# Black-Box k-to-1 PCA Reductions

Arun Jambulapati  Syamantak Kumar  Jerry Li
Shourya Pandey*  Ankit Pensia  Kevin Tian

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

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

- Ubiquitous in statistical estimation, dimensionality reduction
- Indirect access to $\boldsymbol{\Sigma}$ (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
- $\mathbf{M}$, 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?
- **Challenge**: How do approximation errors compound?

## Approximation notion: energy

An orthonormal matrix $\mathbf{U} = (u_1, \ldots, u_k) \in \mathbb{R}^{d\times k}$ is an $\epsilon$-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?**

**Theorem:** [JKLPPT24]
If the deflation algorithm uses $\epsilon$-approximate 1-energy-PCA as $\mathcal{O}_{\text{1-PCA}}$ subroutine, then it outputs an $\epsilon$-approximate $k$-energy-PCA.

*Proof.* We proceed by induction on $i \in [k]$; for disambiguation let $\mathbf{U}_i \in \mathbb{R}^{d\times i}$ denote the horizontal concatenation of the first $i$ calls to $\mathcal{O}_{1\text{PCA}}$, so that $\mathbf{P}_i = \mathbf{I}_d - \mathbf{U}_i\mathbf{U}_i^\top$. Observe that

$$\begin{aligned}\operatorname{Tr}\left(\mathbf{U}_{i+1}^\top \mathbf{M}\mathbf{U}_{i+1}\right) &= \operatorname{Tr}\left(\mathbf{U}_i^\top\mathbf{M}\mathbf{U}_i\right) + u_{i+1}^\top \mathbf{M} u_{i+1}\\ &\geq (1-\epsilon)\|\mathbf{M}\|_i + (1-\epsilon)\|\mathbf{P}_i\mathbf{M}\mathbf{P}_i\|_{\text{op}}\\ &\geq (1-\epsilon)\|\mathbf{M}\|_i + (1-\epsilon)\sigma_{i+1}(\mathbf{M}) = (1-\epsilon)\|\mathbf{M}\|_{i+1}.\end{aligned}$$

## 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 $(\Delta,\Gamma)$-approximate $k$-**correlation**-PCA of a PSD matrix $\mathbf{M} \in \mathbb{R}^{d\times d}$ if

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

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?**

**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)

## Application: robust and heavy-tailed $k$-PCA

**Theorem (Robust heavy-tailed $k$-ePCA)**
Let $p \geq 4$, and let $\mathcal{D}$ be a 2-to-$p$ hypercontractive on $\mathbb{R}^d$ with mean $\mathbb{0}_d$ and covariance $\boldsymbol{\Sigma}$. Let $\epsilon \in (0,\epsilon_0)$, $\delta \in (0,1)$, and $\gamma = \Theta(\epsilon^{1-\frac{2}{p}})$ such that $\gamma \in (0,\gamma_0)$ for absolute constants $\epsilon_0, \gamma_0$. Let $T$ be an $\epsilon$-corrupted set of samples from $\mathcal{D}$ with $|T| = \Theta(\beta(\frac{d\log d + \log(1/\delta)}{\gamma^2}))$ for an appropriate constant, where $\beta := C_p^6 \epsilon^{-\frac{2}{p}}$. Then, Algorithm $\mathcal{A}_k$ with inputs $T$, $\epsilon$, $\gamma$, $\delta$, and $k$ outputs orthonormal $\mathrm{U} \in \mathbb{R}^{d\times k}$ such that with probability $\geq 1-\delta$, U is a $\gamma$-$k$-ePCA of $\boldsymbol{\Sigma}$. The algorithm takes time $O(\frac{ndk}{\gamma^2}\text{polylog}(\frac{d}{\epsilon\delta}))$.

**Theorem (Heavy-tailed $k$-cPCA)**
Let $p \geq 4$ and $\beta \in (0,1)$ be reals. Let $\mathcal{D}$ be a 2-to-$p$ hypercontractive on $\mathbb{R}^d$ with mean $\mathbb{0}_d$ and covariance $\boldsymbol{\Sigma}$. Let $(\Delta,\Gamma) \in (0,1)$ such that $\Delta \cdot \kappa_k(\boldsymbol{\Sigma})^2 \leq \Gamma^2$. Set $\delta = k^{-\Theta(\log k)} \cdot \Delta$ and $\gamma = \Theta(k^{-3}) \cdot \Gamma$ for appropriate constants. Define the quantities

$$\alpha := \left(\frac{C_p^2 \kappa_k(\boldsymbol{\Sigma})\sqrt{k}}{\Gamma\sqrt{\Delta}}\right)^{\frac{1}{p-2}}, \quad R := \Theta(\alpha \operatorname{Tr}(\boldsymbol{\Sigma})).$$

If $n = \Theta(\frac{\alpha d \kappa_k(\boldsymbol{\Sigma})^2}{\delta\gamma^2}\log(\frac{d}{\beta}))$, then the deflation algorithm run with Oja as the 1-PCA oracle returns a $(\Delta,\Gamma)$-$k$-cPCA of $\boldsymbol{\Sigma}$ with probability $\geq 1-\beta$. The algorithm takes time $O(ndk)$.