Computing Systems For Data-Driven Science

Data Science Institute

Non-Stationary Streaming PCA

Daniel Bienstock, Apurv Shukla, Se-Young Yun

Columbia IEOR, KAIST ISE

Data Science Institute
COLUMBIA UNIVERSITY

Objective

Observe $\{x_i\}_{i=1}^{T} \in \mathbb{R}^p$, recover top-$k$ ($< p$) singular vectors of the underlying subspace

Streaming PCA

- Noisy and stationary environment
- Optimal storage and computation requirements

Noisy Power Method

- Power method with noisy observations [1]
- Algorithm:
  - Observe a block B of data every iteration
  - Compute covariance matrix of the observed block of data
  - Multiply with orthonormal basis of previous iteration
  - Obtain an estimate of the current orthonormal basis
- Convergence: Small spectral gap and stationarity

This work

Streaming PCA with noise and non-stationarity

Frequent Directions

- Count-based sketching algorithm for computing prominent singular vectors [2]
- Algorithm for computing top-k singular vectors
  - Maintain $2k$ columns among which k are empty at the beginning of every iteration
  - Assign incoming columns to the empty columns
  - Hard unweighted thresholding of singular values to sketch top-k singular vectors and obtain k-empty columns

Key Idea

Noisy power method + Frequent directions^{++}

Applications

- Portfolio Optimization, Market Structure, Grid Operations, Econometrics, Genomics

Non-Stationary Streaming PCA

- Observe $\{x_i\}_{i=1}^{T}$ and recover underlying subspace by performing computations on at most B vectors
- Spiked Covariance Model [3]: $x_t = Az_t + w_t$

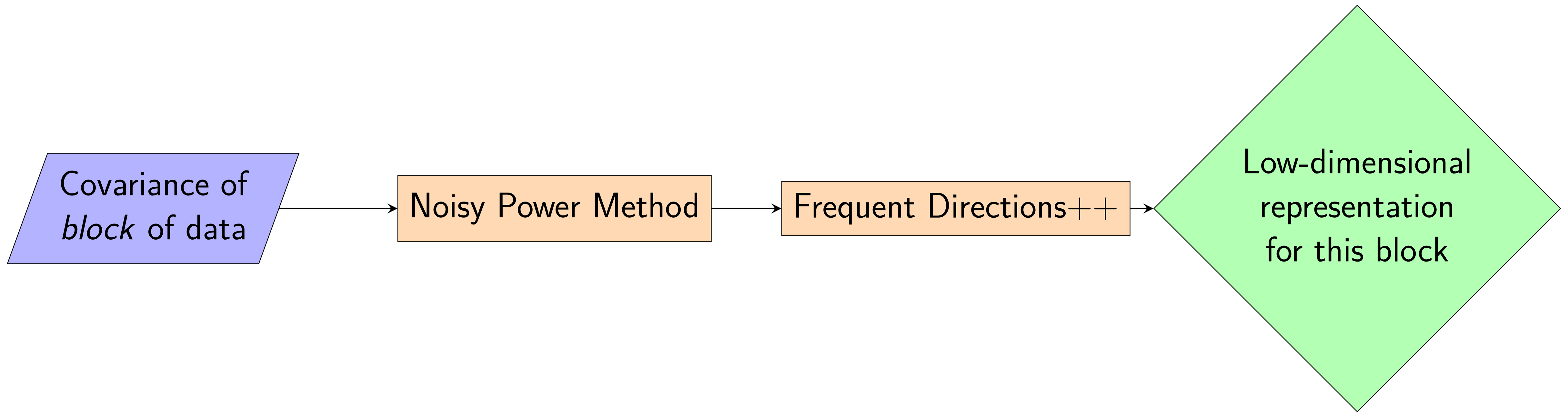
Non-Stationary Spiked Covariance Model

$$x_t = A_t z_t + w_t,\ \|A_t A_t^\top - A_{t-1}A_{t-1}^\top\| \le \gamma,\ \gamma < 1$$

- $A_t \in \mathbb{R}^{p\times k}, z_i \sim \mathcal{N}(0_k, \mathbf{I}_{k\times k}), w_i \sim \mathcal{N}(0_p, \mathbf{I}_{p\times p}), SVD(A_t) = U_t\Sigma_t V_t^\top$
- z_i and w_i are mutually independent of each other and across time

Incorporating Non-stationarity

Exponential smoothing of observed subspaces via sketching and singular value thresholding



Key Results

Exponential Smoothing

Average subspace of matrices is spanned by left singular vectors of sum of corresponding projection matrices

Frequent Directions++

Maintain sketch of singular vectors through weighted thresholding of singular values every iteration

Convergence behaviour

Analysis of convergence behaviour of the proposed algorithm in presence of non-stationarity and noise

Recovery Error

- Distance between recovered and true subspace
- Recovery error decreases to $\gamma^{1/3}$ as $\frac{1}{\sqrt{N}}$ when $N < \gamma^{-2/3}$
- Recovery does not decrease beyond $\gamma^{1/3}$ when $N > \gamma^{-2/3}$

Future Work

- Application of Oja's Algorithm
- Sequential Hypothesis Tests
- Determination of γ

References

[1] Moritz Hardt and Eric Price.
The noisy power method: A meta algorithm with applications.
In *Advances in Neural Information Processing Systems*, pages 2861–2869, 2014.

[2] Mina Ghashami, Edo Liberty, Jeff M Phillips, and David P Woodruff.
Frequent directions: Simple and deterministic matrix sketching.
SIAM Journal on Computing, 45(5):1762–1792, 2016.

[3] Iain M Johnstone.
On the distribution of the largest eigenvalue in principal components analysis.
Annals of statistics, pages 295–327, 2001.

Contact Information

- E-mail: Apurv: apurv.shukla@columbia.edu
  Se-Young: yunseyoung@kaist.ac.kr