Oja's Algorithm for Streaming PCA: Spectral Guarantees for Sparse Matrices

William Guo (UPenn), Advisor: Erik Waingarten (UPenn)

Background

- Principal component analysis (PCA): given a matrix $X \in \mathbb{R}^{n \times d}$, want to find the top eigenvector of its covariance matrix $\frac{1}{n} X^\top X$
- Adversarial streaming setting: given arbitrary data points $x_1, \ldots, x_n$ in a stream, want to approximate $\hat{v}_n$ s.t. $|\langle \hat{v}_n, v_* \rangle| \approx 1$ using $\tilde{O}(d)$ space
- Oja's algorithm: start with random unit vector v_0, updating with learning ruel $v_{i+1} = v_i + \eta x_{i+1} x_{i+1}^\top v_i$
- Want to bound performance in adversarial streams with a logarithmic spectral ratio $R = \lambda_1/\lambda_2 = \sigma_1/\sigma_2$

Algorithm

We analyze the modified version of Oja's algorithm presented in Price & Xun 2024, summarized here:

Algorithm 1 OjaCheckingGrowth - checks if η is too small of a learning rate

Initialize $\hat{v}_0 \leftarrow S^{d-1}$ uniformly at random
for $i = 1$ to n **do**
  Perform Oja's update: $v_{i+1} = v_i + \eta x_{i+1} x_{i+1}^\top v_i$
if $||v_n|| \leq d^{10}$ **then return** $\perp$
else return $\hat{v}_n$

Algorithm 2 AdversarialPCA - full algorithm

Let $b = O(\log nd)$ be the number of bits needed to express each matrix entry X_{ij}
Let $\eta_i = 2^i$, for each $|i| \leq O(b)$
Run OjaCheckingGrowth for each η_i in parallel; simultaneously track $\overline{x} = \arg\max ||x_i||$
Let i^* be the smallest i on which OjaCheckingGrowth outputs $v^{(i)} \neq \perp$
if $\eta^{(i^*)} ||\overline{x}|| > 1$ **then return** $\frac{\overline{x}}{||\overline{x}||}$
else return $\hat{v}^{(i)}$

References

[1] Eric Price, Zhiyang Xun. "Spectral Guarantees for Adversarial Streaming PCA". In *FOCS*, 2024.

[2] Praneeth Kacham, David P. Woodruff. "Approximating the Top Eigenvector in Random Order Streams". In *NeurIPS*, 2024.

Lower Bound

Theorem (lower bound, informal)

For any sufficiently small constant $C > 0$ and spectral ratio $R < C \log d$, there exists an instance on which AdversarialPCA fails to output a suitable $\hat{v}_n$ with high probability.

Proof Overview: Our instance $X \in \mathbb{R}^{n \times d}$ where $n = R + 2$ notably includes $R - 1$ copies of $\frac{1}{\sqrt{R}} e_1$ immediately followed by the row $\frac{1}{\sqrt{R}} e_1 + \frac{1}{\sqrt{R}} e_2$

$$X = \frac{1}{\sqrt{R}} \begin{pmatrix} 1 & -1 & 0 & \ldots & 0 \\ 1 & 0 & 0 & \ldots & 0 \\ \vdots & & & & \\ 1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 0 & \ldots & 0 \\ 0 & 0 & \sqrt{3} & \ldots & 0 \end{pmatrix}$$

- Principal component: e_1 (x_1 ensures symmetry of stream)
- Spectral ratio: $\approx R/3$; max norm vector: x_{R+2}, uncorrelated with e_1
- $|\langle v_n, e_1 \rangle| \simeq (1 + \frac{\eta}{R})^R$, $|\langle v_n, e_2 \rangle| \simeq \frac{\eta}{R}(1 + \frac{\eta}{R})^R$

Intuition:

- η must be large enough for $\langle v_n, e_1 \rangle$ to grow more than a factor of $poly(d)$, so $\eta = \Omega(\log d)$
- Growth in direction e_1 also benefits non-principal directions e_2 (and then e_3), so necessarily $\eta < R$.
- For sufficiently small $R = O(\log d)$, this is a contradiction.

Combined with the upper bound, this spectral ratio requirement is tight (up to a constant factor) for streams with row-sparsity $s = O(1)$.

Upper Bound

We assume $X = MQ^\top$, for orthogonal $Q \in \mathbb{R}^{d \times d}$, and $M \in \mathbb{R}^{n \times d}$ has at most s nonzero entries per row.

Theorem (upper bound)

Given spectral ratio $R = \Omega(s \log d)$, AdversarialPCA outputs $\hat{v}_n$ satisfying $\langle \hat{v}_n, v_* \rangle^2 \geq 1 - O(\frac{\log d}{R})$ with high probability.

Proof Overview: We show the growth's "error" term is $O(\sigma_1)$:

$$\log ||v_n||^2 \geq \eta \sum_{i=1}^{n} \langle x_i, v_{i-1} \rangle^2 \geq \frac{1}{4}\sigma_1 - \eta \sum_{i=1}^{n} \langle x_i, P\hat{v}_{i-1} \rangle^2$$

Here, $P = I - v_* v_*^\top$ projects away from the principal component.

Intuition: Consider when each data point x_i has a nonzero contribution to only 1 non-principal direction w_j, and when $\{1, \ldots, n\}$ can be partitioned into contiguous disjoint sets $S_j = \{i_{j-1} + 1, \ldots, i_j\}$ containing the points x_i contributing to non-principal direction w_j.

We now claim $\eta \sum_{i \in S_j} \langle x_i, P\hat{v}_{i-1} \rangle^2 \leq \sigma_2^2 \frac{||v_{i_j}||^2 - ||v_{i_{j-1}}||^2}{||v_{i_j}||^2}$, using that $\langle x_i, P\hat{v}_{i-1} \rangle^2 = \langle x_i, w_j \rangle^2 \langle \hat{v}_{i-1}, w_j \rangle^2$ and that $\langle \hat{v}_{i-1}, w_j \rangle^2$ is bounded by the fraction of growth from indices in S_j and the total growth thus far.

Since $1 - \frac{1}{x} \leq \log x$, summing this across all j telescopes to $\sigma_2^2 \log ||v_n||^2$.

Future Work

- Improve spectral ratio upper bound for general matrices; unclear whether $O(\log d)$ spectral ratio is obtainable. Note Kacham & Woodruff 2024 showed $O(\log^2 d)$ is obtainable by combining AdversarialPCA and row-norm sketching
- Extend upper/lower bounds to Oja's algorithm for top k principal components