An Empirical Evaluation of Sketched SVD and its Application to Leverage Score Ordering

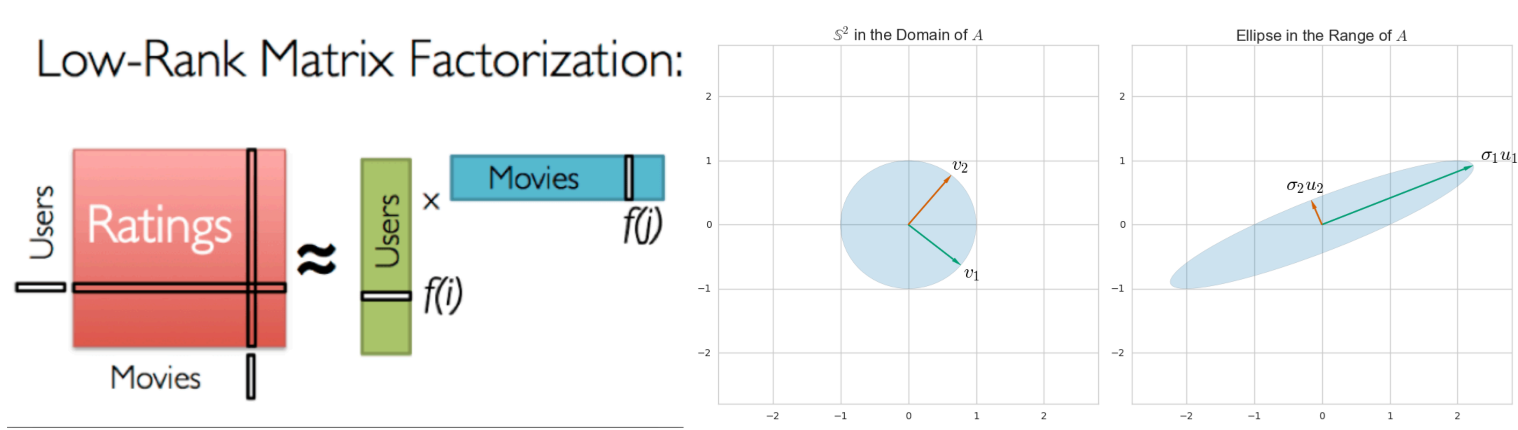
Hui Han Chin, Paul Pu Liang

School of Computer Science, Carnegie Mellon University

hhchin87@gmail.com, pliang@cs.cmu.edu

Introduction

Singular Value Decomposition (SVD): Weighted low rank approximations are used in recommendation systems and leverage scores are heavily used in statistical data analysis. Exact SVD is slow on large-scale datasets but sketching techniques can be used to approximate the SVD [5].



Our approach to the empirical evaluation of Sketched SVD and its application to leverage score ordering:

1. Evaluate sketched SVD on large-scale real world datasets.
2. Provide insights to the practical implementations of sketched SVD.
3. Outline the **Sketched Leverage Score Ordering** algorithm (SLSO).
4. Discuss the effectiveness of the SLSO algorithm on real-world datasets and models.

Preliminaries

Popular choices for the randomized sketching matrix that has the subspace embedding property: 1) Random Gaussian [5], 2) Fast Johnson Lindenstrauss transform (FJLT) [4], 3) Subsampled Randomized Hadamard Transform (SRHT) [1], 4) CountSketch [2], 5) OSNAP [3].

Definition 1. *Subspace Embedding: Let* $\mathbf{A}$ *be a* n *by* d *matrix. A* $(1 \pm \epsilon)$ $\ell 2$ *subspace embedding for the column space of* $\mathbf{A}$ *is a* k *by* n *matrix* $\mathbf{S}$ *such that* $\forall x \in \mathbb{R}^n$,

$$(1-\epsilon)\|\mathbf{A}x\|_2^2 \le \|\mathbf{SA}x\|_2^2 \le (1+\epsilon)\|\mathbf{A}x\|_2^2$$

Theorem 1. *A subspace embedding preserves the set of singular values, of the input matrix* $\mathbf{A}$*. In particular, if* $\mathbf{S}$ *is a* $(1 \pm \epsilon)$ $\ell 2$ *subspace embedding for* $\mathbf{A}$*, then:*

$$\sigma_k(\mathbf{SA}) = (1 \pm O(\epsilon))\sigma_k(\mathbf{A})$$

Leverage score of a data point is a measure of how much an outlier the data point is from other points in data $\mathbf{A}$. The underlying data model is assumed to be linear and the leverage score of the ith point, l_i, is given by ith row norm of the projection matrix of $\mathbf{A}$, i.e

$$\mathbf{H} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{+}\mathbf{A}^T, \; l_i = |[\mathbf{H}]_i|_2^2$$

$$\text{or using SVD, } \; \mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \; l_i = |\mathbf{U}_{i,*}|_2^2$$

Approximation Leverage Score by Sketching

Input : Given $n \times d$ matrix $\mathbf{A}$.
Output : Approximate Leverage score of ith row as l_i.
1. Compute Sketch of $\mathbf{A}$, $\mathbf{SA}$.
2. Compute SVD, $\mathbf{SA} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.
3. Compute $\mathbf{U}^{approx} = \mathbf{A}\mathbf{V}^T\mathbf{\Sigma}^{-1}$.
4. Compute l_i from the first D columns of the ith row $l_i = |\mathbf{U}_i^{approx}|_2^2$.

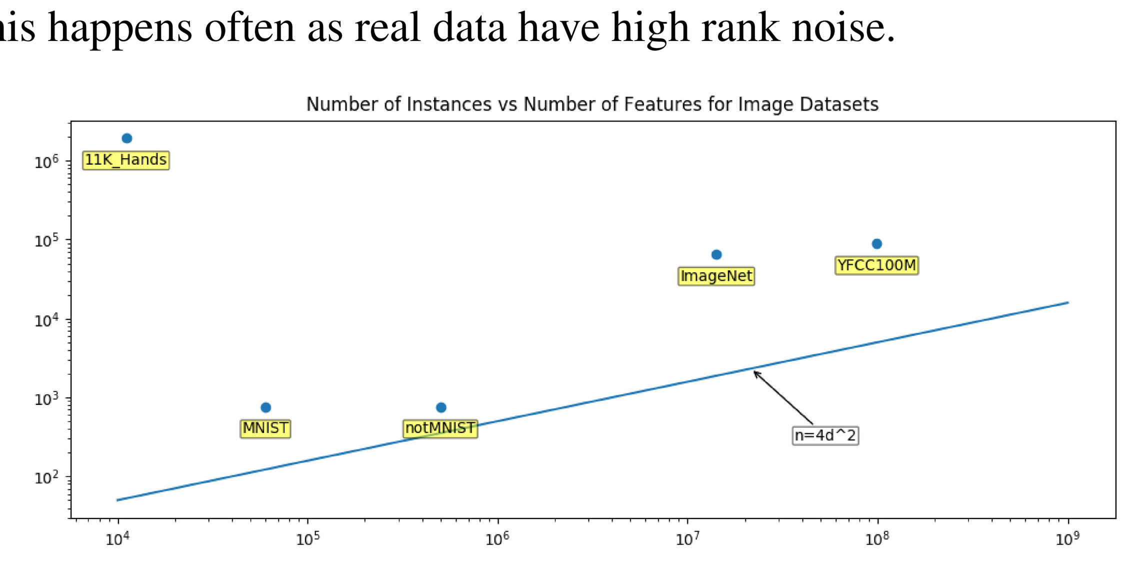
Practical Implementation

Effects of Small Singular Values on Leverage Scores Approximation

Data corrupted by high rank noise would appear as though it has full column rank. This error can cause numerical precision errors when very small singulars are inverted to compute the orthonormal basis from the sketch. i.e step 3, $\mathbf{U}^{\text{approx}} = \mathbf{A}\mathbf{V}^T\mathbf{\Sigma}^{-1}$.

Insights from Practical Implementations of Sketched Leverage Scores

1. SRHT requirement of having rows as a power of 2 was impractical. The memory and timing overheads required to satisfy that requirement is prohibitive.
2. OSNAP is more useful for real world data as it requires less rows.
3. Most curated datasets do not have enough rows (n) to satisfy the CountSketch requirements.
4. Quality of the sketch is highly dependent on the column rank of the data. If the column rank of the data is less than the column dimension, then the approximation error can be unbounded.
5. Small singular values in real data can corrupt the approximate leverage scores returned by sketching. This happens often as real data have high rank noise.



Popular image datasets for ML. None satisfy the $n \ge 4d^2$ requirement for CountSketch matrix.

Sketched Leverage Score Ordering

The approximation of leverage score using sketching is modified to truncate at the small singular values to avoid the numerical issues.

Approximate Leverage Score with Truncation

Input : Given $n \times d$ matrix $\mathbf{A}$, threshold ϵ.
Output : Approximate Leverage score of ith row as l_i.
1. Compute Sketch of $\mathbf{A}$, $\mathbf{SA}$.
2. Compute SVD, $\mathbf{SA} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.
3. Truncate $\mathbf{V}, \mathbf{\Sigma}$ at small singular values less than ϵ. Let this truncated matrices be $\mathbf{V}', \mathbf{\Sigma}'$
4. Compute $\mathbf{U}^{approx} = \mathbf{A}\mathbf{V}'^T\mathbf{\Sigma}'^{-1}$.
5. Compute l_i from the first d columns of the ith row $l_i = |\mathbf{U}_i^{approx}|_2^2$.

Sketched Leverage Score Ordering

Using ideas from curriculum learning, the sketched leverage score is use to generate sampling policies to order the training data:

1. **dec**: Ordered based on strictly decreasing leverage scores. The most important and diverse training points are seen first.
2. **dec, sampling with replacement** (**dec, swr**): Decreasing order, sampling with replacement. The most important and diverse training points are seen first, but with randomness introduced.
3. **dec, sampling without replacement** (**dec, swor**): Same but sampling without replacement.
4. **shuffle** baseline order where models are trained on shuffled training data.

Experiments and Observations

MNIST: image classification on digits handwriting. **SST:** sentiment classification. **CMU-MOSI:** multimodal sentiment analysis over 2 different runs using an early fusion method.

Task	**MNIST** 10 Class Image Classification Accuracy (%)				
Method	LR	NN Small	NN Large	CNN Small	CNN Large
shuffle	92.84	98.50	98.28	98.98	99.34
dec	89.09	98.43	98.34	**99.01**	98.99
dec, swr	92.70	98.42	98.46	**99.01**	99.35
dec, swor	**92.88**	**98.55**	**98.57**	**99.01**	**99.39**

MNIST, "decreasing order without sampling" improves performance.

Task	**SST** 5 Class Sentiment Classification Accuracy (%)				
Method	LR	DAN Small	DAN Large	LSTM Small	LSTM Large
shuffle	42.22	39.68	40.27	42.35	41.67
dec	**42.94**	39.86	**42.76**	42.35	41.31
dec, swr	41.22	**40.72**	40.72	**42.44**	**43.71**
dec, swor	42.26	39.41	40.14	41.27	41.36

SST, "decreasing order with replacement" improves performance.

Task	**CMU-MOSI** Sentiment Analysis											
Method	DAN Large				LSTM Small				LSTM Large			
Metric	A^2	F1	MAE	r	A^2	F1	MAE	r	A^2	F1	MAE	r
shuffle	61.2	59.9	1.314	0.438	73.9	74.0	1.068	0.624	73.3	73.3	1.067	0.604
dec	59.0	56.0	1.365	0.415	73.5	73.5	1.073	**0.626**	73.5	73.4	**1.038**	**0.621**
dec, swr	60.1	58.0	1.336	**0.413**	**74.6**	**74.7**	**1.061**	0.620	**74.1**	**73.9**	1.043	0.612
dec, swor	**64.3**	**64.3**	**1.271**	0.432	73.5	73.5	1.068	0.623	72.9	72.6	1.057	0.600

CMU-MOSI, "sampling in a decreasing order" improves accuracies.

Conclusion

Sketched Leverage Score Ordering, a technique for determining the ordering of data in the training of neural networks by computing leverage scores efficiently via truncated sketching. Our method shows improvements in convergence and results.

References

[1] Christos Boutsidis and Alex Gittens. Improved matrix algorithms via the subsampled randomized hadamard transform. *CoRR*, abs/1204.0062, 2012.

[2] Xiangrui Meng and Michael W. Mahoney. Low-distortion subspace embeddings in input-sparsity time and applications to robust linear regression. In *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing*, STOC '13. ACM, 2013.

[3] Jelani Nelson and Huy L. Nguyễn. Osnap: Faster numerical linear algebra algorithms via sparser subspace embeddings. In *Proceedings of the 2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, FOCS '13. IEEE CS, 2013.

[4] Tamas Sarlos. Improved approximation algorithms for large matrices via random projections. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '06, pages 143–152. IEEE Computer Society, 2006.

[5] David P. Woodruff. Sketching as a tool for numerical linear algebra. *Found. Trends Theor. Comput. Sci.*, 10:1–157, October 2014.