CS 15-859: Algorithms for Big Data Fall 2022

Lecture XX — 2022-12-19

Prof. David Woodruff *Scribe: Mike Xu*

Unless otherwise specified, all materials in this lecture note are derived from the course notes and [CEM+14, MM20].

1 Introduction to Projection-Cost-Preserving Sketches

In lectures 5 & 6, we used *projection-cost-preserving sketches* to improve the communication complexity of distributed low-rank approximation in the arbitrary partition model.

Definition 1.1 (Low Rank Approximation)**.** *Given a matrix $A \in \mathbb{R}^{n\times d}$. A low-rank approximation A' is a rank-k matrix that satisfies $\|A - A'\|_F \leq (1+\epsilon)\|A - A_k\|$, where $A_k = \text{argmin}_{\text{rank}(B)=k}\|A - B\|_F$*

Definition 1.2 (Arbitrary Partition Model)**.** *In the arbitrary partition model, matrix A is partitioned to s servers such that each server i possesses A_i, and $A = \sum_{i=1}^{s} A_i$.*

The technique presented in [BWZ15] improves the communication complexity to $O(nsk + \text{poly}(sk/\epsilon)$ from the previously known upper bound of $O(nsk/\epsilon + \text{poly}(sk/\epsilon)$.

Beyond distributed low-rank approximation, projection-cost-preserving (PCP) sketches have a wide array of applications in linear algebra, data science, and machine learning. This includes k-means clustering, streaming low-rank approximation, and other approximation algorithms in the streaming or distributed model which involves dimensionality reduction.

In short, a PCP sketch preserves the distance of a matrix A's columns to any subspace with k or fewer dimensions.

Definition 1.3 (Projection-Cost Preserving Sketches)**.** *$\tilde{A} \in \mathbb{R}^{n\times m}$ is a (ϵ, c, k)-projection-cost-preserving sketch of A if, for any orthogonal projection matrix P*

$$(1-\epsilon)\|A(I-P)\|_F^2 \leq \left\|\tilde{A}(I-P)\right\|_F^2 + c \leq (1+\epsilon)\|A(I-P)\|_F^2 \tag{1}$$

Where c is a constant independent of P.

Note that unlike the definition presented in class, these are $n \times n$ projection matrices for the column space of A, and not the rows.

Definition 1.3 is also slightly weaker than the one-sided error we saw in class, but the implication still holds that, if $\tilde{P}$ minimizes the sketch version of the problem $\left\|\tilde{A}(I-P)\right\|_F^2$, then $\left\|A(I-\tilde{P})\right\|_F^2 \in (1+O(\epsilon))\|A(I-P)\|_F^2$.

2 Proof Techniques

In [MM20], Musco and Musco present two approaches to proving that a sketch is a PCP sketch. The former involves using well-studied properties of matrices such as subspace embedding and approximate matrix product, and the latter approach uses approximations of the matrix's eigenvalues (*i.e.,* spectral approximation). Because the former is more relevant to the material discussed in class, we will focus on this method. However, it is important to note that much of the structure presented below is relevant to both techniques.

Let $Y = I - P$. For $\tilde{A} = AS$ to be a $(\epsilon, 0, k)$-PCP sketch, it suffices to show

$$| \, \|YA\|_F^2 - \|YAS\|_F^2 \, | \leq \epsilon \, \|YA\|_F^2 \tag{2}$$

We can then rewrite A as $A_k + A_{\setminus k}$, where A_k is the rank-k approximation obtained by truncating the SVD to the first k singular values and vectors. The residual is then $A_{\setminus k} = A - A_k$.

$$\begin{aligned}
& \left| \left\| YA_k + YA_{\setminus k} \right\|_F^2 - \left\| YA_kS + YA_{\setminus k}S \right\|_F^2 \right| \leq \epsilon \, \|YA\|_F^2 \\
\equiv \quad & |\operatorname{tr}((YA_k + YA_{\setminus k})(A_k^T Y^T + A_{\setminus k}^T Y^T)) && (\text{From } \|M\|_F^2 = \operatorname{tr}(MM^T)) \\
& - \operatorname{tr}((YA_kS + YA_{\setminus k}S)(S^T A_k^T Y^T + S^T A_{\setminus k}^T Y^T))| \leq \epsilon \, \|YA\|_F^2 \\
\equiv \quad & |\operatorname{tr}(YA_kA_k^T Y^T) + \operatorname{tr}(YA_{\setminus k}A_{\setminus k}^T Y^T) && (\operatorname{tr}(YA_kA_{\setminus k}^T Y^T) = 0) \\
& - \operatorname{tr}(YA_kSS^T A_k^T Y^T) - \operatorname{tr}(YA_{\setminus k}SS^T A_{\setminus k}^T Y^T) \\
& - \operatorname{tr}(YA_{\setminus k}SS^T A_k^T Y^T) - \operatorname{tr}(YA_kSS^T A_{\setminus k}^T Y^T)| \leq \epsilon \, \|YA\|_F^2
\end{aligned}$$

Therefore, it suffices to show

$$\begin{aligned}
& \left| \operatorname{tr}(YA_kA_k^T Y^T) - \operatorname{tr}(YA_kSS^T A_k^T Y^T) \right| \\
& + \left| \operatorname{tr}(YA_{\setminus k}A_{\setminus k}^T Y^T) - \operatorname{tr}(YA_{\setminus k}SS^T A_{\setminus k}^T Y^T) \right| \\
& + \left| 2 \operatorname{tr}(YA_kSS^T A_{\setminus k}^T Y^T) \right| \leq \epsilon \, \|YA\|_F^2 \\
\equiv \quad & \left| \|YA_k\|_F^2 - \|YA_kS\|_F^2 \right| && (\text{“Head term”}) \\
& + \left| \left\| YA_{\setminus k} \right\|_F^2 - \left\| YA_{\setminus k}S \right\|_F^2 \right| && (\text{“Tail term”}) \\
& + \left| 2 \operatorname{tr}(YA_kSS^T A_{\setminus k}^T Y^T) \right| && (\text{“Cross term”}) \\
& \leq \epsilon \, \|YA\|_F^2 && (3)
\end{aligned}$$

We can now bound the three terms in (3) separately.

2.1 Matrix Approximation Primitives

The rest of the proof involves using the following matrix approximation primitives as discussed in class.

Definition 2.1 (Subspace Embedding)**.** *$S \in \mathbb{R}^{d\times m}$ is an ϵ-subspace embedding for $A \in \mathbb{R}^{n\times d}$ if for all $x \in \mathbb{R}^n$,*

$$\left|\left\|x^T M\right\|_2^2 - \left\|x^T MS\right\|_2^2\right| \leq \epsilon \left\|x^T M\right\|_2^2$$

Definition 2.2 (Approximate Matrix Product)**.** *$S \in \mathbb{R}^{d\times m}$ satisfies an ϵ-approximate matrix product for $A \in \mathbb{R}^{n\times d}, B \in \mathbb{R}^{d\times p}$*

$$\left\|AB^T - ASS^T B\right\|_F^2 \leq \epsilon \|A\|_F \|B\|_F$$

Definition 2.3 (Frobenius Norm Preservation)**.** *$S \in \mathbb{R}^{d\times m}$ satisfies ϵ-frobenius norm preservation if for $A \in \mathbb{R}^{n\times d}$*

$$\left|\|M\|_F^2 - \|MS\|_F^2 \leq \|M\|_F^2\right|$$

Theorem 2.1 (PCP Sketch via Matrix Approximation Primitives)**.** *If $S \in \mathbb{R}^{d\times m}$ satisfies the following:*

1. *$\epsilon/3$-subspace embedding for A_k*
2. *$\frac{\epsilon}{6\sqrt{k}}$-approximate matrix product for $A_{\setminus k}, A_{\setminus k}$*
3. *$\frac{\epsilon}{6\sqrt{k}}$-approximate matrix product for $A_{\setminus k}, V_k$, where V_k is the matrix formed by the top k right singular vectors of A.*
4. *$\frac{\epsilon}{6}$-Frobenius norm preservation for $A_{\setminus k}$*

Then $\tilde{A} = AS$ is an $(\epsilon, 0, k)$-PCP sketch of A.

To prove Theorem 2.1, we will bound the three terms in (3) by $\epsilon/3 \|YA\|_F^2$. It then follows that their sum is bounded by $\epsilon \|YA\|_F^2$, as desired.

2.2 Bounding the Head Term

From the subspace embedding assumption in Theorem 2.1, this follows naturally

$$\left|\|YA_k\|_F^2 - \|YA_kS\|_F^2\right| \leq \frac{\epsilon}{3}\|YA\|_F^2 \tag{4}$$

2.3 Bounding the Tail Term

We want to show the following:

$$\left|\left\|YA_{\setminus k}\right\|_F^2 - \left\|YA_{\setminus k}S\right\|_F^2\right| \leq \frac{\epsilon}{3}\|YA\|_F^2$$

Rewriting the left-hand-side yields

$$\begin{aligned}\left|\left\|(I-P)A_{\setminus k}\right\|_F^2 - \left\|(I-P)A_{\setminus k}S\right\|_F^2\right| &= \left|\left(\left\|A_{\setminus k}\right\|_F^2 - \left\|PA_{\setminus k}\right\|_F^2\right) - \left(\left\|A_{\setminus k}S\right\|_F^2 - \left\|PA_{\setminus k}S\right\|_F^2\right)\right| \\ &\qquad \text{(Pythagorean theorem)} \\ &\leq \left|\left\|A_{\setminus k}\right\|_F^2 - \left\|A_{\setminus k}S\right\|_F^2\right| + \left|\left\|PA_{\setminus k}\right\|_F^2 - \left\|PA_{\setminus k}S\right\|_F^2\right| \end{aligned} \tag{5}$$

We can then bound the first term in (5) using assumption 4 (Frobenius norm preservation) by:

$$\left|\left\|A_{\setminus k}\right\|_F^2 - \left\|A_{\setminus k}S\right\|_F^2\right| \leq \frac{\epsilon}{6}\left\|A_{\setminus k}\right\|_F^2 \leq \frac{\epsilon}{6}\|YA\|_F^2 \tag{6}$$

For the second term in (5), we can bound it as follows:

$$\begin{aligned}
\left|\left\|PA_{\setminus k}\right\|_F^2 - \left\|PA_{\setminus k}S\right\|_F^2\right| &= \left|\operatorname{tr}\left(P[A_{\setminus k}A_{\setminus k}^T A_{\setminus k}SS^T A_{\setminus k}^T]P\right)\right| \\
&= \left|\sum_{i=1}^{k}\lambda_i\right| \leq \sum_{i=1}^{k}|\lambda_i| \leq \sqrt{k}\cdot\left(\sum_{i=1}^{k}\lambda_i^2\right) \\
&\qquad (\lambda_i \text{ denotes } i\text{th eigenvalue, } P \text{ has rank} \leq k) \\
&= \sqrt{k}\cdot\left\|P[A_{\setminus k}A_{\setminus k}^T A_{\setminus k}SS^T A_{\setminus k}^T]P\right\|_F^2 \\
&\leq \sqrt{k}\cdot\left\|A_{\setminus k}A_{\setminus k}^T A_{\setminus k}SS^T A_{\setminus k}^T\right\|_F^2 \\
&\leq \frac{\epsilon}{6}\left\|A_{\setminus k}\right\|_F^2 \leq \frac{\epsilon}{6}\|YA\|_F^2 \qquad (7) \\
&\qquad (\text{Using assumption 2 in Theorem 2.1})
\end{aligned}$$

Combining (6) and (7), we can conclude that

$$\left|\left\|YA_{\setminus k}\right\|_F^2 - \left\|YA_{\setminus k}S\right\|_F^2\right| \leq \frac{\epsilon}{3}\|YA\|_F^2 \tag{8}$$

2.4 Bounding the Cross Term

Bounding the tail term requires showing

$$\left|2\operatorname{tr}(YA_kSS^TA_{\setminus k}^TY^T)\right| \leq \frac{\epsilon}{3}\|YA\|_F^2 \tag{9}$$

This requires using assumption 3 in Theorem 2.1. It is more involved than the proofs for the previous terms. We refer the reader to Section 2 of [MM20] for the full proof.

Combining all three inequalities (4, 8, 9), we can conclude that (3) holds.

3 Constructing A Sketch

It is not clear from the four requirements given by Theorem 2.1 how to construct such a sketch. However, [MM20] show that any matrix satisfying a JL-moment property as discussed in class suffices.

Definition 3.1 ((ϵ,δ,ℓ)-JL Moment Property)**.** *A matris $S \in \mathbb{R}^{d\times m}$ satisfies the (ϵ,δ,ℓ)-JL moment property for all unit vector $x \in \mathbb{R}^d$,*

$$\mathop{\mathbb{E}}_{S}\left|\left\|x^TS\right\|_2^2 - 1\right|^{\ell} \leq \epsilon^{\ell}\delta$$

Lemma 3.1 (PCP from JL-moment property)**.** *If $S \in \mathbb{R}^{d\times m}$ satisfies the $(\frac{\epsilon}{6\sqrt{k}},\delta,\ell)$-JL moment property and the $(\frac{\epsilon}{3},\frac{\delta}{9^k},\ell)$-JL moment property for any $\ell \geq 2$, then with probability $\geq 1-4\delta$, $\tilde{A} = AS$ is an $(\epsilon,0,k)$-PCP sketch of A.*

References

[BWZ15] Christos Boutsidis, David P Woodruff, and Peilin Zhong. Optimal principal component analysis in distributed and streaming models. April 2015.

[CEM+14] Michael B Cohen, Sam Elder, Cameron Musco, Christopher Musco, and Madalina Persu. Dimensionality reduction for k-means clustering and low rank approximation. October 2014.

[MM20] Cameron Musco and Christopher Musco. Projection-Cost-Preserving sketches: Proof strategies and constructions. April 2020.