Fast Exemplar Selection Algorithm for Matrix Approximation and Representation: A Variant oASIS Algorithm

Vinayak Abrol, Pulkit Sharma & Anil Kumar Sao
School of Computing & Electrical Engg., Indian Institute of Technology Mandi, India
votrix13@ieee.org, pulkitsharmay2k@ieee.org, anil@iitmandi.ac.in

Introduction 1

- Fast exemplar selection (FES) is a scalable, deterministic and computationally efficient algorithm for adaptive column sampling.
- FES extracts an incoherent subset that approximates the column span of a matrix $\mathbf{X} \in \mathbb{R}^{n \times l}$
- FES achieves this sequentially by ensuring that the sampled exemplars have a positive definite (PD) Gram matrix.
- To handle larger datasets, FES uses incremental Cholesky decomposition and block matrix inversion algorithms.

Proposed Approach 2

Problem: Aim is to sample a small number of columns of a matrix $\mathbf{X}$ such that

$$\|\mathbf{X} - \mathbf{\Pi X}\|_F = \|\mathbf{X} - \mathbf{X}_k\|_F$$

i.e., error between the target matrix $\mathbf{X}$ and its rank-k approximation $\mathbf{X}_k$

Proposed Approach: A column $\mathbf{x}_i$ from matrix $\mathbf{X}$ can be sampled based on its distance to the space spanned by the sampled set $\mathbf{X}_S$ as

$$i = \underset{i \notin S}{\text{argmax}}\, \|\mathbf{x}_i - \mathbf{\Pi}_S \mathbf{x}_i\|_2^2 = \|\mathbf{x}_i - \mathbf{X}_S \mathbf{X}_S^+ \mathbf{x}_i\|_2^2$$

Assuming columns sampled in $\mathbf{X}_S$ are independent, the above expression can be expanded as

$$\Delta_i = \mathbf{d}_i - \mathbf{a}_i^T (\mathbf{W})^{-1} \mathbf{a}_i$$

where $\mathbf{d}_i = \mathbf{x}_i^T \mathbf{x}_i$, $\mathbf{a}_i = \mathbf{X}_S^T \mathbf{x}_i$ and $\mathbf{W} = \mathbf{X}_S^T \mathbf{X}_S$. The updated Gram matrix after each selection can be computed as (assuming normalized data)

$$\mathbf{W}_{k+1} = \begin{bmatrix} \mathbf{X}_S^T \mathbf{X}_S & \mathbf{a} \\ \mathbf{a}^T & \mathbf{x}_i^T \mathbf{x}_i \end{bmatrix} = \begin{bmatrix} \mathbf{W}_k & \mathbf{a} \\ \mathbf{a}^T & 1 \end{bmatrix}$$

$\mathbf{W}$ will be invertible if it has a unique Cholesky decomposition $\mathbf{W}_k = \mathbf{L}_k \mathbf{L}_k^T$, and the updated Gram matrix can be expressed as

$$\begin{bmatrix} \mathbf{W}_k & \mathbf{a} \\ \mathbf{a}^T & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{L}_k & 0 \\ \mathbf{c}^T & d \end{bmatrix} \begin{bmatrix} \mathbf{L}_k^T & \mathbf{c}^T \\ 0 & d \end{bmatrix} = \begin{bmatrix} \mathbf{L}_k \mathbf{L}_k^T & \mathbf{L}_k \mathbf{c} \\ \mathbf{L}_k^T \mathbf{c}^T & \mathbf{c}^T \mathbf{c} + d^2 \end{bmatrix}$$

which gives us

$$\mathbf{a} = \mathbf{L}_k \mathbf{c} \text{ or } \mathbf{c} = \mathbf{L}_k^{-1} \mathbf{a} \text{ and } \quad d \quad = \sqrt{1 - \mathbf{c}^T \mathbf{c}}$$

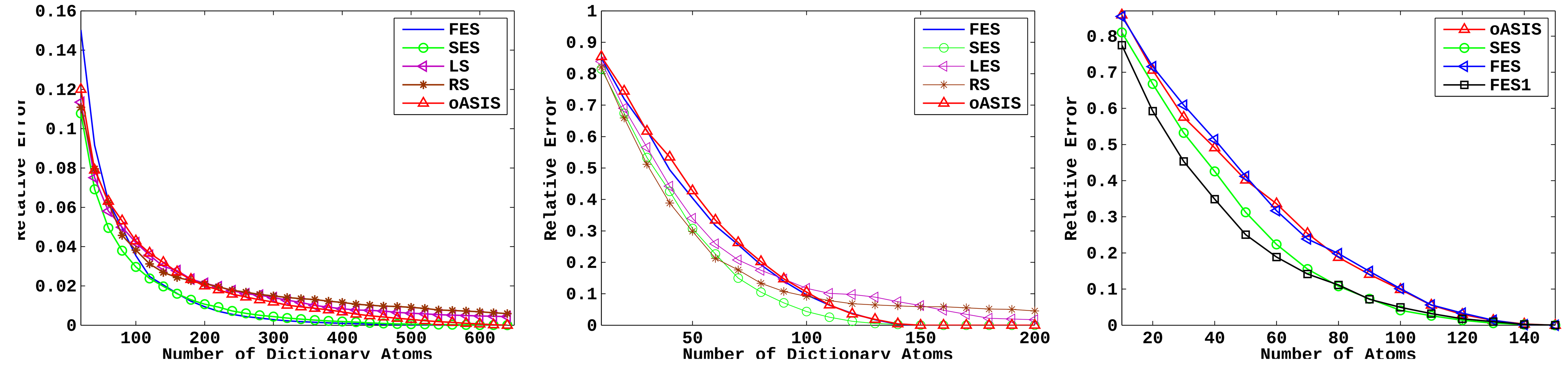
Hence, FES proposes to iteratively sample columns using the criteria $\mathbf{c}^T \mathbf{c} < 1$. This computation can be speed up via approximating $\mathbf{L}_{k+1}^{-1}$ by performing rank-1 updates to the inverse matrix $\mathbf{L}_k^{-1}$ i.e.,

$$\begin{bmatrix} \mathbf{L}_k & 0 \\ \mathbf{c}^T & d \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{L}_k^{-1} & 0 \\ -(1/d)\mathbf{c}^T \mathbf{L}_k^{-1} & 1/d \end{bmatrix}$$

Abbreviations: $\mathbf{x}$ - Signal Vector | $\mathbf{X}$ - Signal Matrix | $\mathbf{X}_S$ - Sampled Matrix | $\mathbf{\Pi}$ - Projection Matrix | $\mathbf{W}$ - Gram Matrix | $\mathbf{L}$ - Cholesky Factor | S - indexes of sampled column

Results 3

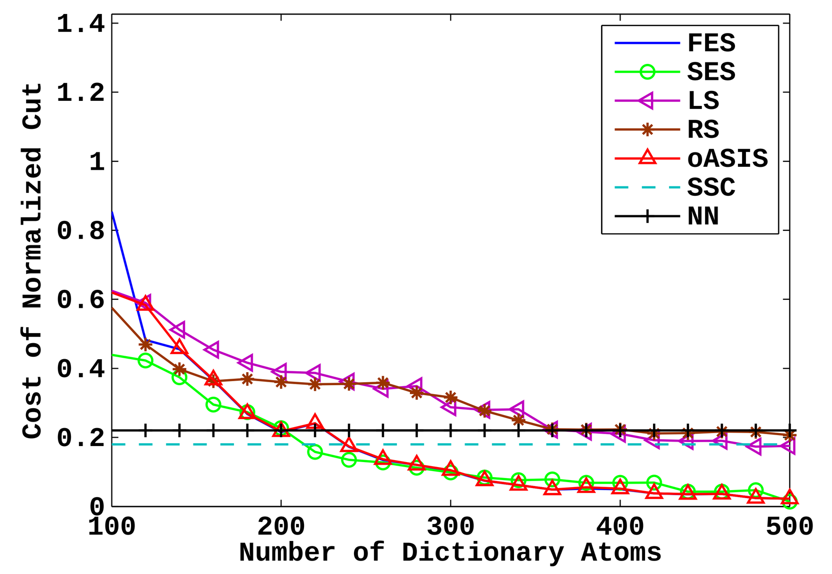
Exact Matrix Recovery



(a) (b) (c)

Approximation error vs number of dictionary atoms for (a) Yale Face, (b) and (c) UoS datasets.

Sparse Representation Based Clustering



Normalized cut ratios for clustering task on the Yale face dataset.

Other applications: Low rank approximation, optimal feature selection, outlier detection, dictionary learning, subspace clustering etc.

References 4

[1] A. Çivril, "Column subset selection problem is UG-hard," *Journal of Computer and System Sciences*, vol. 80, no. 4, pp. 849 – 859, 2014.

[2] A. Çivril and M. Magdon-Ismail, "Column subset selection via sparse approximation of {SVD}," *Theoretical Computer Science*, vol. 421, pp. 1 – 14, 2012.

[3] A. Deshpande and S. Vempala, "Adaptive sampling and fast low-rank matrix approximation," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, ser. APPROX'06/RANDOM'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 292–303.

[4] A. K. Farahat, A. Ghodsi, and M. S. Kamel, "A fast greedy algorithm for generalized column subset selection," *NIPS Workshop on Greedy Algorithms*, December 2013.

[5] R. Patel, T. A. Goldstein, E. L. Dyer, A. Mirhoseini, and R. G. Baraniuk, "oASIS: Adaptive Column Sampling for Kernel Matrix Approximation," *ArXiv e-prints*, may 2015.

Acknowledgment

DST-SERB International Travel Support Scheme
IEEE SPS International Travel Grant