Non-Stationary Streaming PCA

Daniel Bienstock, Apurv Shukla, Se-Young Yun

Columbia IEOR, KAIST ISE

Objective

Observe \( \{x_i\}_{i=1}^N \in \mathbb{R}^p \), 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
  - 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 \( k \) 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++

Key Results

Non-Stationary Streaming PCA

- Observe \( \{x_i\}_{i=1}^N \) and recover underlying subspace by performing computations on at most \( B \) vectors
- Spiked Covariance Model [3]: \( x_i = A_i z_i + w_i \)
- \( A_i \in \mathbb{R}^{p \times k} \), \( z_i \sim \mathcal{N}(0_k, I_{k \times k}) \), \( w_i \sim \mathcal{N}(0_p, I_{p \times p}) \), \( \text{SV}D(A_i) = U_i \Lambda_i V_i^\top \)
- Incorporating Non-stationarity
  - Exponential smoothing of observed subspaces via sketching and singular value thresholding
- Low-dimensional representation for this block

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 \( \gamma \)

References

  The noisy power method: A meta algorithm with applications.

  Frequent directions: Simple and deterministic matrix sketching.

  On the distribution of the largest eigenvalue in principal components analysis.

Contact Information

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