{rank}(\widehat{\mathbf{Z}}) = \mathrm{rank}(\mathbf{Z})$, *then*

$$\||(\mathbf{I} - \mathbf{Z}\mathbf{Z}^T)\mathbf{A}\||_p - \epsilon_Z \, \|| \mathbf{A}\||_p \le \||(\mathbf{I} - \widehat{\mathbf{Z}}\widehat{\mathbf{Z}}^\dagger)\mathbf{A}\||_p$$
$$\le \||(\mathbf{I} - \mathbf{Z}\mathbf{Z}^T)\mathbf{A}\||_p + \epsilon_Z \, \|| \mathbf{A} \||_p \, .$$

2. *If* $\|\mathbf{Z} - \widehat{\mathbf{Z}}\|_2 \le 1/2$, *then* $\mathrm{rank}(\widehat{\mathbf{Z}}) = \mathrm{rank}(\mathbf{Z})$ *and* $\epsilon_Z \le 2 \|\mathbf{Z} - \widehat{\mathbf{Z}}\|_2$.
3. *If, in addition,* $\mathbf{Z} = \mathbf{U}_k$ *are* k *dominant singular vectors of* $\mathbf{A}$, *then*

$$\||(\mathbf{I} - \mathbf{U}_k\mathbf{U}_k^T)\mathbf{A}\||_p \le \||(\mathbf{I} - \widehat{\mathbf{Z}}\widehat{\mathbf{Z}}^\dagger)\mathbf{A}\||_p \le \||(\mathbf{I} - \mathbf{U}_k\mathbf{U}_k^T)\mathbf{A}\||_p + \epsilon_U \, \|| \mathbf{A}\||_p,$$

where $\epsilon_U \equiv \|\widehat{\mathbf{Z}}^\dagger\|_2 \|\mathbf{U}_k - \widehat{\mathbf{Z}}\|_2$.

Proof. Abbreviate $\mathbf{P} \equiv \mathbf{Z}\mathbf{Z}^T$ and $\widehat{\mathbf{P}} \equiv \widehat{\mathbf{Z}}\widehat{\mathbf{Z}}^\dagger$, and write

$$(\mathbf{I} - \widehat{\mathbf{P}})\mathbf{A} = (\mathbf{I} - \mathbf{P})\mathbf{A} + (\mathbf{P} - \widehat{\mathbf{P}})\mathbf{A}.$$

1. Apply the triangle and reverse triangle inequalities, followed by strong submultiplicativity in Lemma 1.3. Then bound the second summand according to item 1 in Theorem 2.1 as follows,

$$\||(\mathbf{P} - \widehat{\mathbf{P}})\mathbf{A}\||_p \le \|\widehat{\mathbf{Z}}^\dagger\|_2 \|\widehat{\mathbf{Z}} - \mathbf{Z}\|_2 \, \|| \mathbf{A}\||_p = \epsilon_Z \, \|| \mathbf{A} \||_p \, .$$

2. This follows from item 2 in Theorem 2.1.
3. In the lower bound, use the optimality of the SVD from Lemma 1.3. $\qquad \square$

2.3. Approximation errors, and perturbations in the matrix. We show that low-rank approximation errors are insensitive to matrix perturbations that are either additive (Theorem 2.3 and Corollary 2.4), or dimension changing (Theorem 2.5).

THEOREM 2.3 (additive perturbations). *Let* $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{m \times n}$, *denote by* $\mathbf{P} \in \mathbb{R}^{m \times m}$ *an orthogonal projector as in* (1.3), *and let* $p \ge 1$ *be an integer. Then*

$$\||(\mathbf{I} - \mathbf{P})\mathbf{A} \||_p - \|| \mathbf{E}\||_p \le \||(\mathbf{I} - \mathbf{P})(\mathbf{A} + \mathbf{E})\||_p \le \||(\mathbf{I} - \mathbf{P})\mathbf{A} \||_p + \|| \mathbf{E} \||_p \, .$$

Proof. Apply the triangle and reverse triangle inequalities, followed by strong submultiplicativity in Lemma 1.3, and the fact that an orthogonal projector $\mathbf{P}$ satisfies $\|\mathbf{I} - \mathbf{P}\|_2 \le 1$. $\qquad \square$

COROLLARY 2.4 (low-rank approximation from singular vectors of $\mathbf{A} + \mathbf{E}$). *Let* $\mathbf{U}_k \in \mathbb{R}^{m \times k}$ *in* (1.2) *be* k *dominant left singular vectors of* $\mathbf{A}$; *and let* $\widehat{\mathbf{U}}_k \in \mathbb{R}^{m \times k}$ *be* k *dominant left singular vectors of* $\mathbf{A} + \mathbf{E}$. *Then*

$$\||(\mathbf{I} - \mathbf{U}_k\mathbf{U}_k^T)\mathbf{A}\||_p \le \||(\mathbf{I} - \widehat{\mathbf{U}}_k\widehat{\mathbf{U}}_k^T)\mathbf{A}\||_p \le \||(\mathbf{I} - \mathbf{U}_k\mathbf{U}_k^T)\mathbf{A} \||_p + 2 \|| \mathbf{E} \||_p \, .$$

Proof. The lower bound follows from the optimality of the SVD of $\mathbf{A}$ in all Schatten p-norms; see Lemma 1.3.

As for the upper bound, set $\mathbf{P} = \widehat{\mathbf{U}}_k \widehat{\mathbf{U}}_k^T$ in the upper bound of Theorem 2.3,

$$\||(\mathbf{I} - \widehat{\mathbf{U}}_k \widehat{\mathbf{U}}_k^T)\mathbf{A}\||_p \leq \||(\mathbf{I} - \widehat{\mathbf{U}}_k \widehat{\mathbf{U}}_k^T)(\mathbf{A} + \mathbf{E})\||_p + \||\mathbf{E}\||_p .$$

Since $\widehat{\mathbf{U}}_k$ are singular vectors of $\mathbf{A} + \mathbf{E}$, the optimality of the SVD of $\mathbf{A} + \mathbf{E}$, see Lemma 1.3, followed by another application of Theorem 2.3, yields

$$\||(\mathbf{I} - \widehat{\mathbf{U}}_k \widehat{\mathbf{U}}_k^T)(\mathbf{A} + \mathbf{E})\||_p = \min_{\mathbf{Z}^T\mathbf{Z} = \mathbf{I}_k} \||(\mathbf{I} - \mathbf{Z}\mathbf{Z}^T)(\mathbf{A} + \mathbf{E})\||_p \leq \||(\mathbf{I} - \mathbf{U}_k \mathbf{U}_k^T)(\mathbf{A} + \mathbf{E})\||_p$$
$$\leq \||(\mathbf{I} - \mathbf{U}_k \mathbf{U}_k^T)\mathbf{A}\||_p + \||\mathbf{E}\||_p .$$

THEOREM 2.5 (perturbations that change the number of columns). *Let* $\mathbf{A} \in \mathbb{R}^{m \times n}$ *and* $\widehat{\mathbf{A}} \in \mathbb{R}^{m \times c}$. *Denote by* $\mathbf{P} \in \mathbb{R}^{m \times m}$ *an orthogonal projector as in* (1.3), *set* $s \equiv \mathrm{rank}(\mathbf{P})$, *and let* $p \geq 1$ *be an even integer. Then*
   1. *Two norm* ($p = \infty$)

$$\left| \|(\mathbf{I} - \mathbf{P})\mathbf{A}\|_2^2 - \|(\mathbf{I} - \mathbf{P})\widehat{\mathbf{A}}\|_2^2 \right| \leq \|\widehat{\mathbf{A}}\widehat{\mathbf{A}}^T - \mathbf{A}\mathbf{A}^T\|_2.$$

   2. *Schatten p-norm* (p *even*)

$$\left| \||(\mathbf{I} - \mathbf{P})\mathbf{A}\||_p^2 - \||(\mathbf{I} - \mathbf{P})\widehat{\mathbf{A}}\||_p^2 \right|$$
$$\leq \min \left\{ \||\widehat{\mathbf{A}}\widehat{\mathbf{A}}^T - \mathbf{A}\mathbf{A}^T\||_{p/2}, \quad \sqrt[p]{m - s} \, \||\widehat{\mathbf{A}}\widehat{\mathbf{A}}^T - \mathbf{A}\mathbf{A}^T\||_p \right\}.$$

   3. *Frobenius norm* ($p = 2$)

$$\left| \|(\mathbf{I} - \mathbf{P})\widehat{\mathbf{A}}\|_F^2 - \|(\mathbf{I} - \mathbf{P})\mathbf{A}\|_F^2 \right|$$
$$\leq \min \left\{ \|\widehat{\mathbf{A}}\widehat{\mathbf{A}}^T - \mathbf{A}\mathbf{A}^T\|_*, \quad \sqrt{m - s}\, \|\widehat{\mathbf{A}}\widehat{\mathbf{A}}^T - \mathbf{A}\mathbf{A}^T\|_F \right\}.$$

Proof. The proof is motivated by that of [8, Theorems 2 and 3]. If $s = m$, then $\mathbf{P} = \mathbf{I}_m$ and the bounds follow from the reverse triangle inequality. So let $s < m$.
   1. *Two-norm.* The invariance of the two-norm under transposition and the triangle inequality imply

$$\|(\mathbf{I} - \mathbf{P})\widehat{\mathbf{A}}\|_2^2 = \|\widehat{\mathbf{A}}^T(\mathbf{I} - \mathbf{P})\|_2^2 = \|(\mathbf{I} - \mathbf{P})\widehat{\mathbf{A}}\widehat{\mathbf{A}}^T(\mathbf{I} - \mathbf{P})\|_2$$
$$= \|(\mathbf{I} - \mathbf{P})\mathbf{A}\mathbf{A}^T(\mathbf{I} - \mathbf{P}) + (\mathbf{I} - \mathbf{P})\left(\widehat{\mathbf{A}}\widehat{\mathbf{A}}^T - \mathbf{A}\mathbf{A}^T\right)(\mathbf{I} - \mathbf{P})\|_2$$
$$\leq \|(\mathbf{I} - \mathbf{P})\mathbf{A}\mathbf{A}^T(\mathbf{I} - \mathbf{P})\|_2 + \|(\mathbf{I} - \mathbf{P})\left(\widehat{\mathbf{A}}\widehat{\mathbf{A}}^T - \mathbf{A}\mathbf{A}^T\right)(\mathbf{I} - \mathbf{P})\|_2.$$

The first summand on the right equals

$$\|(\mathbf{I} - \mathbf{P})\mathbf{A}\mathbf{A}^T(\mathbf{I} - \mathbf{P})\|_2 = \|(\mathbf{I} - \mathbf{P})\mathbf{A}\|_2^2,$$

while the second one can be bounded by submultiplicativity and $\|\mathbf{I} - \mathbf{P}\|_2 \leq 1$,

$$\|(\mathbf{I} - \mathbf{P})\left(\widehat{\mathbf{A}}\widehat{\mathbf{A}}^T - \mathbf{A}\mathbf{A}^T\right)(\mathbf{I} - \mathbf{P})\|_2 \leq \|\mathbf{I} - \mathbf{P}\|_2^2 \|\widehat{\mathbf{A}}\widehat{\mathbf{A}}^T - \mathbf{A}\mathbf{A}^T\|_2$$
$$\leq \|\widehat{\mathbf{A}}\widehat{\mathbf{A}}^T - \mathbf{A}\mathbf{A}^T\|_2.$$

This gives the upper bound

$$\|(\mathbf{I} - \mathbf{P})\widehat{\mathbf{A}}\|_2^2 - \|(\mathbf{I} - \mathbf{P})\mathbf{A}\|_2^2 \leq \|\widehat{\mathbf{A}}\widehat{\mathbf{A}}^T - \mathbf{A}\mathbf{A}^T\|_2.$$

Apply the inverse triangle inequality to show the lower bound,

$$-\|\widehat{\mathbf{A}}\widehat{\mathbf{A}}^T - \mathbf{A}\mathbf{A}^T\|_2 \leq \|(\mathbf{I} - \mathbf{P})\widehat{\mathbf{A}}\|_2^2 - \|(\mathbf{I} - \mathbf{P})\mathbf{A}\|_2^2.$$

2. *Schatten p-norm (p even).* The proof is similar to that of the two-norm, since an even Schatten p-norm is a *Q-norm* [3, Definition IV.2.9], meaning it represents a quadratic gauge function. This can be seen in terms of singular values, where for any matrix $\mathbf{C}$,

$$\|\|\mathbf{C}\|\|_p^p = \sum_j \left(\sigma_j(\mathbf{C})\right)^p = \sum_j \left(\sigma_j(\mathbf{C}\mathbf{C}^T)\right)^{p/2} = \|\|\mathbf{C}\mathbf{C}^T\|\|_{p/2}^{p/2}.$$

Hence

(2.3) $$\|\|\mathbf{C}\|\|_p^2 = \|\|\mathbf{C}\mathbf{C}^T\|\|_{p/2}.$$

Abbreviate $\mathbf{M} \equiv \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T - \mathbf{A}\mathbf{A}^T$, and $\mathbf{B} \equiv \mathbf{I} - \mathbf{P}$ where $\mathbf{B}^T = \mathbf{B}$ and $\|\mathbf{B}\|_2 = 1$. Since singular values do not change under transposition, it follows from (2.3) and the triangle inequality that

(2.4) $$\|\|\mathbf{B}\,\widehat{\mathbf{A}}\|\|_p^2 = \|\|\widehat{\mathbf{A}}^T\,\mathbf{B}\|\|_p^2 = \|\|\mathbf{B}\,\widehat{\mathbf{A}}\widehat{\mathbf{A}}^T\,\mathbf{B}\|\|_{p/2} = \|\|\mathbf{B}\mathbf{A}\mathbf{A}^T\mathbf{B}\,+\,\mathbf{B}\mathbf{M}\mathbf{B}\|\|_{p/2}$$
$$\leq \|\|\mathbf{B}\,\mathbf{A}\mathbf{A}^T\,\mathbf{B}\|\|_{p/2} + \|\|\mathbf{B}\mathbf{M}\mathbf{B}\|\|_{p/2}.$$

Apply (2.3) to the first summand on the right, $\|\|\mathbf{B}\,\mathbf{A}\mathbf{A}^T\,\mathbf{B}\|\|_{p/2} = \|\|\mathbf{B}\,\mathbf{A}\|\|_p^2$, and insert it into the above inequalities,

(2.5) $$\|\|\mathbf{B}\,\widehat{\mathbf{A}}\|\|_p^2 - \|\|\mathbf{B}\,\mathbf{A}\|\|_p^2 \leq \|\|\mathbf{B}\mathbf{M}\mathbf{B}\|\|_{p/2}.$$

1. Derivation of the first term in the minimum: Bound the rightmost term in (2.5) with strong submultiplicativity and $\|\mathbf{B}\|_2 = 1$,

$$\|\|\mathbf{B}\mathbf{M}\mathbf{B}\|\|_{p/2} \leq \|\mathbf{B}\|_2^2\,\|\|\mathbf{M}\|\|_{p/2} \leq \|\|\mathbf{M}\|\|_{p/2},$$

which gives the upper bound

$$\|\|\mathbf{B}\,\widehat{\mathbf{A}}\|\|_p^2 - \|\|\mathbf{B}\,\mathbf{A}\|\|_p^2 \leq \|\|\mathbf{M}\|\|_{p/2}.$$

Apply the inverse triangle inequality in (2.4) to show the lower bound

$$-\|\|\mathbf{M}\|\|_{p/2} \leq \|\|\mathbf{B}\,\widehat{\mathbf{A}}\|\|_p^2 - \|\|\mathbf{B}\,\mathbf{A}\|\|_p^2.$$

2. Derivation of the second term in the minimum: From

$$\operatorname{rank}\left(\mathbf{B}\mathbf{M}\mathbf{B}\right) \leq \operatorname{rank}(\mathbf{B}) = \operatorname{rank}(\mathbf{I} - \mathbf{P}) = m - s > 0$$

follows $\sigma_j(\mathbf{B}) = 1, \ \leq j \leq m - s$. With the nonascending singular value ordering in (1.1), the Schatten p-norm needs to sum over only the largest nonzero $m - s$ singular values. This, together with the singular value inequality [17, (7.3.14)]

$$\sigma_j(\mathbf{B}\mathbf{M}\mathbf{B}) \leq \sigma_1(\mathbf{B})^2\,\sigma_j(\mathbf{M}) = 1 \cdot \sigma_1(\mathbf{M}), \qquad 1 \leq j \leq m - s,$$

gives for the rightmost term in (2.5):

$$\||\mathbf{B\,M\,B}\||_{p/2}^{p/2} = \sum_{j=1}^{m-s} (\sigma_j(\mathbf{B\,M\,B}))^{p/2} \leq \sum_{j=1}^{m-s} 1 \cdot (\sigma_j(\mathbf{M}))^{p/2}.$$

Then apply the Cauchy–Schwarz inequality to the vectors of singular values:

$$\sum_{j=1}^{m-s} 1 \cdot (\sigma_j(\mathbf{M}))^{p/2} \leq \sqrt{m-s} \sqrt{\sum_{j=1}^{m-s} (\sigma_j(\mathbf{M}))^{p}} \leq \sqrt{m-s}\; \||\mathbf{M}\||_p^{p/2}.$$

Merging the last two sequences of inequalities gives

$$\||\mathbf{B\,M\,B}\||_{p/2}^{p/2} \leq \sqrt{m-s}\; \||\mathbf{M}\||_p^{p/2}.$$

Thus $\||\mathbf{B\,M\,B}\||_{p/2} \leq \sqrt[p]{m-s}\; \||\mathbf{M}\||_p$, which can now be substituted into (2.5).

3. *Frobenius norm.* This is the special case $p = 2$ with $\||\mathbf{A}\||_2 = \|\mathbf{A}\|_F$ and $\||\mathbf{A}\||_1 = \|\mathbf{A}\|_*$. $\qquad\qquad\square$

2.4. Approximation error, and Gram matrix approximation. We generalize [8, Theorems 2 and 3] to Schatten p-norms.

THEOREM 2.6. *Let* $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{C} \in \mathbb{R}^{m \times c}$ *with* $s \equiv \mathrm{rank}(\mathbf{C})$; *and let* $p \geq 1$ *be an even integer. Then*

1. *Two-norm* $(p = \infty)$

$$\|(\mathbf{I} - \mathbf{CC}^\dagger)\,\mathbf{A}\|_2^2 \ \leq\ \|\mathbf{AA}^T - \mathbf{CC}^T\|_2.$$

2. *Schatten p-norm* $(p\ even)$

$$\||(\mathbf{I} - \mathbf{CC}^\dagger)\,\mathbf{A}\||_p^2 \ \leq\ \min\left\{ \||\mathbf{AA}^T - \mathbf{CC}^T\||_{p/2}, \quad \sqrt[p]{m-s}\,\||\mathbf{AA}^T - \mathbf{CC}^T\||_p \right\}.$$

3. *Frobenius norm* $(p = 2)$

$$\|(\mathbf{I} - \mathbf{CC}^\dagger)\,\mathbf{A}\|_F^2 \ \leq\ \min\left\{ \|\mathbf{AA}^T - \mathbf{CC}^T\|_*, \quad \sqrt{m-s}\,\|\mathbf{AA}^T - \mathbf{CC}^T\|_F \right\}.$$

Proof. Set $\mathbf{P} = \mathbf{CC}^\dagger$ and $\widehat{\mathbf{A}} = \mathbf{C}$. The properties of the Moore–Penrose inverse imply

$$(\mathbf{I} - \mathbf{P})\,\widehat{\mathbf{A}} = (\mathbf{I} - \mathbf{CC}^\dagger)\,\mathbf{C} = \mathbf{C} - \mathbf{CC}^\dagger\mathbf{C} = \mathbf{0}.$$

When substituting this into Theorem 2.5, the second summand on the left of the bounds drops out.

In addition, for the Frobenius and Schatten p-norm bounds, use $\mathrm{rank}(\mathbf{P}) = \mathrm{rank}(\mathbf{C}) = s$. $\qquad\qquad\square$

Recall *Mirsky's Theorem* [17, Corollary 7.4.9.3], an extension of the Hoffman–Wielandt theorem to any unitarily invariant norm and, in particular, Schatten p-norms: For $\mathbf{A}, \mathbf{H} \in \mathbb{R}^{m \times n}$, the singular values $\sigma_j(\mathbf{AA}^T)$ and $\sigma_j(\mathbf{HH}^T)$, $1 \leq j \leq m$, are also eigenvalues and they satisfy

$$(2.6) \qquad \sum_{j=1}^{m} |\sigma_j(\mathbf{AA}^T) - \sigma_j(\mathbf{HH}^T)|^p \leq \||\mathbf{AA}^T - \mathbf{HH}^T\||_p^p.$$

THEOREM 2.7. *Let* $\mathbf{A} \in \mathbb{R}^{m \times n}$ *and* $\mathbf{C} \in \mathbb{R}^{m \times c}$. *Denote by* $\mathbf{C}_k$ *a best rank-k approximation to* $\mathbf{C}$; *and let* $p \geq 1$ *be an even integer. Then*

1. *Two-norm* ($p = \infty$)

$$\|(\mathbf{I} - \mathbf{C}_k \mathbf{C}_k^\dagger)\,\mathbf{A}\|_2^2 \ \leq \ \|\mathbf{A} - \mathbf{A}_k\|_2^2 + 2\,\|\mathbf{A}\mathbf{A}^T - \mathbf{C}\mathbf{C}^T\|_2.$$

2. *Schatten p-norm* (*p even*)

$$\|\|(\mathbf{I} - \mathbf{C}_k \mathbf{C}_k^\dagger)\,\mathbf{A}\|\|_p^2 \leq \|\|\mathbf{A} - \mathbf{A}_k\|\|_p^2$$
$$+ 2 \min \left\{ \|\|\,\mathbf{A}\mathbf{A}^T - \mathbf{C}\mathbf{C}^T\|\|_{p/2},\ \sqrt[p]{m-k}\,\|\|\mathbf{A}\mathbf{A}^T - \mathbf{C}\mathbf{C}^T\|\|_p \right\}.$$

3. *Frobenius norm* ($p = 2$)

$$\|(\mathbf{I} - \mathbf{C}_k \mathbf{C}_k^\dagger)\,\mathbf{A}\|_F^2 \leq \|\mathbf{A} - \mathbf{A}_k\|_F^2$$
$$+ 2 \min \left\{ \|\mathbf{A}\mathbf{A}^T - \mathbf{C}\mathbf{C}^T\|_*,\ \sqrt{m-k}\,\|\mathbf{A}\mathbf{A}^T - \mathbf{C}\mathbf{C}^T\|_F \right\}.$$

Proof. We first introduce some notation before proving the bounds.

0. *Set-up.* Partition $\mathbf{A} = \mathbf{A}_k + \mathbf{A}_\perp$ and $\mathbf{C} = \mathbf{C}_k + \mathbf{C}_\perp$ to distinguish the respective best rank-k approximations $\mathbf{A}_k$ and $\mathbf{C}_k$. From $\mathbf{A}_k \mathbf{A}_\perp^T = \mathbf{0}$ and $\mathbf{C}_k \mathbf{C}_\perp^T = \mathbf{0}$ follows

(2.7) $$\mathbf{A}\mathbf{A}^T = \mathbf{A}_k \mathbf{A}_k^T + \mathbf{A}_\perp \mathbf{A}_\perp^T, \qquad \mathbf{C}\mathbf{C}^T = \mathbf{C}_k \mathbf{C}_k^T + \mathbf{C}_\perp \mathbf{C}_\perp^T.$$

Since the relevant matrices are symmetric positive semidefinite, eigenvalues are equal to singular values. The dominant ones are

$$\sigma_j(\mathbf{A}_k \mathbf{A}_k^T) = \sigma_j(\mathbf{A}\mathbf{A}^T) = \sigma_j(\mathbf{A})^2, \quad \sigma_j(\mathbf{C}_k \mathbf{C}_k^T) = \sigma_j(\mathbf{C}\mathbf{C}^T) = \sigma_j(\mathbf{C})^2, \quad 1 \leq j \leq k,$$

and the subdominant ones are, with $j \geq 1$,

$$\sigma_j(\mathbf{A}_\perp \mathbf{A}_\perp^T) = \sigma_{k+j}(\mathbf{A}\mathbf{A}^T) = \sigma_{k+j}(\mathbf{A})^2, \quad \sigma_j(\mathbf{C}_\perp \mathbf{C}_\perp^T) = \sigma_{k+j}(\mathbf{C}\mathbf{C}^T) = \sigma_{k+j}(\mathbf{C})^2.$$

To apply Theorem 2.5, set $\widehat{\mathbf{A}} = \mathbf{C}$, $\mathbf{P} = \mathbf{C}_k \mathbf{C}_k^\dagger$. Then $\text{rank}(\mathbf{P}) = \text{rank}(\mathbf{C}_k) = k$ and

$$(\mathbf{I} - \mathbf{P})\,\widehat{\mathbf{A}} = (\mathbf{I} - \mathbf{C}_k \mathbf{C}_k^\dagger)\,(\mathbf{C}_k + \mathbf{C}_\perp) = \mathbf{C}_\perp.$$

Thus

(2.8) $$\|\|(\mathbf{I} - \mathbf{C}_k \mathbf{C}_k^\dagger)\mathbf{C}\|\|_p = \|\|(\mathbf{I} - \mathbf{P})\,\widehat{\mathbf{A}}\|\|_p^2 = \|\|\mathbf{C}_\perp\|\|_p^2.$$

Two-norm. Substituting (2.8) into the two-norm bound in Theorem 2.5 gives

(2.9) $$\|(\mathbf{I} - \mathbf{C}_k \mathbf{C}_k^\dagger)\,\mathbf{A}\|_2^2 \leq \|\mathbf{C}_\perp\|_2^2 + \|\mathbf{A}\mathbf{A}^T - \mathbf{C}\mathbf{C}^T\|_2.$$

With the notation in (2.7), add and subtract $\sigma_{k+1}(\mathbf{A}\mathbf{A}^T) = \|\mathbf{A}_\perp\|_2^2 = \|\mathbf{A} - \mathbf{A}_k\|_2^2$, and then apply Weyl's theorem,

$$\|\mathbf{C}_\perp\|_2^2 = \|\mathbf{C}_\perp \mathbf{C}_\perp^T\|_2 = \sigma_{k+1}(\mathbf{C}\mathbf{C}^T)$$
$$\leq \left| \sigma_{k+1}(\mathbf{C}\mathbf{C}^T) - \sigma_{k+1}(\mathbf{A}\mathbf{A}^T) \right| + \|\mathbf{A}_\perp\|_2^2$$
$$\leq \|\mathbf{A}\mathbf{A}^T - \mathbf{C}\mathbf{C}^T\|_2 + \|\mathbf{A} - \mathbf{A}_k\|_2^2.$$

Substituting this into (2.9) gives

$$\|(\mathbf{I} - \mathbf{C}_k \mathbf{C}_k^\dagger)\,\mathbf{A}\|_2^2 \leq \|\mathbf{A} - \mathbf{A}_k\|_2^2 + 2\,\|\mathbf{A}\mathbf{A}^T - \mathbf{C}\mathbf{C}^T\|_2.$$

Schatten p-norm (p even). Substituting (2.8) into the Schatten p-norm bound in Theorem 2.5 gives

(2.10)
$$\||(\mathbf{I} - \mathbf{C}_k\mathbf{C}_k^\dagger)\,\mathbf{A}\||_p^2 \leq \|||\,\mathbf{C}_\perp\,\||_p^2$$
$$+ \min\left\{ \,\||\,\mathbf{AA}^T - \mathbf{CC}^T\||_{p/2},\ \sqrt[p]{m-k}\,\||\,\mathbf{AA}^T - \mathbf{CC}^T\,\||_p \,\right\}.$$

From (2.5) follows $\||\mathbf{C}_\perp\||_p^2 = \||\mathbf{C}_\perp\mathbf{C}_\perp^T\||_{p/2}$. For a column vector $\mathbf{x}$, let

$$\|\mathbf{x}\|_p = \sqrt[p]{\sum_j |x_j|^{1/p}}$$

be the ordinary vector p-norm, and put the singular values of $\mathbf{C}_\perp\mathbf{C}_\perp^T$ into the vector

$$\mathbf{c}_\perp \equiv \left(\sigma_1(\mathbf{C}_\perp\mathbf{C}_\perp^T) \quad \cdots \quad \sigma_{m-k}(\mathbf{C}_\perp\mathbf{C}_\perp^T)\right)^T.$$

Move from matrix norm to vector norm

$$\||\mathbf{C}_\perp\mathbf{C}_\perp^T\||_{p/2}^{p/2} = \sum_{j=1}^{m-k} \sigma_j(\mathbf{C}_\perp\mathbf{C}_\perp^T)^{p/2} = \sum_{j=1}^{m-k} (\mathbf{c}_\perp)_j^{p/2} = \||\mathbf{c}_\perp\||_{p/2}^{p/2}.$$

Put the singular values of $\mathbf{A}_\perp\mathbf{A}_\perp^T$ into the vector

$$\mathbf{a}_\perp \equiv \left(\sigma_1(\mathbf{A}_\perp\mathbf{A}_\perp^T) \quad \cdots \quad \sigma_{m-k}(\mathbf{A}_\perp\mathbf{A}_\perp^T)\right)^T,$$

and apply the triangle inequality in the vector norm

$$\||\mathbf{C}_\perp\mathbf{C}_\perp^T\||_{p/2} = \||\mathbf{c}_\perp\||_{p/2} \leq \||\mathbf{c}_\perp - \mathbf{a}_\perp\||_{p/2} + \||\,\mathbf{a}_\perp\,\||_{p/2}.$$

Substituting the following expression

$$\||\mathbf{a}_\perp\||_{p/2}^{p/2} = \sum_{j=1}^{m-k} \sigma_j(\mathbf{A}_\perp\mathbf{A}_\perp^T)^{p/2} = \sum_{j=1}^{m-k} \sigma_j(\mathbf{A}_\perp)^p = \||\mathbf{A}_\perp\||_p^p$$

into the previous bound and applying (2.5) again gives

(2.11)
$$\||\,\mathbf{C}_\perp\||_p^2 = \||\mathbf{C}_\perp\mathbf{C}_\perp^T\||_{p/2} \leq \||\mathbf{c}_\perp - \mathbf{a}_\perp\||_{p/2} + \||\,\mathbf{A}_\perp\,\||_p^2.$$

1. Derivation of the first term in the minimum in (2.10):
   Apply Mirsky's Theorem (2.6) to the first summand in (2.11):

$$\||\mathbf{c}_\perp - \mathbf{a}_\perp\||_{p/2}^{p/2} = \sum_{j=1}^{m-k} \left|\sigma_{k+j}(\mathbf{CC}^T) - \sigma_{k+j}(\mathbf{AA}^T)\right|^{p/2}$$

$$\leq \sum_{j=1}^m \left|\sigma_j(\mathbf{CC}^T) - \sigma_j(\mathbf{AA}^T)\right|^{p/2} \leq \||\mathbf{CC}^T - \mathbf{AA}^T\||_{p/2}^{p/2}.$$

Take the $p/2$ square root on both sides,

$$\||\mathbf{c}_\perp - \mathbf{a}_\perp\||_{p/2} \leq \||\mathbf{CC}^T - \mathbf{AA}^T\||_{p/2},$$

and substitute this into (2.11), so that

$$\||\mathbf{C}_\perp\||_p^2 \leq \||\mathbf{A}_\perp\||_p^2 + \||\ \mathbf{C}\mathbf{C}^T - \mathbf{A}\mathbf{A}^T\ \||_{p/2} .$$

The above, in turn, substitute into (2.10) to obtain the first term in the minimum,

$$\||(\mathbf{I} - \mathbf{C}_k\mathbf{C}_k^\dagger)\,\mathbf{A}\||_p^2 \leq \||\mathbf{A}_\perp\||_p^2 + 2\ \||\ \mathbf{C}\mathbf{C}^T - \mathbf{A}\mathbf{A}^T\ \||_{p/2} .$$

2. Derivation of the second term in the minimum in (2.10):
   Consider the first summand in (2.11), but apply the Cauchy–Schwarz inequality before Mirsky's Theorem (2.6):

$$\||\mathbf{c}_\perp - \mathbf{a}_\perp\||_{p/2}^{p/2} = \sum_{j=1}^{m-k} \left|\sigma_{k+j}(\mathbf{C}\mathbf{C}^T) - \sigma_{k+j}(\mathbf{A}\mathbf{A}^T)\right|^{p/2}$$

$$\leq \sqrt{m-k}\ \sqrt{\sum_{j=1}^{m-k} |\sigma_{k+j}(\mathbf{C}\mathbf{C}^T) - \sigma_{k+j}(\mathbf{A}\mathbf{A}^T)|^p}$$

$$\leq \sqrt{m-k}\ \sqrt{\sum_{j=1}^{m} |\sigma_{j}(\mathbf{C}\mathbf{C}^T) - \sigma_{j}(\mathbf{A}\mathbf{A}^T)|^p}$$

$$\leq \sqrt{m-k}\ \||\ \mathbf{C}\mathbf{C}^T - \mathbf{A}\mathbf{A}^T\ \||_p^{p/2} .$$

Take the $p/2$ square root on both sides,

$$\||\mathbf{c}_\perp - \mathbf{a}_\perp\||_{p/2} \leq \sqrt[p]{m-k}\ \||\ \mathbf{C}\mathbf{C}^T - \mathbf{A}\mathbf{A}^T\||_p,$$

and substitute this into (2.11), so that

$$\||\mathbf{C}_\perp\||_p^2 \leq \||\mathbf{A}_\perp\||_p^2 + \sqrt[p]{m-k}\ \||\ \mathbf{C}\mathbf{C}^T - \mathbf{A}\mathbf{A}^T\ \||_p .$$

The above, in turn, substitute into (2.10) to obtain the second term in the minimum,

$$\||(\mathbf{I} - \mathbf{C}_k\mathbf{C}_k^\dagger)\,\mathbf{A}\||_p^2 \leq \||\mathbf{A}_\perp\||_p^2 + 2\sqrt[p]{m-k}\ \||\ \mathbf{C}\mathbf{C}^T - \mathbf{A}\mathbf{A}^T\ \||_p .$$

3. *Frobenius norm.* This is the special case $p = 2$ with $\||\mathbf{A}\||_2 = \|\mathbf{A}\|_F$ and $\||\mathbf{A}\||_1 = \|\mathbf{A}\|_*$. □

3. Approximation errors and angles between subspaces. We consider approximations where the rank of the orthogonal projector is at least as large as the dimension of the dominant subspace, and relate the low-rank approximation error to the subspace angle between projector and target space. After reviewing assumptions and notation (section 3.1), we bound the low-rank approximation error in terms of the subspace angle from below (section 3.2) and from above (section 3.3).

3.1. Assumptions. Given $\mathbf{A} \in \mathbb{R}^{m \times n}$ with a gap after the kth singular value,

$$\|\mathbf{A}\|_2 = \sigma_1(\mathbf{A}) \geq \cdots \geq \sigma_k(\mathbf{A}) > \sigma_{k+1}(\mathbf{A}) \geq \cdots \geq \sigma_r(\mathbf{A}) \geq 0, \qquad r \equiv \min\{m, n\}.$$

The gap assures that the k-dimensional dominant subspace is well posed. Partition the full SVD $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ in section 1.1

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_k & \mathbf{U}_\perp \end{pmatrix}, \qquad \mathbf{V} = \begin{pmatrix} \mathbf{V}_k & \mathbf{V}_\perp \end{pmatrix}, \qquad \mathbf{\Sigma} = \mathrm{diag}\begin{pmatrix} \mathbf{\Sigma}_k & \mathbf{\Sigma}_\perp \end{pmatrix},$$

where the dominant parts are

$$\boldsymbol{\Sigma}_k \equiv \mathrm{diag}\begin{pmatrix} \sigma_1(\mathbf{A}) & \cdots & \sigma_k(\mathbf{A}) \end{pmatrix} \in \mathbb{R}^{k \times k}, \qquad \mathbf{U}_k \in \mathbb{R}^{m \times k}, \qquad \mathbf{V}_k \in \mathbb{R}^{n \times k},$$

and the subdominant ones

$$\boldsymbol{\Sigma}_\perp \in \mathbb{R}^{(m-k) \times (n-k)}, \qquad \mathbf{U}_\perp \in \mathbb{R}^{m \times (m-k)}, \qquad \mathbf{V}_\perp \in \mathbb{R}^{n \times (n-k)}.$$

Thus $\mathbf{A}$ is a "direct sum"

$$\mathbf{A} = \mathbf{A}_k + \mathbf{A}_\perp, \qquad \text{where} \qquad \mathbf{A}_k \equiv \mathbf{U}_k \boldsymbol{\Sigma}_k \mathbf{V}_k^T, \quad \mathbf{A}_\perp \equiv \mathbf{U}_\perp \boldsymbol{\Sigma}_\perp \mathbf{V}_\perp$$

and

$$(3.1) \qquad\qquad\qquad \mathbf{A}_\perp \mathbf{A}_k^\dagger = \mathbf{0} = \mathbf{A}_\perp \mathbf{A}_k^T.$$

The goal is to approximate the k-dimensional dominant left singular vector space,

$$(3.2) \qquad\qquad\qquad \mathbf{P}_k \equiv \mathbf{U}_k \mathbf{U}_k^T = \mathbf{A}_k \mathbf{A}_k^\dagger.$$

To this end, let $\mathbf{P} \in \mathbb{R}^{m \times m}$ be an orthogonal projector as in (1.3), whose rank is at least as large as the dimension of the targeted subspace,

$$\mathrm{rank}(\mathbf{P}) \geq \mathrm{rank}(\mathbf{P}_k).$$

3.2. Subspace angle as a lower bound for the approximation error. We bound the low-rank approximation error from below by the subspace angle and the kth singular value of $\mathbf{A}$, in the two-norm and the Frobenius norm.

THEOREM 3.1. *With the assumptions in section* 3.1, *let $p \geq 1$ be an integer. Then*

$$\||(\mathbf{I} - \mathbf{P})\mathbf{A}\||_p \geq \sigma_k(\mathbf{A}) \,\||\sin\boldsymbol{\Theta}(\mathbf{P},\mathbf{P}_k)\||_p.$$

Proof. From Lemma 1.5, (3.2), (3.1) and Lemma 1.3 follows

$$\begin{aligned}
\||\sin\boldsymbol{\Theta}(\mathbf{P},\mathbf{P}_k)\||_p &= \||(\mathbf{I}-\mathbf{P})\mathbf{P}_k\||_p = \||(\mathbf{I}-\mathbf{P})\mathbf{A}_k\mathbf{A}_k^\dagger\||_p \\
&= \||(\mathbf{I}-\mathbf{P})(\mathbf{A}_k+\mathbf{A}_\perp)\mathbf{A}_k^\dagger\||_p = \||(\mathbf{I}-\mathbf{P})\mathbf{A}\mathbf{A}_k^\dagger\||_p \\
&\leq \|\mathbf{A}_k^\dagger\|_2 \,\||(\mathbf{I}-\mathbf{P})\mathbf{A}\||_p = \||(\mathbf{I}-\mathbf{P})\mathbf{A}\||_p / \sigma_k(\mathbf{A}).
\end{aligned}$$

3.3. Subspace angle as upper bound for the approximation error. We present upper bounds for the low-rank approximation error in terms of the subspace angle, the two-norm (Theorem 3.2), and Frobenius norm (Theorem 3.3).

The bounds are guided by the following observation. In the ideal case, where $\mathbf{P}$ completely captures the target space, we have $\mathrm{range}(\mathbf{P}) = \mathrm{range}(\mathbf{P}_k) = \mathrm{range}(\mathbf{A}_k)$, and

$$\|\sin\boldsymbol{\Theta}(\mathbf{P},\mathbf{P}_k)\|_{2,F} = 0, \qquad \|(\mathbf{I}-\mathbf{P})\mathbf{A}\|_{2,F} = \|\mathbf{A}_\perp\|_{2,F} = \|\boldsymbol{\Sigma}_\perp\|_{2,F},$$

thus suggesting an additive error in the general, nonideal case.

THEOREM 3.2 (two-norm). *With the assumptions in section* 3.1,

$$\|(\mathbf{I}-\mathbf{P})\mathbf{A}\|_2 \leq \|\mathbf{A}\|_2 \,\|\sin\boldsymbol{\Theta}(\mathbf{P},\mathbf{P}_k)\|_2 + \|\mathbf{A}-\mathbf{A}_k\|_2 \,\|\cos\boldsymbol{\Theta}(\mathbf{I}-\mathbf{P},\mathbf{I}-\mathbf{P}_k)\|_2.$$

If also $k \leq \mathrm{rank}(\mathbf{P}) < m - k$, then $\|\cos\boldsymbol{\Theta}(\mathbf{I}-\mathbf{P},\mathbf{I}-\mathbf{P}_k)\|_2 = 1$.

Proof. From $\mathbf{A} = \mathbf{A}_k + \mathbf{A}_\perp$ and the triangle inequality follows

(3.3) $\|(\mathbf{I} - \mathbf{P})\mathbf{A}\|_2 \leq \|(\mathbf{I} - \mathbf{P})\mathbf{A}_k\|_2 + \|(\mathbf{I} - \mathbf{P})\mathbf{A}_\perp\|_2.$

- Bound for the first summand in (3.3):
  Since $\mathrm{rank}(\mathbf{P}) \geq \mathrm{rank}(\mathbf{P}_k)$, Lemma 1.5 implies

$$\|(\mathbf{I} - \mathbf{P})\mathbf{A}_k\|_2 \leq \|(\mathbf{I} - \mathbf{P})\,\mathbf{U}_k\|_2 \|\mathbf{\Sigma}_k\|_2 = \|\mathbf{A}\|_2\,\|(\mathbf{I} - \mathbf{P})\mathbf{P}_k\|_2$$
$$= \|\mathbf{A}\|_2\,\|\sin\mathbf{\Theta}(\mathbf{P}, \mathbf{P}_k)\|_2.$$

Substitute this into (3.3),

(3.4) $\|(\mathbf{I} - \mathbf{P})\mathbf{A}\|_2 \leq \|\mathbf{A}\|_2\,\|\sin\mathbf{\Theta}(\mathbf{P}, \mathbf{P}_k)\|_2 + \|(\mathbf{I} - \mathbf{P})\mathbf{A}_\perp\|_2.$

- Bound for the second summand in (3.3):
  Submultiplicativity implies

$$\|(\mathbf{I} - \mathbf{P})\mathbf{A}_\perp\|_2 \leq \|(\mathbf{I} - \mathbf{P})\,\mathbf{U}_\perp\|_2\,\|\mathbf{\Sigma}_\perp\|_2 = \|\mathbf{A} - \mathbf{A}_k\|_2\,\|(\mathbf{I} - \mathbf{P})\,\mathbf{U}_\perp\|_2.$$

For the last factor, apply the full SVD of $\mathbf{A}$ in section 3.1,

$$\mathrm{range}(\mathbf{U}_\perp) = \mathrm{range}(\mathbf{U}_\perp\mathbf{U}_\perp^T) = \mathrm{range}(\mathbf{U}_k\mathbf{U}_k^T)^\perp = \mathrm{range}(\mathbf{P}_k)^\perp$$
$$= \mathrm{range}(\mathbf{I} - \mathbf{P}_k)$$

so that

$$\|(\mathbf{I} - \mathbf{P})\,\mathbf{U}_\perp\|_2 = \|(\mathbf{I} - \mathbf{P})\,(\mathbf{I} - \mathbf{P}_k)\|_2 = \|\cos\mathbf{\Theta}(\mathbf{I} - \mathbf{P}, \mathbf{I} - \mathbf{P}_k)\|_2.$$

Thus,
$$\|(\mathbf{I} - \mathbf{P})\mathbf{A}_\perp\|_2 \leq \|\mathbf{A} - \mathbf{A}_k\|_2\,\|\cos\mathbf{\Theta}(\mathbf{I} - \mathbf{P}, \mathbf{I} - \mathbf{P}_k)\|_2.$$

Substitute this into (3.4) to obtain the first bound.
- Special case $k \leq \mathrm{rank}(\mathbf{P}) < m - k$:
  Setting $\mathbf{I} - \mathbf{P}_k = \mathbf{Z}_\perp\mathbf{Z}_\perp^T$ and $\mathbf{I} - \mathbf{P} = \widehat{\mathbf{Z}}_\perp\widehat{\mathbf{Z}}_\perp^T$ in Corollary A.2 implies
  $\|\cos\mathbf{\Theta}(\mathbf{I} - \mathbf{P}, \mathbf{I} - \mathbf{P}_k)\|_2 = 1.$ ☐

THEOREM 3.3 (Schatten p-norm). *With the assumptions in section* 3.1, *let* $p \geq 1$
be an integer and $\mathbf{\Gamma} \equiv \cos\mathbf{\Theta}(\mathbf{I} - \mathbf{P}, \mathbf{I} - \mathbf{P}_k)$. *Then*

$$\||(\mathbf{I} - \mathbf{P})\mathbf{A}\||_p \leq \|\mathbf{A}\|_2\,\||\sin\mathbf{\Theta}(\mathbf{P}, \mathbf{P}_k)\,\||_p$$
$$+ \min\left\{\|\mathbf{A} - \mathbf{A}_k\|_2\,\||\,\mathbf{\Gamma}\||_p,\; \||\,\mathbf{A} - \mathbf{A}_k\,\||_p\,\|\mathbf{\Gamma}\|_2\right\}.$$

If also $k \leq \mathrm{rank}(\mathbf{P}) < m - k$, *then*

$$\||(\mathbf{I} - \mathbf{P})\mathbf{A}\||_p \leq \|\mathbf{A}\|_2\,\||\sin\mathbf{\Theta}(\mathbf{P}, \mathbf{P}_k)\,\||_p + \||\,\mathbf{A} - \mathbf{A}_k\,\||_p.$$

Proof. With Lemma 1.3, the analogue of (3.4) is

(3.5) $\||(\mathbf{I} - \mathbf{P})\mathbf{A}\||_p \leq \|\mathbf{A}\|_2\,\||\sin\mathbf{\Theta}(\mathbf{P}, \mathbf{P}_k)\,\||_p + \||\,(\mathbf{I} - \mathbf{P})\mathbf{A}_\perp\,\||_p.$

There are two options to bound $\||(\mathbf{I} - \mathbf{P})\mathbf{A}_\perp\||_p = \||(\mathbf{I} - \mathbf{P})\,\mathbf{U}_\perp\mathbf{\Sigma}_\perp\||_p$, depending on
which factor gets the two-norm. Either

$$\||(\mathbf{I} - \mathbf{P})\,\mathbf{U}_\perp\mathbf{\Sigma}_\perp\||_p \leq \||(\mathbf{I} - \mathbf{P})\,\mathbf{U}_\perp\,\||_p\,\|\mathbf{\Sigma}_\perp\|_2 = \|\mathbf{A} - \mathbf{A}_k\|_2\,\||\,(\mathbf{I} - \mathbf{P})\,\mathbf{U}_\perp\||_p,$$

or

$$\||(\mathbf{I} - \mathbf{P})\mathbf{U}_\perp \mathbf{\Sigma}_\perp\||_p \le \||(\mathbf{I} - \mathbf{P})\mathbf{U}_\perp\||_2 \; \||\mathbf{\Sigma}_\perp\||_p = \||\mathbf{A} - \mathbf{A}_k\||_p \; \||(\mathbf{I} - \mathbf{P})\mathbf{U}_\perp\||_2.$$

As in the proof of Theorem 3.2 one shows

$$\||(\mathbf{I} - \mathbf{P})\mathbf{U}_\perp\||_p = \||\cos\mathbf{\Theta}(\mathbf{I} - \mathbf{P}, \mathbf{I} - \mathbf{P}_k)\||_p,$$

as well as the expression for the special case $k \le \operatorname{rank}(\mathbf{P}) < m - k$. $\qquad\square$

Appendix A. CS decompositions. We review expressions for the CS decompositions from [20, Theorem 8.1] and [29, section 2].

Consider two subspaces $\operatorname{range}(\mathbf{Z})$ and $\operatorname{range}(\widehat{\mathbf{Z}})$ whose dimensions sum up to less than the dimension of the host space. Specifically, let $(\mathbf{Z} \quad \mathbf{Z}_\perp), (\widehat{\mathbf{Z}} \quad \widehat{\mathbf{Z}}_\perp) \in \mathbb{R}^{m \times m}$ be orthogonal matrices where $\mathbf{Z} \in \mathbb{R}^{m \times k}$ and $\widehat{\mathbf{Z}} \in \mathbb{R}^{m \times \ell}$. The CS decomposition of the cross product is

$$(\mathbf{Z} \quad \mathbf{Z}_\perp)^T (\widehat{\mathbf{Z}} \quad \widehat{\mathbf{Z}}_\perp) = \begin{pmatrix} \mathbf{Z}^T\widehat{\mathbf{Z}} & \mathbf{Z}^T\widehat{\mathbf{Z}}_\perp \\ \mathbf{Z}_\perp^T\widehat{\mathbf{Z}} & \mathbf{Z}_\perp^T\widehat{\mathbf{Z}}_\perp \end{pmatrix} = \begin{pmatrix} \mathbf{Q}_{11} & \\ & \mathbf{Q}_{12} \end{pmatrix} \mathbf{D} \begin{pmatrix} \mathbf{Q}_{21} & \\ & \mathbf{Q}_{22} \end{pmatrix},$$

where $\mathbf{Q}_{11} \in \mathbb{R}^{k \times k}$, $\mathbf{Q}_{12} \in \mathbb{R}^{(m-k) \times (m-k)}$, $\mathbf{Q}_{21} \in \mathbb{R}^{\ell \times \ell}$, and $\mathbf{Q}_{22} \in \mathbb{R}^{(m-\ell) \times (m-\ell)}$ are all orthogonal matrices.

THEOREM A.1. *If $k \le \ell < m - k$, then*

$$\mathbf{D} = \left[\begin{array}{cccc|ccc} \mathbf{I}_r & & & & \mathbf{0} & & \\ & \mathbf{C} & & & & \mathbf{S} & \\ & & \mathbf{0} & & & & \mathbf{I}_{k-(r+s)} \\ \hline \mathbf{0} & & & & -\mathbf{I}_{m-(k+\ell)+r} & & \\ & \mathbf{S} & & & & -\mathbf{C} & \\ & & \mathbf{I}_{\ell-(r+s)} & & & & \mathbf{0} \end{array}\right] \begin{array}{l} r \\ s \\ k-(s+r) \\ m-(k+\ell)+r \\ s \\ \ell-(r+s) \end{array}$$

with column labels $r \quad s \quad \ell-(r+s) \quad m-(k+\ell)+r \quad s \quad k-(r+s)$.

Here $\mathbf{C}^2 + \mathbf{S}^2 = \mathbf{I}_s$ with

$$\mathbf{C} = \operatorname{diag}\begin{pmatrix} \cos\theta_1 & \cdots & \cos\theta_s \end{pmatrix}, \quad \mathbf{S} = \operatorname{diag}\begin{pmatrix} \sin\theta_1 & \cdots & \sin\theta_s \end{pmatrix},$$

and

$$r = \dim\left(\operatorname{range}(\mathbf{Z}) \cap \operatorname{range}(\widehat{\mathbf{Z}})\right), \quad m - (k+\ell) + r = \dim\left(\operatorname{range}(\mathbf{Z}_\perp) \cap \operatorname{range}(\widehat{\mathbf{Z}}_\perp)\right),$$

$$\ell - (r+s) = \dim\left(\operatorname{range}(\mathbf{Z}_\perp) \cap \operatorname{range}(\widehat{\mathbf{Z}})\right), \quad k - (r+s) = \dim\left(\operatorname{range}(\mathbf{Z}) \cap \operatorname{range}(\widehat{\mathbf{Z}}_\perp)\right).$$

COROLLARY A.2. *From Theorem A.1 follows*

$$\|\sin\mathbf{\Theta}(\mathbf{Z}, \widehat{\mathbf{Z}})\|_{2,F} = \|\mathbf{Z}^T\widehat{\mathbf{Z}}_\perp\|_{2,F} = \left\| \begin{pmatrix} \mathbf{S} & \\ & \mathbf{I}_{k-(r+s)} \end{pmatrix} \right\|_{2,F},$$

$$\|\cos\mathbf{\Theta}(\mathbf{Z}, \widehat{\mathbf{Z}})\|_{2,F} = \|\mathbf{Z}^T\widehat{\mathbf{Z}}\|_{2,F} = \left\| \begin{pmatrix} \mathbf{I}_r & \\ & \mathbf{C} \end{pmatrix} \right\|_{2,F},$$

$$\|\cos\mathbf{\Theta}(\mathbf{Z}_\perp, \widehat{\mathbf{Z}}_\perp)\|_{2,F} = \|\mathbf{Z}_\perp^T\widehat{\mathbf{Z}}_\perp\|_{2,F} = \left\| \begin{pmatrix} \mathbf{I}_{m-(k+\ell)+r} & \\ & \cos\mathbf{\Theta}(\mathbf{Z}, \widehat{\mathbf{Z}}) \end{pmatrix} \right\|_{2,F}.$$

Acknowledgments. We are grateful to Daniel Kressner for his suggestion to simplify the proof of Corollary 2.4 and thereby extend it to all p-norms.

REFERENCES

[1] K. ABHISEK AND P. DRINEAS, *A note on randomized element-wise matrix sparsification*, preprint, arXiv:1404.0320 [cs.IT].

[2] D. ACHLIOPTAS AND F. MCSHERRY, *Fast computation of low-rank matrix approximations*, J. ACM, 54 (2007), 9.

[3] R. BHATIA, *Matrix analysis*, Graduate Texts in Mathematics, vol. 169, Springer-Verlag, New York, 1997.

[4] CH. DAVIS AND W. M. KAHAN, *Some new bounds on perturbation of subspaces*, Bull. Amer. Math. Soc., 75 (1969), pp. 863–868.

[5] C. DAVIS AND W. M. KAHAN, *The rotation of eigenvectors by a perturbation,* III, SIAM J. Numer. Anal., 7 (1970), pp. 1–46.

[6] P. DRINEAS, I. C. F. IPSEN, E.-M. KONTOPOULO, AND M. MAGDON-ISMAIL, *Structural convergence results for low-rank approximations from block Krylov spaces*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 567–586.

[7] P. DRINEAS, R. KANNAN, AND M. W. MAHONEY, *Fast Monte Carlo algorithms for matrices* I: *Approximating matrix multiplication*, SIAM J. Comput., 36 (2006), pp. 132–157.

[8] P. DRINEAS, R. KANNAN, AND M. W. MAHONEY, *Fast Monte Carlo algorithms for matrices.* II. *Computing a low-rank approximation to a matrix*, SIAM J. Comput., 36 (2006), pp. 158–183.

[9] P. DRINEAS AND M. W. MAHONEY, *RandNLA: Randomized numerical linear algebra*, Commun. ACM, 59 (2016), pp. 80–90.

[10] P. DRINEAS AND A. ZOUZIAS, *A note on element-wise matrix sparsification via a matrix-valued Bernstein inequality*, Inform. Process. Lett., 111 (2011), pp. 385–389.

[11] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, 4th ed., The Johns Hopkins University Press, Baltimore, 2013.

[12] M. GU, *Subspace iteration randomization and singular value problems*, SIAM J. Sci. Comput., 37 (2015), pp. A1139–A1173.

[13] W. HACKBUSCH, *New estimates for the recursive low-rank truncation of block-structured matrices*, Numer. Math., 132 (2016), pp. 303–328.

[14] N. J. HIGHAM, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, 2002.

[15] J. T. HOLODNAK AND I. C. F. IPSEN, *Randomized approximation of the Gram matrix: Exact computation and probabilistic bounds*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 110–137.

[16] R. A. HORN AND C. R. JOHNSON, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

[17] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, 2nd ed., Cambridge University Press, Cambridge, 2013.

[18] I. C. F. IPSEN, *An overview of relative* $\sin \Theta$ *theorems for invariant subspaces of complex matrices*, J. Comput. Appl. Math., 123 (2000), pp. 131–153, Invited paper for the special issue *Numerical Analysis* 2000, Vol. III, Linear Algebra.

[19] C. MUSCO AND C. MUSCO, *Randomized block krylov methods for stronger and faster approximate singular value decomposition*, in Advances in Neural Information Processing Systems 28, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, eds., Curran Associates, Red Hook, NY, 2015, pp. 1396–1404.

[20] C. C. PAIGE AND M. WEI, *History and generality of the* CS *decomposition*, Linear Algebra Appl., 208/209 (1994), pp. 303–326.

[21] Y. SAAD, *On the rates of convergence of the Lanczos and the block-Lanczos methods*, SIAM J. Numer. Anal., 17 (1980), pp. 687–706.

[22] Y. SAAD, *Numerical Methods for Large Eigenvalue Problems*, revised ed., Classics in Applied Mathematics, SIAM, Philadelphia, 2011.

[23] G. W. STEWART, *Error and perturbation bounds for subspaces associated with certain eigenvalue problems*, SIAM Rev., 15 (1973), pp. 727–64.

[24] G. W. STEWART, *On the numerical analysis of oblique projectors*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 309–348.

[25] G. W. STEWART AND J.-G. SUN, *Matrix Perturbation Theory*, Academic Press, San Diego, 1990.

[26] P. Å. WEDIN, *Perturbation bounds in connection with singular value decomposition*, Nordisk Tidskr. Informationsbehandling (BIT), 12 (1972), pp. 99–111.

[27] P. Å. WEDIN, *On angles between subspaces of a finite-dimensional inner product space*, in Matrix Pencils, B. Kågström and A. Ruhe, eds., Lecture Notes in Mathematics 973, Springer-Verlag, New York, 1983, pp. 263–285.

[28] D. P. WOODRUFF, *Sketching as a tool for numerical linear algebra*, Found. Trends Theor. Comput. Sci., 10 (2014), pp. 1–157.

[29] P. ZHU AND A. V. KNYAZEV, *Angles between subspaces and their tangents*, J. Numer. Math., 21 (2013), pp. 325–340.