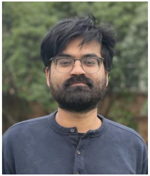
Black-Box k-to-1-PCA Reductions: Theory and Applications

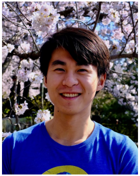
Conference on Learning Theory, 2024

Ankit Pensia



Arun Jambulapati

Syamantak Kumar



Jerry Li

Shourya Pandey

Kevin Tian

Introducing principal component analysis (PCA)

Problem statement. (k-PCA)
Let $\mathcal{D}$ be a distribution on $\mathbb{R}^d$ with covariance matrix Σ.

Input:	(restricted) sample access to $\mathcal{D}$
Output:	(approx.) top-k eigenvectors of Σ.

- Ubiquitous in statistical estimation, dimensionality reduction

Introducing principal component analysis (PCA)

Problem statement. (k-PCA)
Let $\mathcal{D}$ be a distribution on $\mathbb{R}^d$ with covariance matrix Σ.

Input: **(restricted) sample access** to $\mathcal{D}$

Output: (approx.) top-k eigenvectors of Σ.

- Ubiquitous in statistical estimation, dimensionality reduction
- Indirect access to Σ (can't perform matrix-vector products)

Introducing principal component analysis (PCA)

Problem statement. (k-PCA)
Let $\mathcal{D}$ be a distribution on $\mathbb{R}^d$ with covariance matrix Σ.

Input:	**(restricted)** sample access to $\mathcal{D}$
Output:	(approx.) top-k eigenvectors of Σ.

- Ubiquitous in statistical estimation, dimensionality reduction
- Indirect access to Σ (can't perform matrix-vector products)
- Extensively studied under various notions of **restrictions** recently
  - i.i.d. samples, corrupted samples, correlated samples
  - differential privacy, fairness,

Introducing principal component analysis (PCA)

Problem statement. (k-PCA)
Let $\mathcal{D}$ be a distribution on $\mathbb{R}^d$ with covariance matrix Σ.

Input:	(restricted) sample access to $\mathcal{D}$
Output:	(approx.) top-k eigenvectors of Σ.

- Ubiquitous in statistical estimation, dimensionality reduction
- Indirect access to Σ (can't perform matrix-vector products)
- Extensively studied under various notions of restrictions recently
  - i.i.d. samples, corrupted samples, correlated samples
  - differential privacy, fairness,
- However, most works obtain guarantees only for $k = 1$
  - But, many practical applications need $k > 1$

Introducing principal component analysis (PCA)

Problem statement. (k-PCA)
Let $\mathcal{D}$ be a distribution on $\mathbb{R}^d$ with covariance matrix Σ.

Input:	(restricted) sample access to $\mathcal{D}$
Output:	(approx.) top-k eigenvectors of Σ.

- Ubiquitous in statistical estimation, dimensionality reduction
- Indirect access to Σ (can't perform matrix-vector products)
- Extensively studied under various notions of restrictions recently
  - i.i.d. samples, corrupted samples, correlated samples
  - differential privacy, fairness,
- However, most works obtain guarantees only for $k = 1$
  - But, many practical applications need $k > 1$

Can we generalize these existing techniques to $k > 1$?

Introducing deflation: A generic reduction to 1-PCA

Input:

- $k \in [d]$
- $\mathcal{O}_{\text{1-PCA}}$, an arbitrary oracle for (approximate) 1-PCA
- $\boldsymbol{M}$, a $d \times d$ PSD matrix, (access through $\mathcal{O}_{\text{1-PCA}}$)

Introducing **deflation**: A generic reduction to 1-PCA

Input:
- ▷ $k \in [d]$
- ▷ $\mathcal{O}_{\text{1-PCA}}$, an arbitrary oracle for (approximate) 1-PCA
- ▷ $\boldsymbol{M}$, a $d \times d$ PSD matrix, (access through $\mathcal{O}_{\text{1-PCA}}$)

1. $\mathbf{P}_0 \leftarrow \mathbf{I}_d$ (identity projection)
2. For $i \in [k]$:
   - 2.1 $u_i \leftarrow \mathcal{O}_{\text{1-PCA}}(\mathbf{P}_{i-1}\mathbf{M}\mathbf{P}_{i-1})$ (top component in projected space)
   - 2.2 $\mathbf{P}_i \leftarrow \mathbf{P}_{i-1} - u_i u_i^\top$ (updating the projection)

- Repeatedly **deflates** the directions returned by $\mathcal{O}_{\text{1-PCA}}$

Introducing deflation: A generic reduction to 1-PCA

Input:
- $k \in [d]$
- $\mathcal{O}_{\text{1-PCA}}$, an arbitrary oracle for (approximate) 1-PCA
- $\boldsymbol{M}$, a $d \times d$ PSD matrix, (access through $\mathcal{O}_{\text{1-PCA}}$)

1. $\mathbf{P}_0 \leftarrow \mathbf{I}_d$ (identity projection)
2. For $i \in [k]$:
   - 2.1 $u_i \leftarrow \mathcal{O}_{\text{1-PCA}}(\mathbf{P}_{i-1}\mathbf{M}\mathbf{P}_{i-1})$ (top component in projected space)
   - 2.2 $\mathbf{P}_i \leftarrow \mathbf{P}_{i-1} - u_i u_i^\top$ (updating the projection)

Output: $\{u_1, \dots, u_k\}$

- Repeatedly deflates the directions returned by $\mathcal{O}_{\text{1-PCA}}$

Introducing deflation: A generic reduction to 1-PCA

Input:
- ▷ $k \in [d]$
- ▷ $\mathcal{O}_{\text{1-PCA}}$, an arbitrary oracle for (approximate) 1-PCA
- ▷ $\boldsymbol{M}$, a $d \times d$ PSD matrix, (access through $\mathcal{O}_{\text{1-PCA}}$)

1. $\mathbf{P}_0 \leftarrow \mathbf{I}_d$ (identity projection)

2. For $i \in [k]$:

2.1 $u_i \leftarrow \mathcal{O}_{\text{1-PCA}}(\mathbf{P}_{i-1}\mathbf{M}\mathbf{P}_{i-1})$ (top component in projected space)

2.2 $\mathbf{P}_i \leftarrow \mathbf{P}_{i-1} - u_i u_i^\top$ (updating the projection)

Output: $\{u_1, \ldots, u_k\}$

- Repeatedly deflates the directions returned by $\mathcal{O}_{\text{1-PCA}}$
- Importantly, can be performed using **samples** (w/o direct access to $\mathbf{M}$)

Introducing deflation: A generic reduction to 1-PCA

Input:
- ▷ $k \in [d]$
- ▷ $\mathcal{O}_{\text{1-PCA}}$, an arbitrary oracle for (approximate) 1-PCA
- ▷ $\boldsymbol{M}$, a $d \times d$ PSD matrix, (access through $\mathcal{O}_{\text{1-PCA}}$)

1. $\mathbf{P}_0 \leftarrow \mathbf{I}_d$ (identity projection)

2. For $i \in [k]$:

2.1 $u_i \leftarrow \mathcal{O}_{\text{1-PCA}}(\mathbf{P}_{i-1}\mathbf{M}\mathbf{P}_{i-1})$ (top component in projected space)

2.2 $\mathbf{P}_i \leftarrow \mathbf{P}_{i-1} - u_i u_i^\top$ (updating the projection)

Output: $\{u_1, \ldots, u_k\}$

- Repeatedly deflates the directions returned by $\mathcal{O}_{\text{1-PCA}}$
- Importantly, can be performed using samples (w/o direct access to $\mathbf{M}$)
- A **natural** but **not-well-understood** technique

Existing literature on deflation

- If the 1-PCA oracle, $\mathcal{O}_{\text{1-PCA}}$, is exact, then deflation is exact
- **Main question**: What if $\mathcal{O}_{\text{1-PCA}}$ is only **approximately correct**?

Existing literature on deflation

- If the 1-PCA oracle, $\mathcal{O}_{\text{1-PCA}}$, is exact, then deflation is exact
- **Main question**: What if $\mathcal{O}_{\text{1-PCA}}$ is only **approximately correct**?
- **Challenge**: How do approximation errors **compound**?

Existing literature on deflation

- If the 1-PCA oracle, $\mathcal{O}_{\text{1-PCA}}$, is exact, then deflation is exact
- **Main question**: What if $\mathcal{O}_{\text{1-PCA}}$ is only **approximately correct**?
- Challenge: How do approximation errors compound?
- Relatively few concrete guarantees
  - Phrased as an open problem in [MM15]

[MM15] C. Musco, C. Musco. Randomized block krylov methods for stronger and faster approximate SVD. *NeurIPS* (2015)

Existing literature on deflation

- If the 1-PCA oracle, $\mathcal{O}_{\text{1-PCA}}$, is exact, then deflation is exact
- **Main question**: What if $\mathcal{O}_{\text{1-PCA}}$ is only **approximately correct**?
- Challenge: How do approximation errors compound?
- Relatively few concrete guarantees
  - Phrased as an open problem in [MM15]
  - [LZ15]: makes strong eigengap assumptions

[MM15] C. Musco, C. Musco. Randomized block krylov methods for stronger and faster approximate SVD. *NeurIPS* (2015)
[LZ15] R.-C. Li, L.-H. Zhang. Convergence of block lanczos for eigenvalue clusters. *Numerische Mathematik* (2015)

Existing literature on deflation

- If the 1-PCA oracle, $\mathcal{O}_{\text{1-PCA}}$, is exact, then deflation is exact
- **Main question**: What if $\mathcal{O}_{\text{1-PCA}}$ is only **approximately correct**?
- Challenge: How do approximation errors compound?
- Relatively few concrete guarantees
  - Phrased as an open problem in [MM15]
  - [LZ15]: makes strong eigengap assumptions
  - [AL16]: Studies one particular notion of approximation (defined next)

[MM15] C. Musco, C. Musco. Randomized block krylov methods for stronger and faster approximate SVD. *NeurIPS* (2015)
[LZ15] R.-C. Li, L.-H. Zhang. Convergence of block lanczos for eigenvalue clusters. *Numerische Mathematik* (2015)
[AL16] Z. Allen Zhu, Y. Li. LazySVD: Even Faster SVD Decomposition Yet Without Agonizing Pain. *NeurIPS*. 2016

Existing literature on deflation

- If the 1-PCA oracle, $\mathcal{O}_{\text{1-PCA}}$, is exact, then deflation is exact
- **Main question**: What if $\mathcal{O}_{\text{1-PCA}}$ is only **approximately correct**?
- Challenge: How do approximation errors compound?
- Relatively few concrete guarantees
  - Phrased as an open problem in [MM15]
  - [LZ15]: makes strong eigengap assumptions
  - [AL16]: Studies one particular notion of approximation (defined next)
    - However, their focus is on the regime with whitebox/explicit access to $\mathbf{M}$, and achieves different tradeoffs (more later)

[MM15] C. Musco, C. Musco. Randomized block krylov methods for stronger and faster approximate SVD. *NeurIPS* (2015)
[LZ15] R.-C. Li, L.-H. Zhang. Convergence of block lanczos for eigenvalue clusters. *Numerische Mathematik* (2015)
[AL16] Z. Allen Zhu, Y. Li. LazySVD: Even Faster SVD Decomposition Yet Without Agonizing Pain. *NeurIPS*. 2016

Existing literature on deflation

- If the 1-PCA oracle, $\mathcal{O}_{\text{1-PCA}}$, is exact, then deflation is exact
- **Main question**: What if $\mathcal{O}_{\text{1-PCA}}$ is only **approximately correct**?
- Challenge: How do approximation errors compound?
- Relatively few concrete guarantees
  - Phrased as an open problem in [MM15]
  - [LZ15]: makes strong eigengap assumptions
  - [AL16]: Studies one particular notion of approximation (defined next)
    - However, their focus is on the regime with whitebox/explicit access to $\mathbf{M}$, and achieves different tradeoffs (more later)

Can we develop a better understanding of deflation?

[MM15] C. Musco, C. Musco. Randomized block krylov methods for stronger and faster approximate SVD. *NeurIPS* (2015)
[LZ15] R.-C. Li, L.-H. Zhang. Convergence of block lanczos for eigenvalue clusters. *Numerische Mathematik* (2015)
[AL16] Z. Allen Zhu, Y. Li. LazySVD: Even Faster SVD Decomposition Yet Without Agonizing Pain. *NeurIPS*. 2016

Approximate k-PCA

- Multiple notions of approximation
- Inspired by ML/statistics literature, we study two notions:
  1. **energy**-PCA
  2. **correlation**-PCA

Approximate k-PCA

- Multiple notions of approximation
- Inspired by ML/statistics literature, we study two notions:
  1. **energy**-PCA
  2. **correlation**-PCA
- Importantly, both are **gap-free**; no separation between eigenvalues

Approximation notion: energy

An orthonormal matrix $\mathbf{U} = (u_1, \ldots, u_k) \in \mathbb{R}^{d \times k}$ is an ϵ-approximate k-**energy**-PCA of a PSD matrix $\mathbf{M} \in \mathbb{R}^{d \times d}$ if

$$\sum_{i=1}^{k} u_i^\top \mathbf{M} u_i \geq (1 - \epsilon) \sum_{i=1}^{k} \lambda_i(\mathbf{M})$$

- Maximum amount of energy/variance: $\sum_{i=1}^{k} \lambda_i(\mathbf{M})$
  - Achieved when u_i's are leading eigenvectors

Approximation notion: energy

An orthonormal matrix $\mathbf{U} = (u_1, \dots, u_k) \in \mathbb{R}^{d \times k}$ is an ϵ-approximate k-**energy**-PCA of a PSD matrix $\mathbf{M} \in \mathbb{R}^{d \times d}$ if

$$\sum_{i=1}^{k} u_i^\top \mathbf{M} u_i \geq (1-\epsilon) \sum_{i=1}^{k} \lambda_i(\mathbf{M})$$

- Maximum amount of energy/variance: $\sum_{i=1}^{k} \lambda_i(\mathbf{M})$
  - Achieved when u_i's are leading eigenvectors

Is deflation energy-(PCA)-efficient?

$\sum_{i=1}^{k} u_i^\top \mathbf{M} u_i \geq (1-\epsilon) \sum_{i=1}^{k} \lambda_i(\mathbf{M})$

Our result for energy-PCA

Theorem: [JKLPPT24]

If the deflation algorithm uses ϵ-approximate **1**-energy-PCA as $\mathcal{O}_{\text{1-PCA}}$ subroutine, then it outputs an ϵ-approximate k-energy-PCA.

- Deflation is *lossless* for energy approximation!

$\sum_{i=1}^{k} u_i^\top \mathbf{M} u_i \geq (1-\epsilon) \sum_{i=1}^{k} \lambda_i(\mathbf{M})$

Our result for energy-PCA

Theorem: [JKLPPT24]

If the deflation algorithm uses ϵ-approximate **1**-energy-PCA as $\mathcal{O}_{\text{1-PCA}}$ subroutine, then it outputs an ϵ-approximate k-energy-PCA.

- Deflation is *lossless* for energy approximation!
- **Application**: Stronger results for outlier-robust PCA, heavy-tailed PCA

$\sum_{i=1}^{k} u_i^\top \mathbf{M} u_i \geq (1-\epsilon) \sum_{i=1}^{k} \lambda_i(\mathbf{M})$

Our result for energy-PCA

Theorem: [JKLPPT24]

If the deflation algorithm uses ϵ-approximate 1-energy-PCA as $\mathcal{O}_{\text{1-PCA}}$ subroutine, then it outputs an ϵ-approximate k-energy-PCA.

- Deflation is *lossless* for energy approximation!
- Application: Stronger results for outlier-robust PCA, heavy-tailed PCA
- Three-line proof using induction

$\sum_{i=1}^{k} u_i^\top \mathbf{M} u_i \geq (1-\epsilon) \sum_{i=1}^{k} \lambda_i(\mathbf{M})$

Our result for energy-PCA

Theorem: [JKLPPT24]

If the deflation algorithm uses ϵ-approximate **1**-energy-PCA as $\mathcal{O}_{\text{1-PCA}}$ subroutine, then it outputs an ϵ-approximate k-energy-PCA.

- Deflation is *lossless* for energy approximation!
- Application: Stronger results for outlier-robust PCA, heavy-tailed PCA
- Three-line proof using induction

$$\sum_{i=1}^{k+1} u_i^\top \mathbf{M} u_i \geq (1-\epsilon) \left(\sum_{i=1}^{k} \lambda_i(\mathbf{M}) \right) + u_{k+1}^\top \mathbf{M} u_{k+1}$$

$\sum_{i=1}^{k} u_i^\top \mathbf{M} u_i \geq (1-\epsilon) \sum_{i=1}^{k} \lambda_i(\mathbf{M})$

Our result for energy-PCA

Theorem: [JKLPPT24]

If the deflation algorithm uses ϵ-approximate **1**-energy-PCA as $\mathcal{O}_{\text{1-PCA}}$ subroutine, then it outputs an ϵ-approximate k-energy-PCA.

- Deflation is *lossless* for energy approximation!
- Application: Stronger results for outlier-robust PCA, heavy-tailed PCA
- Three-line proof using induction

$$
\begin{aligned}
\sum_{i=1}^{k+1} u_i^\top \mathbf{M} u_i &\geq (1-\epsilon) \left(\sum_{i=1}^{k} \lambda_i(\mathbf{M}) \right) + u_{k+1}^\top \mathbf{M} u_{k+1} \\
&\geq \qquad \dots \qquad + (1-\epsilon) \cdot \lambda_1 \left(k\text{-deflated version of } \mathbf{M} \right)
\end{aligned}
$$

$\sum_{i=1}^{k} u_i^\top \mathbf{M} u_i \geq (1-\epsilon) \sum_{i=1}^{k} \lambda_i(\mathbf{M})$

Our result for energy-PCA

Theorem: [JKLPPT24]

If the deflation algorithm uses ϵ-approximate 1-energy-PCA as $\mathcal{O}_{\text{1-PCA}}$ subroutine, then it outputs an ϵ-approximate k-energy-PCA.

- Deflation is *lossless* for energy approximation!
- Application: Stronger results for outlier-robust PCA, heavy-tailed PCA
- Three-line proof using induction

$$\sum_{i=1}^{k+1} u_i^\top \mathbf{M} u_i \geq (1-\epsilon)\left(\sum_{i=1}^{k} \lambda_i(\mathbf{M})\right) + u_{k+1}^\top \mathbf{M} u_{k+1}$$

$$\geq \quad \dots \quad + (1-\epsilon) \cdot \lambda_1\left(k\text{-deflated version of } \mathbf{M}\right)$$

$$\geq \quad \dots \quad + (1-\epsilon) \cdot \lambda_{k+1}(\mathbf{M})$$

$\sum_{i=1}^{k} u_i^\top \mathbf{M} u_i \geq (1-\epsilon) \sum_{i=1}^{k} \lambda_i(\mathbf{M})$

Our result for energy-PCA

Theorem: [JKLPPT24]

If the deflation algorithm uses ϵ-approximate 1-energy-PCA as $\mathcal{O}_{\text{1-PCA}}$ subroutine, then it outputs an ϵ-approximate k-energy-PCA.

- Deflation is *lossless* for energy approximation!
- Application: Stronger results for outlier-robust PCA, heavy-tailed PCA
- Three-line proof using induction

$$\sum_{i=1}^{k+1} u_i^\top \mathbf{M} u_i \geq (1-\epsilon)\left(\sum_{i=1}^{k} \lambda_i(\mathbf{M})\right) + u_{k+1}^\top \mathbf{M} u_{k+1}$$

$$\geq \quad \dots \quad + (1-\epsilon) \cdot \lambda_1\left(k\text{-deflated version of } \mathbf{M}\right)$$

$$\geq \quad \dots \quad + (1-\epsilon) \cdot \lambda_{k+1}(\mathbf{M}) = (1-\epsilon) \sum_{i=1}^{k+1} \lambda_i(\mathbf{M})$$

Approximation notion: correlation

- Geometric notion: output has low correlation with low eigendirections

An orthonormal matrix $\mathbf{U} \in \mathbb{R}^{d\times k}$ is a (Δ, Γ)-approximate k-**correlation**-PCA of a PSD matrix $\mathbf{M} \in \mathbb{R}^{d\times d}$ if

Approximation notion: correlation

- Geometric notion: output has low correlation with low eigendirections

An orthonormal matrix $\mathbf{U} \in \mathbb{R}^{d\times k}$ is a (Δ, Γ)-approximate k-**correlation**-PCA of a PSD matrix $\mathbf{M} \in \mathbb{R}^{d\times d}$ if

$$\left\| \left(\mathbf{V}^{<(1-\Gamma)\lambda_k}\right)^\top \mathbf{U} \right\|_{\mathrm{Fr}}^2 \leq \Delta,$$

where $\mathbf{V}^{<(1-\Delta)\lambda_k}$ is the orthonormal matrix of eigenvectors of $\mathbf{M}$ with eigenvalues less than $(1-\Gamma)\lambda_k$.

$\sum_{i=1}^{k} u_i^\top \mathbf{M} u_i \geq (1-\epsilon) \sum_{i=1}^{k} \lambda_i(\mathbf{M})$

Approximation notion: correlation

- Geometric notion: output has low correlation with low eigendirections

An orthonormal matrix $\mathbf{U} \in \mathbb{R}^{d \times k}$ is a (Δ, Γ)-approximate k-**correlation**-PCA of a PSD matrix $\mathbf{M} \in \mathbb{R}^{d \times d}$ if

$$\left\| \left(\mathbf{V}^{<(1-\Gamma)\lambda_k} \right)^\top \mathbf{U} \right\|_{\mathrm{Fr}}^2 \leq \Delta,$$

where $\mathbf{V}^{<(1-\Delta)\lambda_k}$ is the orthonormal matrix of eigenvectors of $\mathbf{M}$ with eigenvalues less than $(1-\Gamma)\lambda_k$.

- (Relation with energy PCA) Up to some **loss in parameters**,
  - 1-correlation-PCA $\Longrightarrow$ 1-energy-PCA
  - k-energy-PCA $\Longrightarrow$ k-correlation-PCA

Approximation notion: correlation

- Geometric notion: output has low correlation with low eigendirections

An orthonormal matrix $\mathbf{U} \in \mathbb{R}^{d\times k}$ is a (Δ,Γ)-approximate k-**correlation**-PCA of a PSD matrix $\mathbf{M} \in \mathbb{R}^{d\times d}$ if

$$\left\| \left(\mathbf{V}^{<(1-\Gamma)\lambda_k}\right)^\top \mathbf{U} \right\|_{\mathrm{Fr}}^2 \leq \Delta,$$

where $\mathbf{V}^{<(1-\Delta)\lambda_k}$ is the orthonormal matrix of eigenvectors of $\mathbf{M}$ with eigenvalues less than $(1-\Gamma)\lambda_k$.

- (Relation with energy PCA) Up to some **loss in parameters**,
  - 1-correlation-PCA $\implies$ 1-energy-PCA
  - k-energy-PCA $\implies$ k-correlation-PCA
- Our energy-PCA result $\implies$ deflation performs **(lossy)** correlation-PCA

Approximation notion: correlation

- Geometric notion: output has low correlation with low eigendirections

An orthonormal matrix $\mathbf{U} \in \mathbb{R}^{d\times k}$ is a (Δ, Γ)-approximate k-**correlation**-PCA of a PSD matrix $\mathbf{M} \in \mathbb{R}^{d\times d}$ if

$$\left\| \left(\mathbf{V}^{<(1-\Gamma)\lambda_k}\right)^{\top} \mathbf{U} \right\|_{\mathrm{Fr}}^2 \leq \Delta,$$

where $\mathbf{V}^{<(1-\Delta)\lambda_k}$ is the orthonormal matrix of eigenvectors of $\mathbf{M}$ with eigenvalues less than $(1-\Gamma)\lambda_k$.

- (Relation with energy PCA) Up to some **loss in parameters**,
  - 1-correlation-PCA $\implies$ 1-energy-PCA
  - k-energy-PCA $\implies$ k-correlation-PCA
- Our energy-PCA result $\implies$ deflation performs **(lossy)** correlation-PCA

Can deflation avoid this parameter loss?

Our result for correlation-PCA

- Suppose $\mathcal{O}_{\text{1-PCA}}$ in the deflation algorithm is a (δ, γ)-approximate PCA.
- **Key question:** How large can δ and γ be while ensuring that deflation outputs a (Δ, Γ)-approximate PCA?[1]

[1] For the simplicity of this talk, we assume $\lambda_1 \asymp \lambda_k$.

Our result for correlation-PCA

- Suppose $\mathcal{O}_{\text{1-PCA}}$ in the deflation algorithm is a (δ, γ)-approximate PCA.
- **Key question:** How large can δ and γ be while ensuring that deflation outputs a (Δ, Γ)-approximate PCA?[1]
  - [AL16] $\gamma = O(\Gamma)$ but $\delta = O_k(\Delta^2\Gamma^2)$
  - (Using relation with energy-PCA) $\delta = \gamma = O_k(\Delta\Gamma)$
  - Is there a **lossless** guarantee? By lossless, we mean $\delta \propto \Delta$ and $\gamma \propto \Gamma$

[1] For the simplicity of this talk, we assume $\lambda_1 \asymp \lambda_k$.

[AL16] Z. Allen Zhu, Y. Li. LazySVD: Even Faster SVD Decomposition Yet Without Agonizing Pain. *NeurIPS*. 2016

Our result for correlation-PCA

- Suppose $\mathcal{O}_{\text{1-PCA}}$ in the deflation algorithm is a (δ, γ)-approximate PCA.
- **Key question:** How large can δ and γ be while ensuring that deflation outputs a (Δ, Γ)-approximate PCA?[1]
  - [AL16] $\gamma = O(\Gamma)$ but $\delta = O_k(\Delta^2\Gamma^2)$
  - (Using relation with energy-PCA) $\delta = \gamma = O_k(\Delta\Gamma)$
  - Is there a **lossless** guarantee? By lossless, we mean $\delta \propto \Delta$ and $\gamma \propto \Gamma$

Theorem: Informal [JKLPPT24]

[1] For the simplicity of this talk, we assume $\lambda_1 \asymp \lambda_k$.

[AL16] Z. Allen Zhu, Y. Li. LazySVD: Even Faster SVD Decomposition Yet Without Agonizing Pain. *NeurIPS*. 2016

Our result for correlation-PCA

- Suppose $\mathcal{O}_{\text{1-PCA}}$ in the deflation algorithm is a (δ, γ)-approximate PCA.
- **Key question:** How large can δ and γ be while ensuring that deflation outputs a (Δ, Γ)-approximate PCA?[1]
  - [AL16] $\gamma = O(\Gamma)$ but $\delta = O_k(\Delta^2\Gamma^2)$
  - (Using relation with energy-PCA) $\delta = \gamma = O_k(\Delta\Gamma)$
  - Is there a **lossless** guarantee? By lossless, we mean $\delta \propto \Delta$ and $\gamma \propto \Gamma$

Theorem: Informal [JKLPPT24]

- **(Lossless)** If $\Delta = O(\Gamma^2)$, then can take $\delta = \Theta_k(\Delta)$, $\gamma = \Theta_k(\Gamma)$.

[1] For the simplicity of this talk, we assume $\lambda_1 \asymp \lambda_k$.

[AL16] Z. Allen Zhu, Y. Li. LazySVD: Even Faster SVD Decomposition Yet Without Agonizing Pain. *NeurIPS*. 2016

Our result for correlation-PCA

- Suppose $\mathcal{O}_{\text{1-PCA}}$ in the deflation algorithm is a (δ, γ)-approximate PCA.
- **Key question:** How large can δ and γ be while ensuring that deflation outputs a (Δ, Γ)-approximate PCA?[1]
  - [AL16] $\gamma = O(\Gamma)$ but $\delta = O_k(\Delta^2\Gamma^2)$
  - (Using relation with energy-PCA) $\delta = \gamma = O_k(\Delta\Gamma)$
  - Is there a **lossless** guarantee? By lossless, we mean $\delta \propto \Delta$ and $\gamma \propto \Gamma$

Theorem: Informal [JKLPPT24]

- **(Lossless)** If $\Delta = O(\Gamma^2)$, then can take $\delta = \Theta_k(\Delta)$, $\gamma = \Theta_k(\Gamma)$.

- Dependence on k can likely be improved (currently quasipolynomial)

[1] For the simplicity of this talk, we assume $\lambda_1 \asymp \lambda_k$.

[AL16] Z. Allen Zhu, Y. Li. LazySVD: Even Faster SVD Decomposition Yet Without Agonizing Pain. *NeurIPS*. 2016

Our result for correlation-PCA

- Suppose $\mathcal{O}_{\text{1-PCA}}$ in the deflation algorithm is a (δ, γ)-approximate PCA.
- **Key question:** How large can δ and γ be while ensuring that deflation outputs a (Δ, Γ)-approximate PCA?[1]
  - [AL16] $\gamma = O(\Gamma)$ but $\delta = O_k(\Delta^2\Gamma^2)$
  - (Using relation with energy-PCA) $\delta = \gamma = O_k(\Delta\Gamma)$
  - Is there a **lossless** guarantee? By lossless, we mean $\delta \propto \Delta$ and $\gamma \propto \Gamma$

Theorem: Informal [JKLPPT24]

- **(Lossless)** If $\Delta = O(\Gamma^2)$, then can take $\delta = \Theta_k(\Delta)$, $\gamma = \Theta_k(\Gamma)$.
- **(Lossy)** If $\Delta = \Omega(\Gamma^2)$, then deflation can be lossy even for $k = 2$.

- Dependence on k can likely be improved (currently quasipolynomial)

[1] For the simplicity of this talk, we assume $\lambda_1 \asymp \lambda_k$.

[AL16] Z. Allen Zhu, Y. Li. LazySVD: Even Faster SVD Decomposition Yet Without Agonizing Pain. *NeurIPS*. 2016

Our result for correlation-PCA

- Suppose $\mathcal{O}_{\text{1-PCA}}$ in the deflation algorithm is a (δ, γ)-approximate PCA.
- **Key question:** How large can δ and γ be while ensuring that deflation outputs a (Δ, Γ)-approximate PCA?[1]
  - [AL16] $\gamma = O(\Gamma)$ but $\delta = O_k(\Delta^2\Gamma^2)$
  - (Using relation with energy-PCA) $\delta = \gamma = O_k(\Delta\Gamma)$
  - Is there a **lossless** guarantee? By lossless, we mean $\delta \propto \Delta$ and $\gamma \propto \Gamma$

Theorem: Informal [JKLPPT24]

- **(Lossless)** If $\Delta = O(\Gamma^2)$, then can take $\delta = \Theta_k(\Delta)$, $\gamma = \Theta_k(\Gamma)$.
- **(Lossy)** If $\Delta = \Omega(\Gamma^2)$, then deflation can be lossy even for $k = 2$.

- Dependence on k can likely be improved (currently quasipolynomial)
- Characterizes the regime of lossless deflation

[1] For the simplicity of this talk, we assume $\lambda_1 \asymp \lambda_k$.

[AL16] Z. Allen Zhu, Y. Li. LazySVD: Even Faster SVD Decomposition Yet Without Agonizing Pain. *NeurIPS*. 2016

Conclusion

- We studied blackbox reduction from k-pca to 1-pca
- Studied two notions of approximate PCA
  - energy-PCA: proved that deflation is lossless
  - correlation-PCA: characterized the regime when deflation is lossless
- Applied these reductions to get improved algorithms for outlier-robust, heavy-tailed settings
- Open questions:
  - Dependence on k in correlation-PCA
  - Other notions of approximation
  - Further applications of this framework

Thank You!