CS395T: Continuous Algorithms, Part XI
Low-rank approximation

Kevin Tian

1 Principal component analysis

In this lecture, we introduce the topic of *low-rank approximation*. Broadly speaking, the goal of low-rank approximation is to approximate a matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$, potentially of rank as large as $\min(n, d)$, as well as possible with a matrix $\widetilde{\mathbf{A}}$ of much smaller rank.

Because of the Eckart-Young-Mirsky theorem (Theorem 1), this is often conflated with the goal of *principal component analysis*. In order to state this result, we first require the following definition:

$$\mathcal{U}_k := \left\{ \mathbf{U} \in \mathbb{R}^{d \times k} \mid \mathbf{U}^\top \mathbf{U} = \mathbf{I}_k \right\}. \tag{1}$$

In other words, $\mathcal{U}_k$ is the set of $d \times k$ matrices with orthonormal columns. We can now define the k-principal component analysis (k-PCA) problem, which asks to return $\mathbf{V} \in \mathcal{U}_k$ satisfying

$$\left\langle \mathbf{V}\mathbf{V}^\top, \mathbf{M} \right\rangle = \max_{\mathbf{U} \in \mathcal{U}_k} \left\langle \mathbf{U}\mathbf{U}^\top, \mathbf{M} \right\rangle, \tag{2}$$

where $\mathbf{M} \in \mathbb{S}^{d \times d}_{\succeq \mathbf{0}}$. It is a well-known fact in numerical linear algebra that the optimal solution to (2) is a basis for any eigenspace corresponding to the k largest eigenvalues of $\mathbf{M}$.

Lemma 1. *Let $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top = \sum_{i \in [d]} \boldsymbol{\lambda}_i \mathbf{u}_i \mathbf{u}_i^\top$ be the eigendecomposition of $\mathbf{M} \in \mathbb{S}^{d \times d}_{\succeq \mathbf{0}}$, where the $\{\boldsymbol{\lambda}_i\}_{i \in [d]}$ are nonincreasing, let $[m] \subseteq [d]$ correspond to indices $i \in [d]$ where $\boldsymbol{\lambda}_i \geq \boldsymbol{\lambda}_k$, and let $[\ell] \subseteq [m]$ correspond to indices $i \in [d]$ where $\boldsymbol{\lambda}_i > \boldsymbol{\lambda}_k$. Then $\mathbf{V} \in \mathcal{U}_k$ solves (2) optimally iff*

$$\text{Span}\left(\{\mathbf{u}_i\}_{i \in [\ell]} \right) \subseteq \text{Span}\left(\mathbf{V} \right) \subseteq \text{Span}\left(\{\mathbf{u}_i\}_{i \in [m]} \right). \tag{3}$$

Proof. First, the von Neumann trace inequality (Theorem 6, Part VI) shows that

$$\max_{\mathbf{U} \in \mathcal{U}_k} \left\langle \mathbf{U}\mathbf{U}^\top, \mathbf{M} \right\rangle \leq \sum_{i \in [k]} \boldsymbol{\lambda}_i,$$

since $\mathbf{U}\mathbf{U}^\top$ has exactly k eigenvalues equal to 1, and the rest are 0. Moreover, examining the proof of Theorem 6, Part V implies that if we let $\{\mathbf{v}_i\}_{i \in [k]}$ denote the columns of $\mathbf{V}$, then the extremal value above is attained iff $|\langle \mathbf{u}_{\sigma(i)}, \mathbf{v}_i \rangle| = 1$ for all $i \in [k]$ and a permutation $\sigma : [d] \to [d]$ such that $\{\boldsymbol{\lambda}_{\sigma(i)}\}_{i \in [k]}$ has the same sum as $\{\boldsymbol{\lambda}_i\}_{i \in [k]}$, which is equivalent to the condition (3). □

Lemma 1 shows that solving (2) is computationally tractable (i.e., performable in polynomial time via eigendecomposition) for any $k \in [d]$. This is perhaps somewhat surprising, given that even the $k = 1$ case of (2) asks to maximize a convex function (i.e., $\mathbf{u}^\top \mathbf{M} \mathbf{u}$ for $\|\mathbf{u}\|_2 \leq 1$) over a convex set, which is a nonconvex optimization (indeed, a concave minimization) problem.

In fact, the following theorem due to Eckart-Young and Mirsky [EY36, Mir60], as alluded to earlier, shows that optimally performing k-PCA simultaneously solves a broad range of low-rank approximation problems beyond the quadratic form maximization problem in (2).

Theorem 1 (Eckart-Young-Mirsky)**.** *Let $\mathbf{A} \in \mathbb{R}^{n \times d}$ with $n \geq d$, and let $\|\cdot\|$ be a unitarily-invariant norm.*[1] *Then letting $\mathbf{V} \in \mathcal{U}_k$ attain the maximum value in (2) for $\mathbf{M} := \mathbf{A}^\top \mathbf{A}$, we have*

$$\left\| \mathbf{A} - \mathbf{A}\mathbf{V}\mathbf{V}^\top \right\| \leq \left\| \mathbf{A} - \widetilde{\mathbf{A}} \right\|, \text{ for all rank-}k\ \widetilde{\mathbf{A}} \in \mathbb{R}^{n \times d}. \tag{4}$$

[1] We defined unitarily-invariant norms in Part VI, Section 2.2.

Proof. We only prove the cases where $\|\cdot\| = \|\cdot\|_{\mathrm{op}}$ and $\|\cdot\| = \|\cdot\|_{\mathrm{F}}$ here, deferring a proof of the general case to [Mir60], which draws upon some of the analysis tools from Part VI.

Let $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ be the singular value decomposition of $\mathbf{A}$ (Corollary 3, Part VI), where $\mathbf{\Sigma} = \mathbf{diag}(\boldsymbol{\sigma})$ and we assume $\boldsymbol{\sigma}$ is in nonincreasing order. Moreover, let the columns of $\mathbf{U}$ be denoted $\{\mathbf{u}_i\}_{i\in[d]} \subset \mathbb{R}^n$, and similarly let $\{\mathbf{v}_i\}_{i\in[d]} \subset \mathbb{R}^d$ be the columns of $\mathbf{V}$. Observe that

$$\mathbf{M} = \mathbf{A}^\top\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top, \text{ where } \mathbf{\Lambda} = \mathbf{\Sigma}^2.$$

Thus, by Lemma 1, the optimal solution to (2) is given by the first k columns of $\mathbf{V}$, corresponding to the k largest singular values of $\mathbf{A}$ (breaking ties arbitrarily). Thus, our goal is to prove that

$$\mathbf{A}\left(\sum_{i\in[k]}\mathbf{v}_i\mathbf{v}_i^\top\right) = \left(\sum_{j\in[d]}\boldsymbol{\sigma}_j\mathbf{u}_j\mathbf{v}_j^\top\right)\left(\sum_{i\in[k]}\mathbf{v}_i\mathbf{v}_i^\top\right) = \sum_{i\in[k]}\boldsymbol{\sigma}_i\mathbf{u}_i\mathbf{v}_i^\top$$

is the optimal low-rank approximation to $\mathbf{A}$ in the sense of (4), when $\|\cdot\| \in \{\|\cdot\|_{\mathrm{op}}, \|\cdot\|_{\mathrm{F}}\}$.

For the operator norm, we proceed as follows. Let $\widetilde{\mathbf{A}} = \mathbf{X}\mathbf{Y}^\top$ be rank-k, with $\mathbf{X} \in \mathbb{R}^{n\times k}$ and $\mathbf{Y} \in \mathbb{R}^{d\times k}$. Let $\mathbf{V}_{[k+1]:}$ denote the first $k+1$ columns of $\mathbf{V}$. Because $\mathrm{Span}(\mathbf{Y})$ is k-dimensional, there must be some $\mathbf{v} \in \mathbb{R}^d$ in $\mathrm{Span}(\mathbf{V}_{[k+1]:})$ such that $\mathbf{Y}^\top\mathbf{v} = \mathbf{0}_k$. Without loss of generality, let $\|\mathbf{v}\|_2 = 1$, so that by orthonormality of the columns of $\mathbf{V}$, $\mathbf{v} = \mathbf{V}_{[k+1]:}\mathbf{w}$ where $\|\mathbf{w}\|_2 = 1$. Then,

$$\begin{aligned}\left\|\mathbf{A} - \widetilde{\mathbf{A}}\right\|_{\mathrm{op}} &\geq \left\|\left(\mathbf{A} - \widetilde{\mathbf{A}}\right)\mathbf{v}\right\|_2 = \|\mathbf{A}\mathbf{v}\|_2 \\ &= \left\|\mathbf{A}\mathbf{V}_{[k+1]:}\mathbf{w}\right\|_2 = \left\|\sum_{i\in[k+1]}\boldsymbol{\sigma}_i\mathbf{w}_i\mathbf{u}_i\right\|_2 = \sqrt{\sum_{i\in[k+1]}\boldsymbol{\sigma}_i^2\mathbf{w}_i^2} \geq \boldsymbol{\sigma}_{k+1},\end{aligned}$$

where the minimal value in the last inequality is achieved by $\mathbf{w}$ with all its mass on the $(k+1)^{\text{th}}$ coordinate. Finally, observe that $\|\mathbf{A} - \widetilde{\mathbf{A}}\|_{\mathrm{op}} = \boldsymbol{\sigma}_{k+1}$ is achieved by $\widetilde{\mathbf{A}} = \sum_{i\in[k]}\boldsymbol{\sigma}_i\mathbf{u}_i\mathbf{v}_i^\top$.

For the Frobenius norm, again let $\widetilde{\mathbf{A}} \in \mathbb{R}^{n\times d}$ be any rank-k matrix. Then for all $i \geq 1$,

$$\begin{aligned}\boldsymbol{\sigma}_i\left(\mathbf{A} - \widetilde{\mathbf{A}}\right) &= \boldsymbol{\sigma}_i\left(\mathbf{A} - \widetilde{\mathbf{A}}\right) + \boldsymbol{\sigma}_{k+1}\left(\widetilde{\mathbf{A}}\right) \\ &= \boldsymbol{\sigma}_1\left(\mathbf{A} - \widetilde{\mathbf{A}} - \mathbf{B}\right) + \boldsymbol{\sigma}_1\left(\widetilde{\mathbf{A}} - \widetilde{\mathbf{A}}\right) \\ &\geq \boldsymbol{\sigma}_1\left(\mathbf{A} - \left(\widetilde{\mathbf{A}} + \mathbf{B}\right)\right) \geq \boldsymbol{\sigma}_{i+k}\left(\mathbf{A}\right),\end{aligned}$$

for some rank-$(i-1)$ $\mathbf{B} \in \mathbb{R}^{n\times d}$, where the first inequality used that $\boldsymbol{\sigma}_1$ is the operator norm (which obeys the triangle inequality), and the second inequality used our earlier characterization of the operator norm and the fact that $\widetilde{\mathbf{A}} + \mathbf{B}$ is rank-$(k+i-1)$. Hence,

$$\left\|\mathbf{A} - \widetilde{\mathbf{A}}\right\|_{\mathrm{F}}^2 \geq \sum_{i\in[d-k]}\boldsymbol{\sigma}_i\left(\mathbf{A} - \widetilde{\mathbf{A}}\right)^2 \geq \sum_{i=k+1}^{d}\boldsymbol{\sigma}_i(\mathbf{A})^2.$$

It is straightforward to verify that equality is achieved above by taking $\widetilde{\mathbf{A}} = \sum_{i\in[k]}\boldsymbol{\sigma}_i\mathbf{u}_i\mathbf{v}_i^\top$. □

Theorem 1 shows that developing algorithms for computing low-rank approximations to possibly asymmetric $\mathbf{A} \in \mathbb{R}^{n\times d}$ in any unitarily-invariant norm reduces to efficiently performing k-PCA on a PSD matrix (i.e., $\mathbf{M} = \mathbf{A}^\top\mathbf{A}$). The rest of these notes focus on this latter task. In fact, we primarily focus on the $k = 1$ case for simplicity. However, throughout we will discuss how our methods extend to the case of general k, and indeed, the focus of Section 5 is how to use approximate 1-PCA algorithms in a black-box fashion to approximate k-PCA as well.

In the rest of these notes, $\mathbf{M}$ will always be a target matrix in $\mathbb{S}^{d\times d}_{\succeq\mathbf{0}}$ that we wish to perform PCA on. We denote its eigendecomposition (breaking ties arbitrarily) by

$$\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top = \sum_{i\in[d]}\boldsymbol{\lambda}_i\mathbf{u}_i\mathbf{u}_i^\top, \text{ where } \boldsymbol{\lambda} \text{ has nonincreasing coordinates.} \tag{5}$$

2 Krylov methods

In this section, we focus on algorithms for computing an approximate 1-PCA to $\mathbf{M} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$, which only access $\mathbf{M}$ through matrix-vector products. This access model is interesting for several reasons. First, it is well-motivated in applications where $\mathbf{M}$ is *implicit*. For example, if we are targetting low-rank approximations to $\mathbf{A} \in \mathbb{R}^{n\times d}$, the cost of actually computing $\mathbf{M} = \mathbf{A}^\top \mathbf{A}$ scales as $nd^{\omega-1}$ in theory and nd^2 in practice (cf. discussion in Remark 1, Part IX). However, we can simulate matrix-vector queries with $\mathbf{M}$ via two multiplications through $\mathbf{A}$, requiring just $O(\mathrm{nnz}(\mathbf{A}))$ time. More generally, when $\mathbf{M}$ is a small power or otherwise simple function of a matrix, explicitly forming $\mathbf{M}$ can be significantly more expensive than matrix-vector products. Finally, given the restricted nature of matrix-vector products, it is often possible to establish strong lower bounds on the performance of algorithms in this query model [BCW22, BN23].

We begin by describing the *power method*, perhaps the most famous algorithm for approximate 1-PCA (which, as seen in Lemma 1, is equivalent to top eigenvector computation).

Theorem 2 (Power method, gapped variant)**.** *Let $\mathbf{M} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ have eigendecomposition (5), and suppose for some $\Gamma \in (0,1)$, it is the case that $\boldsymbol{\lambda}_2 \leq (1-\Gamma)\boldsymbol{\lambda}_1$. Further, let $\delta, \Delta \in (0,1)$, and $p \geq \frac{8}{\Gamma}\log(\frac{32d}{\delta\Delta})$. Then, with probability $\geq 1-\delta$, we have that $\langle \hat{\mathbf{u}}, \mathbf{u}_1\rangle^2 \geq 1-\Delta$,[2] where*

$$\hat{\mathbf{u}} := \frac{\mathbf{M}^p\mathbf{g}}{\|\mathbf{M}^p\mathbf{g}\|_2} \text{ for } \mathbf{g} \sim \mathcal{N}(\mathbf{0}_d, \mathbf{I}_d).$$

Proof. For all $i \in [d]$, $\langle \mathbf{g}, \mathbf{u}_i\rangle$ is an independently-distributed random variable $\sim \mathcal{N}(0,1)$.[3] Each random variable is 1-sub-Gaussian, so with probability $\geq 1-\frac{\delta}{2}$, Theorem 1, Part VI shows

$$|\langle \mathbf{g}, \mathbf{u}_i\rangle| \leq \sqrt{2\log\left(\frac{2d}{\delta}\right)}, \text{ for all } i \in [d].$$

Moreover, we can directly show that with probability $\geq 1-\frac{\delta}{2}$, $|\langle \mathbf{g}, \mathbf{u}_1\rangle| \geq \frac{\delta}{4}$: indeed for any $r > 0$,

$$\Pr_{Z\sim\mathcal{N}(0,1)}[Z \in [-r,r]] = \frac{1}{\sqrt{2\pi}}\int_{-r}^{r}\exp\left(-\frac{s^2}{2}\right)\mathrm{d}s \leq \int_{-r}^{r}\mathrm{d}s \leq 2r.$$

Thus, union bounding on the above two events, we have that with probability $\geq 1-\delta$,

$$\frac{\langle \mathbf{g}, \mathbf{u}_i\rangle^2}{\langle \mathbf{g}, \mathbf{u}_1\rangle^2} \leq \frac{32\log\left(\frac{d}{\delta}\right)}{\delta^2} =: R, \text{ for all } 2 \leq i \leq d. \tag{6}$$

Condition on (6) in the remainder of the proof. Now, let $\mathbf{P} := \mathbf{M}^p$ and observe that $\boldsymbol{\lambda}_1(\mathbf{P}) = \boldsymbol{\lambda}_1^p \geq (1+\Gamma)^p\boldsymbol{\lambda}_2^p \geq \frac{dR}{\Delta}\boldsymbol{\lambda}_2(\mathbf{P})$ for our choice of p. Thus, we have

$$\begin{aligned}
\|\mathbf{P}\mathbf{g}\|_2^2 &= \sum_{i\in[d]}\langle \mathbf{P}\mathbf{g}, \mathbf{u}_i\rangle^2 = \sum_{i\in[d]}\boldsymbol{\lambda}_i^p\langle \mathbf{g}, \mathbf{u}_i\rangle^2 \\
&= \boldsymbol{\lambda}_1^p\langle \mathbf{g}, \mathbf{u}_1\rangle^2\left(1 + \sum_{i=2}^{d}\left(\frac{\boldsymbol{\lambda}_i^p}{\boldsymbol{\lambda}_1^p}\right)\left(\frac{\langle \mathbf{g}, \mathbf{u}_i\rangle^2}{\langle \mathbf{g}, \mathbf{u}_1\rangle^2}\right)\right) \\
&\leq \boldsymbol{\lambda}_1^p\langle \mathbf{g}, \mathbf{u}_1\rangle^2\left(1 + dR\cdot\frac{\Delta}{dR}\right) = (1+\Delta)\boldsymbol{\lambda}_1^p\langle \mathbf{g}, \mathbf{u}_1\rangle^2.
\end{aligned}$$

Finally, the desired claim follows from

$$\langle \hat{\mathbf{u}}, \mathbf{u}_1\rangle^2 = \frac{\langle \mathbf{P}\mathbf{g}, \mathbf{u}_1\rangle^2}{\|\mathbf{P}\mathbf{g}\|_2^2} = \frac{\boldsymbol{\lambda}_1^p\langle \mathbf{g}, \mathbf{u}_1\rangle^2}{\|\mathbf{P}\mathbf{g}\|_2^2} \geq \frac{1}{1+\Delta} \geq 1-\Delta.$$

□

[2] In this context, it is more reasonable to track the squared quantity than $\langle \hat{\mathbf{u}}, \mathbf{u}_1\rangle$ directly, because $-\mathbf{u}_1$ is also a top eigenvector of $\mathbf{M}$, so we should accept either as a 1-PCA solution.

[3] This is clear when $\{\mathbf{u}_i\}_{i\in[d]}$ is the standard basis vectors $\{\mathbf{e}_i\}_{i\in[d]}$; the general case follows by rotational invariance of the Gaussian density (alternatively, direct computation on the PDF of multivariate Gaussians).

Theorem 1 has a simple and intuitive message. Assuming the existence of a *gap* in the spectrum of $\mathbf{M}$, i.e., that $\mathbf{u}_1$ is "obviously" the top eigenvector by at least a factor of $1-\Gamma$, we can amplify this gap by powering up the matrix $\mathbf{M}$ into $\mathbf{P} = \mathbf{M}^p$. In particular, $\mathbf{P}$ has the same eigenvectors as $\mathbf{M}$, but has a much larger gap, which is enough to outweigh differences in the initial random correlations between $\mathbf{g} \sim \mathcal{N}(\mathbf{0}_d, \mathbf{I}_d)$ and the eigenvectors $\{\mathbf{u}_i\}_{i\in[d]}$.

One could also ask: how does the power method perform when there is no gap in the $\{\boldsymbol{\lambda}_i\}_{i\in[d]}$? In this case, a goal such as $\langle \hat{\mathbf{u}}, \mathbf{u}_1\rangle^2 \geq 1-\Delta$ (as in Theorem 1) may not even be well-posed. For example, if $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_2$, then either $\hat{\mathbf{u}} = \mathbf{u}_1$ or $\hat{\mathbf{u}} = \mathbf{u}_2$ perfectly solves 1-PCA, but these vectors are orthogonal. Nonetheless, one could hope for the output $\hat{\mathbf{u}}$ to at least lie in the span of the "good candidates" for an approximate top eigenvector, dodging the orthogonal small eigenspace. This motivates our next definition for approximate PCA in the gap-free setting.

Definition 1 (Correlation 1-PCA)**.** *Let $\mathbf{M} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ have eigendecomposition (5), and let $\Gamma, \Delta \in (0,1)$. Further, let $\ell \in [d]$ satisfy $\boldsymbol{\lambda}_\ell > (1-\Gamma)\boldsymbol{\lambda}_1 \geq \boldsymbol{\lambda}_{\ell+1}$. We say that $\hat{\mathbf{u}} \in \mathcal{U}_1$ is a (Γ,Δ)-approximate correlation-1-PCA (or, (Γ,Δ)-1-cPCA) of $\mathbf{M}$ if*

$$\sum_{i\in[\ell]} \langle \hat{\mathbf{u}}, \mathbf{u}_i\rangle^2 \geq 1-\Delta.$$

Intuitively, the notion of approximation in Definition 1 penalizes any mass that $\hat{\mathbf{u}}$ puts outside the "large eigenvectors" $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_\ell$, but allows $\hat{\mathbf{u}}$ to vary arbitrarily within their span. This is a suitable generalization in the gap-free setting, treating eigenvectors that stay above the gap as equally-acceptable solutions. As $(\Gamma,\Delta) \to (0,0)$, we recover that $\hat{\mathbf{u}}$ must become a top eigenvector of $\mathbf{M}$. With this definition in hand, we give an analog to Theorem 1 in the gap-free setting.

Theorem 3 (Power method, gap-free variant)**.** *Let $\mathbf{M} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ have eigendecomposition (5), let $\delta, \Delta, \Gamma \in (0,1)$, and $p \geq \frac{8}{\Gamma}\log(\frac{32d}{\delta\Delta})$. Then, with probability $\geq 1-\delta$, we have that $\hat{\mathbf{u}}$ is a (Γ,Δ)-1-cPCA of $\mathbf{M}$, where*

$$\hat{\mathbf{u}} := \frac{\mathbf{M}^p\mathbf{g}}{\|\mathbf{M}^p\mathbf{g}\|_2} \text{ for } \mathbf{g} \sim \mathcal{N}(\mathbf{0}_d, \mathbf{I}_d).$$

Proof. Throughout this proof, we follow notation in Definition 1, and further, we condition on (6) holding, which gives the failure probability. Let

$$L := \sum_{i\in[\ell]} \langle \hat{\mathbf{u}}, \mathbf{u}_i\rangle^2, \; S := \sum_{i=\ell+1}^{d} \langle \hat{\mathbf{u}}, \mathbf{u}_i\rangle^2$$

be the correlations of $\hat{\mathbf{u}}$ with the large and small eigenspaces of $\mathbf{M}$, respectively. As in the proof of Theorem 1, under (6), we have that for $\mathbf{P} = \mathbf{M}^p$,

$$\sum_{i=\ell+1}^{d} \langle \mathbf{P}\mathbf{g}, \mathbf{u}_i\rangle^2 = \boldsymbol{\lambda}_1^p \langle \mathbf{g}, \mathbf{u}_1\rangle^2 \left(\sum_{i=\ell+1}^{d} \left(\frac{\boldsymbol{\lambda}_i^p}{\boldsymbol{\lambda}_1^p}\right)\left(\frac{\langle \mathbf{g}, \mathbf{u}_i\rangle^2}{\langle \mathbf{g}, \mathbf{u}_1\rangle^2}\right)\right) \leq \Delta\boldsymbol{\lambda}_1^p \langle \mathbf{g}, \mathbf{u}_1\rangle^2.$$

Thus, because $\hat{\mathbf{u}} \propto \mathbf{M}^p\mathbf{g}$ up to a common normalization factor,

$$\frac{L}{S} \geq \frac{\langle \hat{\mathbf{u}}, \mathbf{u}_1\rangle^2}{S} = \frac{\langle \mathbf{P}\mathbf{g}, \mathbf{u}_1\rangle^2}{\sum_{i=\ell+1}^{d} \langle \mathbf{P}\mathbf{g}, \mathbf{u}_i\rangle^2} \geq \frac{1}{\Delta}.$$

Finally, because $L + S = \|\mathbf{P}\mathbf{g}\|_2^2$, we have the desired

$$\sum_{i\in[\ell]} \langle \hat{\mathbf{u}}, \mathbf{u}_i\rangle^2 = \frac{L}{L+S} \geq \frac{1}{1+\Delta} \geq 1-\Delta.$$

□

The runtime of Theorem 3 is dominated by $p \approx \frac{1}{\Gamma}$ matrix-vector multiplications through $\mathbf{M}$. As alluded to in Section 1.2, Part VII, we can improve upon this runtime by using low-degree polynomial approximations to $\mathbf{M}^p$. This can be done explicitly (by directly applying the polynomial), or implicitly (via the Lanczos method, i.e., Theorem 1, Part VII).

Corollary 1. *In the setting of Theorem 3, let* $p \geq \frac{8}{\Gamma}\log(\frac{64d}{\delta\Delta})$. *There is an algorithm that uses*

$$m = \sqrt{2p\log\left(\frac{96}{\delta\Delta}\right)} = O\left(\frac{1}{\sqrt{\Gamma}}\log\left(\frac{d}{\delta\Delta}\right)\right)$$

matrix-vector multiplications through $\mathbf{M}$ *and* $O(m^2)$ *additional time, and with probability* $\geq 1-\delta$, *returns* $\hat{\mathbf{u}}$ *that is a* (Γ, Δ)*-1-cPCA of* $\mathbf{M}$.

Proof. We give a proof suppressing dependence on the ϵ parameter in Theorem 1, Part VII, and assuming that the error bound is $2\delta_k$ (as in the exact arithmetic setting of Section 5, Part VII), rather than $O(k \cdot \delta_k + \epsilon)$. These parts of the theorem statement are used to handle issues arising from working in finite-precision arithmetic; we defer the full proof to Theorem 18, [MMS18].

Recall from Lemma 1, Part VII, that there is a polynomial q of degree m satisfying

$$\sup_{x\in[-1,1]} |q(x) - x^p| \leq 2\exp\left(-\frac{m^2}{2p}\right) \leq \frac{\delta\Delta}{48}.$$

Thus, letting $r(x) := \boldsymbol{\lambda}_1^p q(\frac{x}{\boldsymbol{\lambda}_1})$, we have

$$\sup_{x\in[0,\boldsymbol{\lambda}_1]} |r(x) - x^p| \leq \frac{\delta\Delta}{48}\boldsymbol{\lambda}_1^p. \tag{7}$$

Next, for notational simplicity, let $\mathbf{n} := \mathbf{M}^p\mathbf{g}$ and $D := \|\mathbf{n}\|_2$, so that the output of Theorem 3 is $\frac{\mathbf{n}}{D}$. From the proof of Theorem 2, except with probability δ, we have that

$$D \geq |\langle \mathbf{P}\mathbf{g}, \mathbf{u}_1\rangle| = \boldsymbol{\lambda}_1^p |\langle \mathbf{g}, \mathbf{u}_1\rangle| \geq \frac{\delta\boldsymbol{\lambda}_1^p}{4}. \tag{8}$$

Now, let $\tilde{\mathbf{n}}$ be the output of the Lanczos method (Theorem 1, Part VII) with $k \leftarrow m$, $\mathbf{A} \leftarrow \mathbf{M}$, and $f(x) \leftarrow x^p$. By combining (7) and (8),

$$\|\tilde{\mathbf{n}} - \mathbf{n}\|_2 = \|\tilde{\mathbf{n}} - \mathbf{M}^p\mathbf{g}\|_2 \leq \frac{\delta\Delta}{24}\boldsymbol{\lambda}_1^p \leq \frac{\Delta D}{6}.$$

Therefore, letting $\widetilde{D} := \|\tilde{\mathbf{n}}\|_2$, and $\hat{\mathbf{u}} := \frac{\tilde{\mathbf{n}}}{\widetilde{D}}$ be our output vector,

$$\left\|\frac{\mathbf{n}}{D} - \frac{\tilde{\mathbf{n}}}{\widetilde{D}}\right\|_2 \leq \frac{1}{D}\|\mathbf{n} - \tilde{\mathbf{n}}\|_2 + \left|\frac{1}{D} - \frac{1}{\widetilde{D}}\right|\|\tilde{\mathbf{n}}\|_2 \leq \frac{\Delta}{6} + \frac{\frac{\Delta}{6}(1+\frac{\Delta}{6})}{1-\frac{\Delta}{6}} \leq \frac{\Delta}{2}.$$

Finally, let $\boldsymbol{\Pi} := \sum_{i=\ell+1}^{d}\mathbf{u}_i\mathbf{u}_i^\top$ be the projection matrix onto the small eigenspace of $\mathbf{M}$, as in Definition 1. Theorem 3 with our choice of p implies $\|\boldsymbol{\Pi}(\frac{\mathbf{n}}{D})\|_2 \leq \frac{\Delta}{2}$, so we have the desired claim:

$$\|\boldsymbol{\Pi}\hat{\mathbf{u}}\|_2 \leq \left\|\boldsymbol{\Pi}\left(\hat{\mathbf{u}} - \frac{\mathbf{n}}{D}\right)\right\|_2 + \left\|\boldsymbol{\Pi}\left(\frac{\mathbf{n}}{D}\right)\right\|_2 \leq \left\|\hat{\mathbf{u}} - \frac{\mathbf{n}}{D}\right\|_2 + \frac{\Delta}{2} \leq \Delta.$$

□

Up to low-order terms, Corollary 1 improves upon Theorem 3's runtime by a $\approx \Gamma^{-1/2}$ factor. These Krylov method-based algorithms admit various extensions: for example, they generalize to approximate k-PCA for $k > 1$ [MM15, AZL16], the low-order poly(m) additive runtime terms can be removed [AZL16], and even the leading-order term of $\approx \mathcal{T}_{\text{mv}}(\mathbf{M}) \cdot \text{poly}(\frac{1}{\Gamma})$ can be improved (as discussed in Section 4). The first of these extensions, i.e., the generalization to k-PCA, is fairly straightforward to obtain by slightly modifying the proofs of Theorems 2 and 3, and Corollary 1.

There has been recent work studying the optimality of Krylov methods for PCA and low-rank approximation. A particularly surprising result [BCW22] shows that Frobenius norm low-rank approximation, i.e., producing a rank-k projection matrix $\widehat{\boldsymbol{\Pi}} \in \mathbb{S}_{\succeq \mathbf{0}}^{d\times d}$ such that

$$\left\|\mathbf{A} - \mathbf{A}\widehat{\boldsymbol{\Pi}}\right\|_{\text{F}} \leq (1+\epsilon) \min_{\substack{\boldsymbol{\Pi}\in\mathbb{S}_{\succeq\mathbf{0}}^{d\times d} \\ \text{rank}(\boldsymbol{\Pi})=k \\ \boldsymbol{\lambda}(\boldsymbol{\Pi})\in\{0,1\}^d}} \|\mathbf{A} - \mathbf{A}\boldsymbol{\Pi}\|_{\text{F}}, \tag{9}$$

is achievable using only $\approx \epsilon^{-1/3}$ matrix-vector products. This improves upon direct applications of the Lanczos method (e.g., Corollary 1), which as shown in [MM15] use $\approx \epsilon^{-1/2}$ matrix-vector products to achieve a guarantee such as (9). For $k = 1$, [BN23] demonstrated that $\approx \epsilon^{-1/3}$ is the optimal matrix-vector query complexity for low-rank approximation in $\|\cdot\|_{\mathrm{F}}$, and $\approx \epsilon^{-1/2}$ is optimal in $\|\cdot\|_{\mathrm{op}}$; however, characterizing the landscape for general k remains open.

3 Notions of approximation

Definition 1 is not the only notion of approximate PCA commonly seen in the literature. In this section, we introduce another standard definition, and compare it to Definition 1.

Definition 2 (Energy 1-PCA)**.** *Let $\mathbf{M} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ have eigendecomposition (5), and let $\epsilon \in (0,1)$. We say that $\hat{\mathbf{u}}$ is an ϵ-approximate energy-1-PCA (or, ϵ-1-ePCA) of $\mathbf{M}$ if $\hat{\mathbf{u}} \in \mathcal{U}_1$, and*

$$\hat{\mathbf{u}}^\top \mathbf{M} \hat{\mathbf{u}} \geq (1-\epsilon) \max_{\mathbf{u} \in \mathcal{U}_1} \mathbf{u}^\top \mathbf{M} \mathbf{u} = (1-\epsilon)\boldsymbol{\lambda}_1.$$

Definition 2 is somewhat more straightforward than Definition 1; it simply requires that $\hat{\mathbf{u}}$ approximately solves the optimization problem (2) when $k = 1$, agnostic to the presence of a gap in $\boldsymbol{\lambda}$. Recall that in the exact case $\epsilon = 0$, we have by Lemma 1 that $\hat{\mathbf{u}}$ is a top eigenvector of $\mathbf{M}$. Moreover, guarantees for Definition 1 transfer to those for Definition 2 (and vice versa).

Lemma 2. *If $\hat{\mathbf{u}}$ is an ϵ-1-ePCA of $\mathbf{M} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$, it is a $(\Gamma, \frac{\epsilon}{\Gamma})$-1-cPCA of $\mathbf{M}$ for any $\Gamma \in (0,1)$.*

Proof. Following the notation in (5) and Definition 1, let $\mathbf{L} \in \mathbb{R}^{d\times \ell}$ consist of the first ℓ columns of $\mathbf{U}$ (i.e., the ℓ largest eigenvectors), and $\mathbf{S} \in \mathbb{R}^{d\times(d-\ell)}$ consist of the remaining columns. Further, let $\Delta := \left\|\mathbf{S}^\top \hat{\mathbf{u}}\right\|_2^2$, so the claim is $\Delta \leq \frac{\epsilon}{\Gamma}$. By the matrix Hölder inequality (Eq. (12), Part VI),

$$\begin{aligned}(1-\epsilon)\boldsymbol{\lambda}_1 \leq \hat{\mathbf{u}}^\top \mathbf{M} \hat{\mathbf{u}} &= \left\langle \hat{\mathbf{u}}\hat{\mathbf{u}}^\top, \mathbf{L}\mathbf{L}^\top \mathbf{M} \mathbf{L}\mathbf{L}^\top \right\rangle + \left\langle \hat{\mathbf{u}}\hat{\mathbf{u}}^\top, \mathbf{S}\mathbf{S}^\top \mathbf{M} \mathbf{S}\mathbf{S}^\top \right\rangle \\ &\leq \left\|\mathbf{L}^\top \hat{\mathbf{u}}\hat{\mathbf{u}}^\top \mathbf{L}\right\|_{\mathrm{tr}} \left\|\mathbf{S}^\top \mathbf{M} \mathbf{S}\right\|_{\mathrm{op}} + \left\|\mathbf{S}^\top \hat{\mathbf{u}}\hat{\mathbf{u}}^\top \mathbf{S}\right\|_{\mathrm{tr}} \left\|\mathbf{S}^\top \mathbf{M} \mathbf{S}\right\|_{\mathrm{op}} \\ &\leq (1-\Delta)\boldsymbol{\lambda}_1 + \Delta(1-\Gamma)\boldsymbol{\lambda}_1 = (1-\Delta\Gamma)\boldsymbol{\lambda}_1.\end{aligned}$$

The conclusion follows by rearranging and solving for Δ. □

Lemma 3. *If $\hat{\mathbf{u}}$ is a (Γ, Δ)-1-ePCA of $\mathbf{M} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$, it is a $(\Gamma + \Delta)$-1-ePCA of $\mathbf{M}$.*

Proof. Following the notation in the proof of Lemma 2,

$$\begin{aligned}\left\langle \hat{\mathbf{u}}\hat{\mathbf{u}}^\top, \mathbf{M} \right\rangle &= \left\langle \hat{\mathbf{u}}\hat{\mathbf{u}}^\top, \mathbf{L}\mathbf{L}^\top \mathbf{M} \mathbf{L}\mathbf{L}^\top \right\rangle + \left\langle \hat{\mathbf{u}}\hat{\mathbf{u}}^\top, \mathbf{S}\mathbf{S}^\top \mathbf{M} \mathbf{S}\mathbf{S}^\top \right\rangle \\ &\geq \left\langle \mathbf{L}^\top \hat{\mathbf{u}}\hat{\mathbf{u}}^\top \mathbf{L}, \mathbf{L}^\top \mathbf{M} \mathbf{L} \right\rangle \\ &\geq (1-\Delta)(1-\Gamma)\boldsymbol{\lambda}_1 \geq (1-\Gamma-\Delta)\boldsymbol{\lambda}_1.\end{aligned}$$

In the last line, we used that $\left\|\mathbf{L}^\top \hat{\mathbf{u}}\right\|_2^2 \geq 1 - \Delta$ by assumption, and that $\mathbf{L}^\top \hat{\mathbf{u}}$ is a vector in the span of $\mathbf{L}^\top \mathbf{M} \mathbf{L}$, whose smallest eigenvalue is at least $(1-\Gamma)\boldsymbol{\lambda}_1$. □

We give a short application of these conversion results, showing that we can efficiently estimate the top eigenvalue of a matrix via matrix-vector queries.

Corollary 2. *In the setting of Theorem 3, let $\delta, \epsilon \in (0,1)$, and let $p \geq \frac{16}{\epsilon}\log(\frac{128d}{\delta\epsilon})$. There is an algorithm that uses*

$$m = \sqrt{2p \log\left(\frac{192}{\delta\epsilon}\right)} = O\left(\frac{1}{\sqrt{\epsilon}} \log\left(\frac{d}{\delta\epsilon}\right)\right)$$

matrix-vector multiplications through $\mathbf{M}$ and $O(m^2)$ additional time, and with probability $\geq 1-\delta$, returns $\hat{\mathbf{u}}$ that is a ϵ-1-cPCA of $\mathbf{M}$. Assuming the success of this procedure, with $O(\mathcal{T}_{\mathrm{mv}}(\mathbf{M}) + d)$ additional time, we can compute $\hat{\lambda}$ satisfying $\boldsymbol{\lambda}_1 \geq \hat{\lambda} \geq (1-\epsilon)\boldsymbol{\lambda}_1$.

Proof. For the first conclusion, we set $\Gamma = \Delta = \frac{\epsilon}{2}$ in Corollary 1, and apply Lemma 3. For the second conclusion, it is enough to output $\hat{\mathbf{u}}^\top \mathbf{M} \hat{\mathbf{u}}$ which takes $O(\mathcal{T}_{\mathrm{mv}}(\mathbf{M}) + d)$ time. □

4 Shift-and-invert preconditioning

In this section, we describe the *shift-and-invert preconditioning* framework for computing approximate top eigenvectors of a matrix. This framework reduces top eigenvector computation to approximately solving a small number of well-conditioned linear systems. These subproblems in turn are amenable to stochastic optimization techniques, e.g., the stochastic variance-reduced gradient method of Section 6, Part III, that can improve the runtimes obtained in Section 2.

The basic idea of shift-and-invert is the following observation: for any $\lambda \geq \boldsymbol{\lambda}_1$, the matrix $\lambda \mathbf{I}_d - \mathbf{M}$ is positive semidefinite and has the eigenvalues $\{\lambda - \boldsymbol{\lambda}_i\}_{i \in [d]}$ in nondecreasing order (so $\lambda - \boldsymbol{\lambda}_d$ is largest). Thus, the power method applied to the *shifted* matrix $\lambda \mathbf{I}_d - \mathbf{M}$, for judiciously chosen λ, can be used to estimate the bottom eigenvector of $\mathbf{M}$. However, we are more interested in its application to PCA, which follows because the top eigenvector of the *shift-and-inverted* matrix $(\lambda \mathbf{I}_d - \mathbf{M})^{-1}$ is again $\mathbf{u}_1$. The upside is that $(\lambda \mathbf{I}_d - \mathbf{M})^{-1}$ may be much better-conditioned than $\mathbf{M}$, and hence a few applications of it is enough to estimate $\mathbf{u}_1$.

Here we describe a shift-and-invert preconditioning framework for top eigenvector approximation, adapted from [GHJ+16, AZL16]. Specifically, for some $\delta, \Gamma, \Delta \in (0, 1)$ fixed throughout, our goal is to return a (Γ, Δ)-1-cPCA of $\mathbf{M}$ with probability $\geq 1 - \delta$. For simplicity, we assume we can exactly solve linear systems. The main point of [GHJ+16, AZL16] is that the algorithm is robust to inexact solves, and that this can be used to obtain various applications (e.g., approximate k-PCA).

Estimating the top eigenvalue. The first step is to tighly estimate $\boldsymbol{\lambda}_1$. We show how to obtain $\hat{\lambda}$ satisfying $(1 + \frac{\Gamma}{4})\boldsymbol{\lambda}_1 \leq \hat{\lambda} \leq (1 + \frac{\Gamma}{2})\boldsymbol{\lambda}_1$, assuming we start with a rough initial estimate $\hat{\lambda}_0$ that satisfies $\boldsymbol{\lambda}_1 \leq \hat{\lambda}_0 \leq R\boldsymbol{\lambda}_1$ for some parameter R. We proceed in a sequence of iterations $0 \leq t < T$ maintaining an invariant $\hat{\lambda} \geq \boldsymbol{\lambda}_1$, where in each iteration t we compute a value

$$\alpha_t \in \left[\frac{1}{2}\left(\hat{\lambda}_t - \boldsymbol{\lambda}_1\right), \hat{\lambda}_t - \boldsymbol{\lambda}_1\right]. \tag{10}$$

We then update $\hat{\lambda}_{t+1} \leftarrow \hat{\lambda}_t - \alpha$, which clearly preserves the invariant that $\hat{\lambda}_t \geq \boldsymbol{\lambda}_1$ always, assuming (10) holds. Our termination criterion is $\alpha_T \leq \frac{\Gamma}{18}\hat{\lambda}_T$. This implies

$$\hat{\lambda}_T \leq \boldsymbol{\lambda}_1 + 2\alpha_T \leq \boldsymbol{\lambda}_1 + \frac{\Gamma}{9}\hat{\lambda}_T \implies \hat{\lambda}_T \leq \frac{1}{1 - \frac{\Gamma}{9}}\boldsymbol{\lambda}_1 \leq \left(1 + \frac{\Gamma}{6}\right)\boldsymbol{\lambda}_1,$$

from which we can set $\hat{\lambda} \leftarrow (1 + \frac{\Gamma}{4})\hat{\lambda}_T$ and obtain the desired $(1 + \frac{\Gamma}{4})\boldsymbol{\lambda}_1 \leq \hat{\lambda} \leq (1 + \frac{\Gamma}{2})\boldsymbol{\lambda}_1$.

The first key observation is that after few iterations, the stopping criterion $\alpha_T \leq \frac{\Gamma}{18}\hat{\lambda}_T$ must be met. Suppose this were not the case in some iteration t. Then,

$$\hat{\lambda}_t \geq \boldsymbol{\lambda}_1 + \alpha \geq \boldsymbol{\lambda}_1 + \frac{\Gamma}{18}\hat{\lambda}_t \geq \left(1 + \frac{\Gamma}{18}\right)\boldsymbol{\lambda}_1. \tag{11}$$

On the other hand, in each iteration the update makes multiplicative progress towards $\boldsymbol{\lambda}_1$:

$$\hat{\lambda}_{t+1} - \boldsymbol{\lambda}_1 = \hat{\lambda}_t - \alpha - \boldsymbol{\lambda}_1 \leq \left(\hat{\lambda}_t - \boldsymbol{\lambda}_1\right) - \frac{1}{2}\left(\hat{\lambda}_t - \boldsymbol{\lambda}_1\right) = \frac{1}{2}\left(\hat{\lambda}_t - \boldsymbol{\lambda}_1\right).$$

Thus, after at most $T = O(\log\left(\frac{R}{\Gamma}\right))$ iterations, the algorithm must terminate.

The second key observation is that to produce an estimate α_t, it is enough to apply Corollary 2 to the matrix $\mathbf{B}_t := (\hat{\lambda}_t \mathbf{I}_d - \mathbf{M})^{-1}$, with $\epsilon \leftarrow \frac{1}{2}$. This is because the top eigenvalue of $\mathbf{B}_t$ is $(\hat{\lambda}_t - \boldsymbol{\lambda}_1)^{-1}$, i.e., the inverse of what α_t in (10) wants to estimate. Corollary 2 requires $O(\log(\frac{dR}{\delta\Gamma}))$ (being conservative with the logarithmic factor) linear system solves in $\mathbf{B}_t$. Further, we claim that $\mathbf{B}_t$ is always well-conditioned before termination: by applying (11),

$$\frac{\boldsymbol{\lambda}_1(\mathbf{B}_t)}{\boldsymbol{\lambda}_d(\mathbf{B}_t)} \leq \frac{\hat{\lambda}_t}{\hat{\lambda}_t - \boldsymbol{\lambda}_1} \leq \frac{1 + \frac{\Gamma}{18}}{\frac{\Gamma}{18}} \leq \frac{19}{\Gamma}. \tag{12}$$

The condition number bound (12) will help us bound the runtime of linear system solves later.

Shift-and-inverse widens the gap. Next, recall that following the notation in Definition 1, our goal in (Γ, Δ)-1-cPCA is to return a vector $\hat{\mathbf{u}}$ satisfying

$$\left\|\mathbf{S}^\top \hat{\mathbf{u}}\right\|_2^2 \le \Delta^2, \text{ where } \mathbf{S}\mathbf{S}^\top = \sum_{i=\ell+1}^{d} \mathbf{u}_i \mathbf{u}_i^\top \tag{13}$$

is the projection matrix onto the eigenvectors of $\mathbf{M}$ below the gap. To see why the shift-and-invert preconditioning framework is beneficial computationally, observe that Krylov methods, e.g., Corollary 1, depend only mildly on the Δ parameter, but have a polynomial dependence on Γ^{-1}. Hence it is in our best interest to improve the gap parameter Γ via a transformation.

Fortunately, the transformation $\mathbf{M} \to \mathbf{B} := (\hat{\lambda}\mathbf{I}_d - \mathbf{M})^{-1}$ does exactly this, where $\hat{\lambda}$ is our previously-computed estimate satisfying $(1 + \frac{\Gamma}{4})\boldsymbol{\lambda}_1 \le \hat{\lambda} \le (1 + \frac{\Gamma}{2})\boldsymbol{\lambda}_1$. In particular,

$$\boldsymbol{\lambda}_{\ell+1}(\mathbf{B}) = \frac{1}{\hat{\lambda} - \boldsymbol{\lambda}_{\ell+1}} \le \frac{1}{\boldsymbol{\lambda}_1 - (1-\Gamma)\boldsymbol{\lambda}_1} = \frac{1}{\Gamma\boldsymbol{\lambda}_1},$$

while on the other hand,

$$\boldsymbol{\lambda}_1(\mathbf{B}) = \frac{1}{\hat{\lambda} - \boldsymbol{\lambda}_1} \ge \frac{2}{\Gamma\boldsymbol{\lambda}_1}.$$

Moreover, the eigenvectors of $\mathbf{B}$ and $\mathbf{M}$ are exactly the same, and appear in the same order. Thus, the guarantee (13) follows by computing a $(\frac{1}{2}, \Delta)$-1-cPCA of $\mathbf{B}$. By Corollary 1, this only requires $O(\log(\frac{d}{\delta\Delta}))$ matrix-vector multiplications through $\mathbf{B}$. By using similar logic to (12), we have

$$\frac{\boldsymbol{\lambda}_1(\mathbf{B})}{\boldsymbol{\lambda}_d(\mathbf{B})} = O\left(\frac{1}{\Gamma}\right).$$

Instantiating the framework. All told, before accounting for approximation error, we have reduced computing (Γ, Δ)-1-cPCA of a matrix $\mathbf{M}$ to solving $O(\log(\frac{dR}{\delta\Gamma\Delta}))$ linear systems in matrices of the form $\hat{\lambda}\mathbf{I}_d - \mathbf{M}$. Moreover, these matrices always have a condition number $O(\frac{1}{\Gamma})$.

By accounting for approximation error, [GHJ+16] show that solving linear systems with accelerated gradient descent (cf. Theorem 2, Part V and Lemma 11, Part II) already gives a runtime of $\approx \mathcal{T}_{\text{mv}}(\mathbf{M}) \cdot \Gamma^{-1/2}$. This shaves a low-order poly$(\frac{1}{\Gamma})$ term from Corollary 1's runtime.

To obtain further runtime improvements, [GHJ+16, AZL16] focus on the case $\mathbf{M} = \mathbf{A}^\top\mathbf{A}$ for some $\mathbf{A} \in \mathbb{R}^{n\times d}$ with rows $\{\mathbf{a}_i\}_{i\in[n]} \subset \mathbb{R}^d$. It can be shown that accelerated variance reduced methods (introduced in Section 6, Part III) can solve a linear system in $\mathbf{B} = \hat{\lambda}\mathbf{I}_d - \mathbf{A}^\top\mathbf{A}$ using

$$\approx \text{nnz}(\mathbf{A}) + \frac{\text{nnz}(\mathbf{A})^{\frac{3}{4}}(d \cdot \text{sr}(\mathbf{A}))^{\frac{1}{4}}}{\sqrt{\Gamma}} \tag{14}$$

time, where we hide logarithmic factors in the target error, and define the *stable rank* of $\mathbf{A}$ by

$$\text{sr}(\mathbf{A}) := \frac{\|\mathbf{A}\|_{\text{F}}^2}{\|\mathbf{A}\|_{\text{op}}^2} = \frac{\sum_{i\in[d]} \boldsymbol{\lambda}_i(\mathbf{M})^2}{\boldsymbol{\lambda}_1(\mathbf{M})^2}.$$

Observe that $\text{sr}(\mathbf{A}) \le \text{rank}(\mathbf{A}) \le d$, and in general, low-rank approximation is well-motivated when $\text{sr}(\mathbf{A})$ is small. In particular, when $\text{nnz}(\mathbf{A}) \approx nd$, the runtime (14) improves upon the $O(nd\cdot\Gamma^{-1/2})$ time required by accelerated gradient descent by a factor of $\approx (\frac{n}{\text{sr}(\mathbf{A})})^{1/4}$. By using the shift-and-invert framework, [GHJ+16] shows that the entire cost of (Γ, Δ)-1-cPCA is thus proportional to (14) up to logarithmic factors; for a range of moderate Γ, this is input-sparsity time.

5 Deflation methods

In this section, we overview a reduction-based approach for approximately performing k-PCA known as *deflation* (see, e.g., [Mac08]). This approach iteratively peels off approximate 1-PCAs to a residual matrix via orthogonal projection. Concretely, let $\mathcal{O} : \mathbb{S}^{d\times d}_{\succeq \mathbf{0}} \to \mathcal{U}_1$ be an algorithm that

returns an approximate top eigenvector to its input $\mathbf{M}$, for an approximation notion to be defined. Deflation methods for k-PCA initialize $\mathbf{\Pi}_0 \leftarrow \mathbf{I}_d$ and $i \leftarrow 1$, and iterate

$$\mathbf{v}_i \leftarrow \mathcal{O}\left(\mathbf{\Pi}_{i-1}\mathbf{M}\mathbf{\Pi}_{i-1}\right), \ \mathbf{\Pi}_i \leftarrow \mathbf{\Pi}_{i-1} - \mathbf{v}_i\mathbf{v}_i^\top, \text{ for } i \in [k]. \tag{15}$$

We assume that $\mathbf{v}_i \in \mathrm{Span}(\mathbf{\Pi}_{i-1})$ in each iteration i, which is essentially without loss of generality when accessing $\mathbf{\Pi}_{i-1}\mathbf{M}\mathbf{\Pi}_{i-1}$ via matrix-vector products. Thus, (15) returns the columns $\{\mathbf{v}_i\}_{i\in[k]}$ of an orthonormal matrix $\mathbf{V} \in \mathcal{U}_k$. Further, because distinct eigenspaces are orthogonal, Lemma 1 shows that when $\mathcal{O}$ computes top eigenvectors exactly, the output $\mathbf{V}$ is also an exact solution to the k-PCA problem (2). This is a black-box reduction from exact k-PCA to exact 1-PCA.

What is the quality of the output $\mathbf{V} \in \mathcal{U}_k$ from deflation methods, when $\mathcal{O}$ is approximate? In Section 5.1, we motivate this question via statistical settings, where polynomial dependences on accuracy parameters are necessary. In Sections 5.2 and 5.3, we survey results of [AZL16, JKL+24], who characterized the lossiness of deflation methods in different parameter regimes.

5.1 Statistical PCA

Consider the following *statistical PCA* problem: there is a distribution $\mathcal{D}$ over $\mathbb{R}^d$, with mean zero (i.e., $\mathbb{E}_{\mathbf{x}\sim\mathcal{D}}[\mathbf{x}] = \mathbf{0}_d$).[4] Our goal is to estimate the top eigenvector of the covariance matrix, $\mathbf{\Sigma} := \mathbb{E}_{\mathbf{x}\sim\mathcal{D}}[\mathbf{x}\mathbf{x}^\top]$, from samples $\{\mathbf{x}_i\}_{i\in[n]} \sim_{\text{i.i.d.}} \mathcal{D}$. Notably, in this problem we do not have access to $\mathbf{\Sigma}$, and instead must use empirical estimates computed from our dataset $\{\mathbf{x}_i\}_{i\in[n]}$.

Often, in machine learning applications, statistical PCA is our actual goal, so that we can learn an "important subspace" of $\mathcal{D}$ for use in downstream tasks, e.g., low-rank approximation or clustering. The *offline PCA* problem we have studied thus far (i.e., computing an approximate top eigenvector of the explicit matrix $\widehat{\mathbf{\Sigma}} := \frac{1}{n}\sum_{i\in[n]} \mathbf{x}_i\mathbf{x}_i^\top$) is only solved as a proxy for statistical PCA.

To analyze this strategy, we make the following assumptions about $\mathcal{D}$:

$$\left\|\mathbb{E}_{\mathbf{x}\sim\mathcal{D}}\left[\|\mathbf{x}\|_2^2\,\mathbf{x}\mathbf{x}^\top\right]\right\|_{\text{op}} \le \sigma^2, \text{ and } \|\mathbf{x}\|_2 \le R \text{ with probability 1 over } \mathbf{x} \sim \mathcal{D}. \tag{16}$$

For intuition, suppose $\mathcal{D} = \mathcal{N}(\mathbf{0}_d, \mathbf{\Sigma})$ for some $\mathbf{\Sigma} \preceq \mathbf{I}_d$. Then one can show that (16) holds with $\sigma^2 = O(d)$, using that Gaussian distributions satisfy the following 2-to-4 hypercontractivity bound:

$$\mathbb{E}\,\langle \mathbf{x}, \mathbf{u}\rangle^4 \le O(1) \text{ for all } \|\mathbf{u}\|_2 = 1.$$

Moreover, clipping the distribution so that $R \approx \sqrt{d}$ negligibly changes the covariance. More generally, even for *heavy-tailed* hypercontractive distributions where the latter bound in (16) fails and we only have the former, the bias of clipping to enforce $\|\mathbf{x}\|_2 \le R$ can usually be directly bounded, see e.g., Lemma 14, [JKL+24]. In this section, we will simply assume (16) holds.

Our primary tool used to compare the empirical and true covariances, $\widehat{\mathbf{\Sigma}}$ and $\mathbf{\Sigma}$, is a variant of an eigenspace perturbation result by [Wed72], which we adapt from Lemma B.3, [AZL16]. Intuitively, it says that if we lightly perturb a matrix $\mathbf{M}$, then eigenspaces of $\mathbf{M}$ that originally had a gap between them still remain mostly-uncorrelated after the perturbation.

Lemma 4 (Gap-free Wedin's theorem)**.** *Let $\epsilon, \lambda, \tau > 0$, and let $\mathbf{M}, \widehat{\mathbf{M}} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$ have $\|\mathbf{M} - \widehat{\mathbf{M}}\|_{\text{op}} \le \epsilon$. Let $\mathbf{L}, \mathbf{S}$ have eigenvectors of $\mathbf{M}$ with eigenvalues $> \lambda$ and $\le \lambda$ as columns respectively, so $\mathbf{L}\mathbf{L}^\top + \mathbf{S}\mathbf{S}^\top = \mathbf{I}_d$. Similarly, let $\widehat{\mathbf{L}}, \widehat{\mathbf{S}}$ have eigenvectors of $\widehat{\mathbf{M}}$ with eigenvalues $> \lambda + \tau$ and $\le \lambda + \tau$ as columns respectively. Then,*

$$\left\|\mathbf{S}^\top\widehat{\mathbf{L}}\right\|_{\text{op}} \le \frac{\epsilon}{\tau}.$$

Proof. For convenience, let us write the entire eigendecompositions of $\mathbf{M}, \widehat{\mathbf{M}}$, as

$$\mathbf{M} = \mathbf{L}\mathbf{\Lambda}_L\mathbf{L}^\top + \mathbf{S}\mathbf{\Lambda}_S\mathbf{S}^\top, \ \widehat{\mathbf{M}} = \widehat{\mathbf{L}}\widehat{\mathbf{\Lambda}}_L\widehat{\mathbf{L}}^\top + \widehat{\mathbf{S}}\widehat{\mathbf{\Lambda}}_S\widehat{\mathbf{S}}^\top,$$

[4] This zero mean assumption is without loss of generality in the context of statistical PCA. Otherwise, we can define a modified distribution $\mathcal{D}'$ where a draw from $\mathcal{D}'$ takes $\mathbf{x}, \mathbf{x}' \sim_{\text{i.i.d.}} \mathcal{D}$ and returns $\mathbf{x} - \mathbf{x}'$. Then, $\mathcal{D}'$ has mean $\mathbf{0}_d$, and has the same covariance matrix as $\mathcal{D}$ up to scaling, so we can solve PCA on $\mathcal{D}'$ instead.

so $\|\mathbf{\Lambda}_S\|_{\mathrm{op}} \le \lambda$, and $\|\widehat{\mathbf{\Lambda}}_L^{-1}\|_{\mathrm{op}} \le \frac{1}{\lambda+\tau}$. Letting $\mathbf{R} := \mathbf{M} - \widehat{\mathbf{M}}$, we have by orthogonality of $\mathbf{S}, \mathbf{L}$ that

$$\begin{aligned}
\mathbf{\Lambda}_S \mathbf{S}^\top &= \mathbf{S}^\top \mathbf{M} = \mathbf{S}^\top \left(\widehat{\mathbf{M}} + \mathbf{R}\right) \\
\implies \mathbf{\Lambda}_S \mathbf{S}^\top \widehat{\mathbf{L}} &= \mathbf{S}^\top \widehat{\mathbf{M}}^\top \widehat{\mathbf{L}} + \mathbf{S}^\top \mathbf{R} \widehat{\mathbf{L}} = \mathbf{S}^\top \widehat{\mathbf{L}} \widehat{\mathbf{\Lambda}}_L + \mathbf{S}^\top \mathbf{R} \widehat{\mathbf{L}} \\
\implies \mathbf{\Lambda}_S \mathbf{S}^\top \widehat{\mathbf{L}} \widehat{\mathbf{\Lambda}}_L^{-1} &= \mathbf{S}^\top \widehat{\mathbf{L}} + \mathbf{S}^\top \mathbf{R} \widehat{\mathbf{L}} \widehat{\mathbf{\Lambda}}_L^{-1}.
\end{aligned}$$

Thus, by taking operator norms of both sides,

$$\begin{aligned}
\frac{\lambda}{\lambda+\tau} \left\|\mathbf{S}^\top \widehat{\mathbf{L}}\right\|_{\mathrm{op}} &\ge \|\mathbf{\Lambda}_S\|_{\mathrm{op}} \left\|\mathbf{S}^\top \widehat{\mathbf{L}}\right\|_{\mathrm{op}} \left\|\widehat{\mathbf{\Lambda}}_L^{-1}\right\|_{\mathrm{op}} \\
&\ge \left\|\mathbf{S}^\top \widehat{\mathbf{L}}\right\|_{\mathrm{op}} - \left\|\mathbf{S}^\top \mathbf{R} \widehat{\mathbf{L}}\right\|_{\mathrm{op}} \left\|\widehat{\mathbf{\Lambda}}_L^{-1}\right\|_{\mathrm{op}} \ge \left\|\mathbf{S}^\top \widehat{\mathbf{L}}\right\|_{\mathrm{op}} - \frac{\epsilon}{\lambda+\tau},
\end{aligned}$$

and rearranging yields the bound $\|\mathbf{S}^\top \widehat{\mathbf{L}}\|_{\mathrm{op}} \le \frac{\epsilon}{\tau}$ as desired. □

We are now ready to analyze a simple strategy for statistical PCA: take enough i.i.d. samples from $\mathcal{D}$, and apply Theorem 3 (or Corollary 1) to the empirical covariance.

Proposition 1. *Let $\mathcal{D}$ be a distribution on $\mathbb{R}^d$ with mean $\mathbf{0}_d$ and covariance $\mathbf{\Sigma}$, assume that $\mathcal{D}$ satisfies (16), and let $\delta, \Gamma, \Delta \in (0,1)$. Given n samples $\{\mathbf{x}_i\}_{i\in[n]} \sim_{\text{i.i.d.}} \mathcal{D}$, for*

$$n \ge \left(\frac{64\sigma^2}{\boldsymbol{\lambda}_1(\mathbf{\Sigma})^2 \Delta \Gamma^2} + \frac{32R^2}{\boldsymbol{\lambda}_1(\mathbf{\Sigma})\sqrt{\Delta}\Gamma}\right) \log\left(\frac{2d}{\delta}\right),$$

any $(\frac{\Gamma}{6}, \frac{\Delta}{4})$-1-cPCA for $\widehat{\mathbf{\Sigma}} := \frac{1}{n}\sum_{i\in[n]} \mathbf{x}_i \mathbf{x}_i^\top$ is a (Γ, Δ)-1-cPCA for $\mathbf{\Sigma}$ with probability $\ge 1-\delta$.

Proof. Our first goal is to bound $\|\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}}\|_{\mathrm{op}}$, so we may apply Lemma 4. For all $i \in [n]$, define a random matrix $\mathbf{Z}_i := \frac{1}{n}(\mathbf{x}_i \mathbf{x}_i^\top - \mathbf{\Sigma})$, so that $\widehat{\mathbf{\Sigma}} = \sum_{i\in[n]} \mathbf{Z}_i$. Moreover, note that for all $i \in [n]$,

$$\begin{aligned}
\left\|\mathbb{E}\mathbf{Z}_i^2\right\|_{\mathrm{op}} &= \frac{1}{n^2}\left\|\mathbb{E}\left[\mathbf{x}_i \mathbf{x}_i^\top \mathbf{x}_i \mathbf{x}_i^\top\right] - \mathbf{\Sigma}^2\right\|_{\mathrm{op}} \\
&\le \frac{1}{n^2}\left\|\mathbb{E}\left[\mathbf{x}_i \mathbf{x}_i^\top \mathbf{x}_i \mathbf{x}_i^\top\right]\right\|_{\mathrm{op}} = \frac{1}{n^2}\left\|\mathbb{E}\left[\|\mathbf{x}_i\|_2^2 \mathbf{x}_i \mathbf{x}_i^\top\right]\right\|_{\mathrm{op}} \le \frac{\sigma^2}{n^2},
\end{aligned}$$

using the first bound in (16). The first line also used that $\mathbb{E}[\mathbf{x}_i \mathbf{x}_i^\top \mathbf{x}_i \mathbf{x}_i^\top] \succeq \mathbf{\Sigma}^2$, because

$$\mathbb{E}\left[\mathbf{u}^\top \mathbf{x}_i \mathbf{x}_i^\top \mathbf{x}_i \mathbf{x}_i^\top \mathbf{u}\right] = \mathbb{E}\left[\left\|\mathbf{x}_i \mathbf{x}_i^\top \mathbf{u}\right\|_2^2\right] \ge \left\|\mathbb{E}\left[\mathbf{x}_i \mathbf{x}_i^\top \mathbf{u}\right]\right\|_2^2 = \mathbf{u}^\top \mathbf{\Sigma}^2 \mathbf{u} \text{ for all } \mathbf{u} \in \mathbb{R}^d,$$

by convexity of $\|\cdot\|_2^2$ applied to the random vector $\mathbf{x}_i \mathbf{x}_i^\top \mathbf{u}$. Similarly,

$$\|\mathbf{Z}_i\|_{\mathrm{op}} = \frac{1}{n}\left\|\mathbf{x}_i \mathbf{x}_i^\top - \mathbf{\Sigma}\right\|_{\mathrm{op}} \le \frac{1}{n}\left\|\mathbf{x}_i \mathbf{x}_i^\top\right\|_{\mathrm{op}} + \frac{1}{n}\|\mathbf{\Sigma}\|_{\mathrm{op}} \le \frac{2R^2}{n} \text{ with probability 1,}$$

by the second bound in (16), and since $\|\mathbf{\Sigma}\|_{\mathrm{op}} \le R^2$ by convexity of $\|\cdot\|_{\mathrm{op}}$. We can now apply the matrix Bernstein inequality (Theorem 11, Part VI) with $\mathbf{Z} \leftarrow \mathbf{\Sigma}$, $c \leftarrow \frac{2R^2}{n}$, and $\sigma^2 \leftarrow \frac{\sigma^2}{n}$ to obtain

$$\Pr\left[\left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|_{\mathrm{op}} \ge t\right] \le 2d \exp\left(-\min\left(\frac{nt^2}{4\sigma^2}, \frac{nt}{8R^2}\right)\right) \le \delta, \text{ for } t = \frac{\sqrt{\Delta}\Gamma \boldsymbol{\lambda}_1(\mathbf{\Sigma})}{4}.$$

Assume that the event above does not hold. By the gap-free Wedin's theorem (Lemma 4) applied with $\mathbf{M} \leftarrow \mathbf{\Sigma}$, $\widehat{\mathbf{M}} \leftarrow \widehat{\mathbf{\Sigma}}$, $\lambda \leftarrow (1-\Gamma)\boldsymbol{\lambda}_1(\mathbf{\Sigma})$, and $\tau \leftarrow \frac{\Gamma}{2}\boldsymbol{\lambda}_1(\mathbf{\Sigma})$, we obtain

$$\left\|\widehat{\mathbf{L}}^\top \mathbf{S}\right\|_{\mathrm{op}} \le \frac{2}{\Gamma \boldsymbol{\lambda}_1(\mathbf{\Sigma})} \cdot \frac{\sqrt{\Delta}\Gamma \boldsymbol{\lambda}_1(\mathbf{\Sigma})}{4} = \frac{1}{2}\sqrt{\Delta}.$$

Here, we followed notation in Lemma 4, so $\mathbf{S}$ spans the eigenspace of $\mathbf{M}$ below $(1-\Gamma)\boldsymbol{\lambda}_1(\mathbf{\Sigma})$, and $\widehat{\mathbf{L}}$ spans the eigenspace of $\widehat{\mathbf{M}}$ above $(1-\frac{\Gamma}{2})\boldsymbol{\lambda}_1(\mathbf{\Sigma})$. Furthermore, note that $\boldsymbol{\lambda}_1(\widehat{\mathbf{\Sigma}}) \ge (1-\frac{\Gamma}{4})\boldsymbol{\lambda}_1(\mathbf{\Sigma})$, and $(1-\frac{\Gamma}{6})(1-\frac{\Gamma}{4})\boldsymbol{\lambda}_1(\mathbf{\Sigma}) \ge (1-\frac{\Gamma}{2})\boldsymbol{\lambda}_1(\mathbf{\Sigma})$. Thus if $\hat{\mathbf{u}}$ is a $(\frac{\Gamma}{6}, \frac{\Delta}{4})$-1-cPCA for $\widehat{\mathbf{\Sigma}}$, it must satisfy

$$\left\|\widehat{\mathbf{S}}^\top \hat{\mathbf{u}}\right\|_2^2 \le \frac{\Delta}{4},$$

where $\widehat{\mathbf{S}}$ is the complement to $\widehat{\mathbf{L}}$ as in Lemma 4. Combining the above two displays, we have the desired claim that $\hat{\mathbf{u}}$ is a (Γ, Δ)-cPCA, because

$$\left\|\mathbf{S}^\top \hat{\mathbf{u}}\right\|_2^2 \le \left(\left\|\mathbf{S}^\top \widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top \hat{\mathbf{u}}\right\|_2 + \left\|\mathbf{S}^\top \widehat{\mathbf{S}}\widehat{\mathbf{S}}^\top \hat{\mathbf{u}}\right\|_2\right)^2 \le \left(\left\|\mathbf{S}^\top \widehat{\mathbf{L}}\right\|_{\mathrm{op}} + \left\|\widehat{\mathbf{S}}^\top \hat{\mathbf{u}}\right\|_2\right)^2 \le \Delta^2.$$

□

Let us give an example to understand Proposition 1. As discussed earlier, after a mild amount of clipping, Gaussian distributions $\mathcal{D}$ satisfy $\frac{\sigma^2}{\boldsymbol{\lambda}_1(\boldsymbol{\Sigma})^2}, \frac{R^2}{\boldsymbol{\lambda}_1(\boldsymbol{\Sigma})} \lesssim d$. Thus, Proposition 1 states that

$$n \approx \frac{d \log(d)}{\Delta \Gamma^2} \tag{17}$$

samples are needed for the empirical covariance to serve as a good proxy in statistical PCA.

One interesting qualitative aspect of this bound is its polynomial dependence on both Γ^{-1} and Δ^{-1}. This is in contrast to the offline setting (e.g., Theorem 3, Corollary 1), which depended on $\log(\frac{1}{\Delta})$. In fact, it is known that these polynomial dependences are necessary in the statistical setting (Theorem 3.1, [VL13]), motivating the question of fine-grained guarantees for how the error parameters Γ, Δ blow up in deflation methods for cPCA.

Indeed, the design of approximate 1-PCA algorithms, let alone k-PCA algorithms, becomes even more complicated when additional constraints are added (see [JKL+24] for a list of well-studied statistical PCA problems, including streaming, dependent sample, robust, and private variants). Thus, deflation methods are attractive to algorithm designers as a way to focus on the least complicated 1-PCA case, presuming we can bound their degradation in quality.

5.2 Black-box ePCA

The good news is that if we shift our notion to approximation to ePCA (Definition 2), deflation methods result in *no blowup* of the approximation parameter. Concretely, following the notation in (2), say that $\mathbf{V} \in \mathcal{U}_k$ is an ϵ-k-ePCA of $\mathbf{M}$, if

$$\left\langle \mathbf{V}\mathbf{V}^\top, \mathbf{M} \right\rangle \ge (1-\epsilon) \max_{\mathbf{U} \in \mathcal{U}_k} \left\langle \mathbf{U}\mathbf{U}^\top, \mathbf{M} \right\rangle.$$

Then we have the following black-box reduction from k-ePCA to 1-ePCA.

Proposition 2. *Let $\mathbf{M} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$, and for any projection matrix $\mathbf{\Pi} \in \mathbb{S}^{d\times d}_{\succeq \mathbf{0}}$, let $\mathcal{O}$ be an oracle that takes input $\mathbf{\Pi}\mathbf{M}\mathbf{\Pi}$ and returns $\mathbf{v}$, an ϵ-1-ePCA to $\mathbf{\Pi}\mathbf{M}\mathbf{\Pi}$ satisfying $\mathbf{v} \in \mathrm{Span}(\mathbf{\Pi})$. Further, let $\mathbf{V} \in \mathcal{U}_k$ concatenate $\{\mathbf{v}_i\}_{i\in[k]}$ resulting from iterating (15). Then $\mathbf{V}$ is an ϵ-k-ePCA to $\mathbf{M}$.*

Proof. We proceed by induction on $i \in [k]$. Let $\mathbf{V}_i$ denote the horizontal concatenation of the first i calls to $\mathcal{O}$, so that $\mathbf{\Pi}_i = \mathbf{I}_d - \mathbf{V}_i\mathbf{V}_i^\top$. The inductive hypothesis tells us

$$\left\langle \mathbf{V}_i\mathbf{V}_i^\top, \mathbf{M} \right\rangle \ge (1-\epsilon) \max_{\mathbf{U} \in \mathcal{U}_i} \left\langle \mathbf{U}\mathbf{U}^\top, \mathbf{M} \right\rangle = \sum_{j\in[i]} \boldsymbol{\lambda}_j(\mathbf{M}),$$

where we applied Lemma 1 to compute the right-hand side. This then implies

$$\begin{aligned}
\mathrm{Tr}\left(\mathbf{V}_{i+1}^\top \mathbf{M} \mathbf{V}_{i+1}\right) &= \mathrm{Tr}\left(\mathbf{V}_i^\top \mathbf{M} \mathbf{V}_i\right) + \mathbf{v}_{i+1}^\top \mathbf{M} \mathbf{v}_{i+1} \\
&\ge (1-\epsilon)\left(\sum_{j\in[i]} \boldsymbol{\lambda}_j(\mathbf{M})\right) + \mathbf{v}_{i+1}^\top \mathbf{M} \mathbf{v}_{i+1} \\
&\ge (1-\epsilon)\left(\sum_{j\in[i]} \boldsymbol{\lambda}_j(\mathbf{M})\right) + (1-\epsilon)\left\|\mathbf{\Pi}_i \mathbf{M} \mathbf{\Pi}_i\right\|_{\mathrm{op}} \\
&\ge (1-\epsilon)\left(\sum_{j\in[i+1]} \boldsymbol{\lambda}_j(\mathbf{M})\right).
\end{aligned}$$

The second line used the inductive hypothesis on $\mathbf{V}_i$, the third line used the 1-ePCA guarantee on $\mathbf{v}_{i+1}$, and the last line applied the Cauchy interlacing theorem (Corollary 4, Part VI). □

5.3 Black-box cPCA

The story gets murkier when it comes to deflation methods for cPCA. In this discussion, let $\gamma, \delta \in (0,1)$, and suppose that $\mathcal{O}$ in (2) returns a (γ,δ)-1-cPCA to its input. Further, define $\kappa_k := \frac{\boldsymbol{\lambda}_1}{\boldsymbol{\lambda}_k}$, where $\mathbf{M}$ has eigendecomposition (5). Our goal is to understand when the deflation method (2) returns a set of vectors $\{\mathbf{v}_i\}_{i\in[k]}$ that is a (Γ,Δ)-k-cPCA, and bound how large Γ, Δ are as a function of the original parameters γ, δ, as well as potentially k, κ_k.

The baseline is to use Proposition 2 alongside our conversion results, Lemmas 2 and 3. In fact, one can show a variant of Lemma 2 that says any ϵ-k-ePCA of $\mathbf{M}$ is also an $(\frac{\epsilon k \kappa_k}{\Gamma}, \Gamma)$-$k$-cPCA of $\mathbf{M}$ (Lemma 1, [JKL+24]). Directly plugging this into Proposition 2 implies it is enough to take $\epsilon = \frac{\Gamma\Delta}{k\kappa_k}$, so Lemma 3 shows we can take $\gamma = \delta = O(\frac{\Gamma\Delta}{k\kappa_k})$ in our 1-cPCA oracle $\mathcal{O}$.

In the statistical setting with i.i.d. Gaussian data, putting these parameters into (17) results in

$$n \approx d\log(d) \cdot \frac{k^3\kappa_k^3}{\Gamma^3\Delta^3}$$

samples required to solve 1-cPCA to the level needed for deflation to yield a (Γ,Δ)-k-cPCA. Improving upon this in some parameter regimes, [AZL16] showed that it is enough to take

$$\gamma = \frac{\Gamma}{2},\ \delta = \Theta\left(\frac{\Gamma^2\Delta^2}{k^4\kappa_k^2}\right),$$

for (15) to give a (Γ,Δ)-k-cPCA, which in the case of (17) needs $\approx d\log(d) \cdot \frac{k^4\kappa_k^2}{\Delta^2\Gamma^4}$ samples for $\mathcal{O}$. The upshot of the [AZL16] result is that it works very well in the *offline* cPCA setting, where polylog$(\frac{1}{\Delta})$ rates are possible given an explicit matrix (Theorem 3, Corollary 1). Thus, the blowup of $\delta \to \Delta$ in their reduction is less of an issue, and $\gamma \approx \Gamma$ is the salient feature.

Returning to the statistical setting, it is known that the optimal sample complexity of solving (Γ,Δ)-k-cPCA in one shot (rather than via deflation methods) scales as $\approx \frac{\kappa_k}{\Gamma^2\Delta}$ [VL13], which is tight up to the dependence on k [AL17]. This matches the dependences of the 1-cPCA case in terms of Γ, Δ. Thus, ambitiously, could we hope for *lossless* or near-lossless cPCA reductions, more in line with what we showed in Proposition 2 (where ϵ did not blow up at all)?

This question was studied recently by [JKL+24], who showed the answer is actually no: if $\Gamma \lesssim \kappa_k\sqrt{\Delta}$, deflation methods fail to give a lossless cPCA reduction, even if $k = 2$ and $d = 3$. This shows a qualitative separation between our approximation notions in Definitions 1 and 2. More generally, assuming that we are in the opposite regime $\Gamma \gtrsim \kappa_k\sqrt{\Delta}$, a lossless reduction actually is possible for any constant k (Theorem 2, [JKL+24]). Unfortunately, the parameters in [JKL+24] reduction lose a $k^{\Theta(\log k)}$ factor, and it is an open problem to improve the k dependence to polynomial.

Source material

This lecture is based on the author's own experience working in the field.

References

[AL17] Zeyuan Allen-Zhu and Yuanzhi Li. First efficient convergence for streaming k-pca: A global, gap-free, and near-optimal rate. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017*, pages 487–492. IEEE Computer Society, 2017.

[AZL16] Zeyuan Allen-Zhu and Yuanzhi Li. Even faster SVD decomposition yet without agonizing pain. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016*, pages 974–982, 2016.

[BCW22] Ainesh Bakshi, Kenneth L. Clarkson, and David P. Woodruff. Low-rank approximation with $1/\varepsilon^{1/3}$ matrix-vector products. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, 2022*, pages 1130–1143. ACM, 2022.

[BN23] Ainesh Bakshi and Shyam Narayanan. Krylov methods are (nearly) optimal for low-rank approximation. In *64th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2023*, pages 2093–2101. IEEE, 2023.

[EY36] Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211–218, 1936.

[GHJ+16] Dan Garber, Elad Hazan, Chi Jin, Sham M. Kakade, Cameron Musco, Praneeth Netrapalli, and Aaron Sidford. Faster eigenvector computation via shift-and-invert preconditioning. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 2626–2634. JMLR.org, 2016.

[JKL+24] Arun Jambulapati, Syamantak Kumar, Jerry Li, Shourya Pandey, Ankit Pensia, and Kevin Tian. Black-box k-to-1-pca reductions: Theory and applications. In *The Thirty Seventh Annual Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 2564–2607. PMLR, 2024.

[Lia23] Xin Liang. On the optimality of the oja's algorithm for online pca. *Statistics and Computing*, 33(3):62, 2023.

[Mac08] Lester W. Mackey. Deflation methods for sparse PCA. In *Advances in Neural Information Processing Systems 21, Proceedings of the Twenty-Second Annual Conference on Neural Information Processing Systems*, pages 1017–1024. Curran Associates, Inc., 2008.

[Mir60] Leon Mirsky. Symmetric gauge functions and unitarily invariant norms. *Quart. J. Math. Oxford*, 11:50–59, 1960.

[MM15] Cameron Musco and Christopher Musco. Randomized block krylov methods for stronger and faster approximate singular value decomposition. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015*, pages 1396–1404, 2015.

[MMS18] Cameron Musco, Christopher Musco, and Aaron Sidford. Stability of the lanczos method for matrix function approximation. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018*, pages 1605–1624. SIAM, 2018.

[VL13] Vincent Q. Vu and Jing Lei. Minimax sparse principal subspace estimation in high dimensions. *Annals of Statistics*, 41(6):2905–2947, 2013.

[Wed72] Per-Ake Wedin. Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics*, 12(1):99–111, 1972.