Topic

Applications of Low-Rank Updating Singular Value Decomposition in Incremental Principal Component Analysis (PCA)

Members

C14125018 蔡宇德

C14126111 陳柏諭

C14121048 鄭丞佑

C14121145 陳柏任

2. Introduction

a. Motivation

In many modern applications such as image recognition, online recommendation systems, and streaming data analysis, new data arrive continuously over time. Recomputing the full Singular Value Decomposition (SVD) each time new data appear is computationally expensive: for an $m \times n$ matrix, traditional SVD requires $O(\min(m,n)^2 \max(m,n))$ operations, which becomes prohibitive for real-time systems when data accumulate incrementally [1].

To address this issue, **Low-Rank Updating SVD** provides an efficient way to update the decomposition without recomputing it from scratch. Brand [1] demonstrated that the incremental method is "orders of magnitude faster" than batch methods such as Lanczos SVD for large-scale matrices. One of the most important applications of this technique is **Incremental Principal Component Analysis (Incremental PCA)**, where the principal components of a dataset can be efficiently updated as new samples arrive. This allows PCA to be applied in real-time environments, such as face recognition systems, online anomaly detection, and adaptive signal processing, where batch processing is impractical [3].

b. Literature Survey

The concept of low-rank modification of SVD was first systematically discussed in Brand [1], which introduced efficient algorithms for updating SVD after low-rank perturbations. The paper presents both theoretical foundations and practical implementations of the thin SVD update method.

Other foundational works include:

- **Golub and Van Loan** [4] : Provides the theoretical basis for matrix factorization and numerical stability in SVD computations.

- **Levy and Lindenbaum** [5]: Presents early forms of incremental PCA using sequential Karhunen-Loeve basis extraction, demonstrating applications to image processing.

- **Ross et al.** [2]: Shows real-time visual tracking using low-rank SVD updates with mean reconstruction errors of 7.93×10^{-2} per pixel in practical tracking scenarios.

- **Zhao et al.** [3]: Provides mathematical proofs of error bounds for incremental PCA, showing that the approximation error is $O(\varepsilon)$ where ε is the truncation error in the singular value approximation.

These studies demonstrate that low-rank SVD updates are both theoretically elegant and practically powerful, enabling efficient online learning with bounded approximation error.

c. Potential Theoretical Background

This project will utilize several key mathematical concepts:

i. Singular Value Decomposition (SVD)

Any matrix $A \in \mathbb{R}^{m \times n}$ can be decomposed as [1]:

$$A = U\Sigma V^T$$

where $U \in \mathbb{R}^{m \times r}$ and $V \in \mathbb{R}^{n \times r}$ are orthogonal matrices, and $\Sigma \in \mathbb{R}^{r \times r}$ is diagonal with non-negative singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$.

ii. Low-Rank Update Formula

When new data arrive in the form of a low-rank perturbation $A' = A + BC^T$ where $B \in \mathbb{R}^{m \times k}$ and $C \in \mathbb{R}^{n \times k}$ with $k \ll \min(m, n)$, the updated SVD can be efficiently computed using Brand's algorithm [1]:

- Project the new data onto the existing subspace and its orthogonal complement
- Perform SVD on a smaller $(r + k) \times (r + k)$ matrix
- Rotate the original basis to obtain the updated decomposition

This reduces computational complexity from $O(mn\min(m, n))$ to $O((m + n)r^2 + r^3)$ per update.

iii. Incremental PCA Principle

PCA is equivalent to computing the dominant singular vectors of the data matrix. When new samples $x_{n+1}, x_{n+2}, \ldots$ arrive, the data matrix can be incrementally updated [3]:

$$X_{\text{new}} = [X_{\text{old}}, x_{\text{new}}]$$

Low-rank SVD updates allow the principal components to be adapted efficiently without full recomputation [1] [3].

iv. Error Bound Analysis

Mathematical proofs in [3] show that the approximation error between incremental and batch PCA is bounded. Specifically, if the singular value approximation error is $\varepsilon = \sigma_{k+1}$ (the first truncated singular value), then the error in estimated eigenvectors is:

$$\|v_i - w_i\| = O(\varepsilon)$$

where v_i is the incremental estimate and w_i is the true eigenvector from batch PCA. This theoretical guarantee ensures that incremental methods maintain accuracy while achieving computational efficiency [3].

3. Expected Results

This project aims to:

- **Implement** a simplified version of Incremental PCA using Low-Rank Updating SVD in MATLAB or Python (NumPy) [1] [2].

- **Demonstrate efficiency** compared with recomputing SVD from scratch. **Hypothesis**: Based on Brand's experiments [1], we expect to observe substantial computational time reduction (e.g., for a 3000×3000 matrix with rank-50 approximation, Brand reports approximately 200 seconds for incremental SVD versus 1400 seconds for batch Lanczos SVD).

- **Apply the algorithm** to the **ORL Face Database** [6] (40 subjects, 400 images total, each 92×112 pixels), showing how the principal components evolve as data increases incrementally.

- **Provide performance metrics** including:
  - **Computation time** comparison (incremental vs. full SVD)
  - **Reconstruction error** measured by mean squared error or relative Frobenius norm
  - **Number of retained components** determined by reconstruction error threshold, following the approach in [2].
  - **Error bound verification** comparing theoretical bounds with empirical observations [3].

Expected outcome: The results should demonstrate that low-rank updating SVD significantly reduces computational cost [1] while maintaining high accuracy in PCA reconstruction, with approximation errors consistent with theoretical bounds [3].

4. Gantt Chart

Task Description	Week 1	Week 2	Week 3	Week 4	Week 5
Literature review and understanding SVD updating algorithms (Brand [1], Ross et al. [2])	████████				
Implement basic SVD and low-rank update in MATLAB/Python; download and preprocess ORL Face Database [6]		████████			
Implement Incremental PCA with low-rank SVD updating; test on synthetic data			████████		
Run experiments on ORL Face Database; collect computation time, reconstruction error, and error bound metrics				████████	
Analyze results, verify error bounds, create comparison plots, and prepare final report and presentation					████████

Note: The shaded cells (████████) indicate the primary work period for each task. Tasks may have minor overlap in actual execution.

References

[1] M. Brand, Fast low-rank modifications of the thin singular value decomposition, Linear Algebra Appl. 415 (2006), no. 1, 20–30.

[2] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, Incremental learning for robust visual tracking, Int. J. Comput. Vis. 77 (2008), no. 1-3, 125–141.

[3] H. Zhao, P. C. Yuen, and J. T. Kwok, A novel incremental principal component analysis and its application for face recognition, IEEE Trans. Systems Man Cybernet. B 36 (2006), no. 4, 873–886.

[4] G. H. Golub and C. F. Van Loan, Matrix Computations, 4th ed., Johns Hopkins University Press, Baltimore, 2013.

[5] A. Levy and M. Lindenbaum, Sequential Karhunen-Loeve basis extraction and its application to images, IEEE Trans. Image Process. 9 (2000), no. 8, 1371–1374.

[6] F. S. Samaria and A. C. Harter, Parameterisation of a stochastic model for human face identification, Proc. 2nd IEEE Workshop on Applications of Computer Vision, Sarasota, FL, 1994, pp. 138–142.