CS 15-859: Algorithms for Big Data	Fall 2020
Lecture 6.1 — October 15	
Prof. David Woodruff	*Scribe: Jeffrey Chen*

1 Distributed Low Rank Approximation Recap

Recall that in our communication model, we have one coordinator C with two-way communication to a set of players $P_1, \ldots, P_s$. The players receive input $A_1, \ldots, A_s$ such that $A = \sum_{i=1}^{s} A_i$. The goal is for each player P_i to output a common rank-k projection P such that $\|A - AP\|_F \leq (1+\epsilon)\|A - A_k\|_F$, where A_k is the best rank k approximation for A. We previously showed with the FSS and KVW protocols that we can do this with $O(skd/\epsilon)$ communication; however, we would like to improve this bound to $O(skd) + \text{poly}(sk/\epsilon)$. This motivates the BWZ protocol, which will achieve this bound.

We will use projection-cost preserving (PCP) sketches [CEM+14] in our analysis.

Theorem 1. *Let A be an $n \times d$ matrix, and let S be a random $k/\epsilon^2 \times n$ matrix. Then there exists a scalar $c \geq 0$ such that for all k-dimensional projection matrices P,*

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

We omit the proof for this theorem; to read more, see [CEM+14].

To see what this implies, let $I - \tilde{P}$ be the minimizer of $\|SA(I - P)\|_F^2$ and let $I - P^*$ be the minimizer of $\|A(I - P)\|_F^2$. Then

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

Note that $[SA]_k = SA\tilde{P}$ and $A_k = AP^*$. The previous inequality essentially states $\|A(I - \tilde{P})\|_F^2 \leq (1 + \epsilon)\|A - A_k\|_F^2$. Then

$$\|SA - [SA]_k\|_F^2 + c = \|SA(I - \tilde{P})\|_F^2 + c \leq (1 + \epsilon)\|A(I - \tilde{P})\|_F^2 \leq (1 + O(\epsilon))\|A - A_k\|_F^2$$

Our protocol so far was as follows:

1. Let S be a $k/\epsilon^2 \times n$ projection-cost preserving sketch.
2. Let T be a $d \times k/\epsilon^2$ projection-cost preserving sketch. (Note that each server can have the same S and T sketches by each receiving the same small pseudorandom seed from the coordinator.)
3. Server t sends matrix SA^tT to the coordinator.
4. Coordinator sends matrix $SAT = \sum_{t=1}^{s} SA^tT$ to each server.
5. Each server computes $k/\epsilon^2 \times k$ matrix U of top k left singular vectors of SAT. (Note that SAT is a small $k/\epsilon^2 \times k/\epsilon^2$ matrix, so computing the SVD is cheap.)
6. Server t sends matrix $U^T SA^t$ to the coordinator.

7. Coordinator returns matrix $U^T SA = \sum_{t=1}^{s} U^T SA^t$ to output. The projection P will project onto the row span of $U^T SA$.

In step 6, the matrix $U^T SA^t$ is a $k \times d$ matrix, so s servers each sending a $k \times d$ matrix gives a communication cost of $O(skd)$. Previously, the servers sent SA^t, which are $k/\epsilon \times d$ matrices, meaning the cost was $O(skd/\epsilon)$, so we saved a factor of $1/\epsilon$. In total, this two-round protocol has cost $s \cdot \text{poly}(k/\epsilon) + O(skd)$ time, $s \cdot \text{poly}(k/\epsilon)$ for the first round and $O(skd)$ for the second round.

1.1 Intuition

Before showing correctness of this protocol, we develop some intuition for these matrices. Intuitively, U looks like the top k left singular vectors of SA. This is because we can think of T as a PCP for SA on the right, so by equation (1),

$$\|(I-P)SA\|_F^2 \leq \|(I-P)SAT\|_F^2 + c \leq (1+\epsilon)\|(I-P)SA\|_F^2$$

U contains the top k left singular vectors of SAT, so the rank-k projection that minimizes $\|(I-P)SAT\|_F^2$ is $P = UU^T$ that projects onto the vectors of U. Plugging this P into equation (2), we get $\|(I-P)SA\|_F^2 \leq (1+\epsilon)\|[SA]_k - SA\|_F^2$. This shows that projecting onto U gives a good approximation to the best rank-k approximation of SA, which is given by the top k singular vectors of SA.

If U_k is the top k left singular vectors of SA and $SA = U\Sigma V^T$ is ths SVD of SA, then $U_k U\Sigma V^T = \Sigma_k V_k^T$ would be the top k right singular vectors of SA scaled by its singular values. Therefore, $U^T SA$ (here we are referring to the original U that is the top k left singular values of SAT) looks like the top k right singular values of SA scaled by its singular values. The coordinator outputs the rows of $U^T SA$ to project onto, which looks like the right singular vectors of SA. Note that the rank-k projection $\tilde{P}$ that minimizes $\|SA(I-P)\|_F^2$ is given by projecting onto the right singular vectors of SA, which is $\tilde{P} = V_k V_k^T$. By equation (2), $\|A(I-\tilde{P}\| \leq (1+\epsilon)\|A(I-P^*)\|_F^2$, so $\tilde{P}$ is a good approximation to the optimal projection P^* that we want. Since $U^T SA$ looks like the top k right singular vectors of SA, we should intuitively project onto the rows of $U^T SA$ to get a good answer. We now formalize this intuition.

1.2 Proof

Let columns of W be an orthonormal basis for row span of $U^T SA$. Let $P = WW^T$ be the projection onto W. We want to show that

$$\|A - AP\|_F^2 \leq (1 + O(\epsilon))\|A - A_k\|_F^2$$

Recall that T is a PCP for SA. Let $\tilde{P} = \arg\min_{\text{rank-}k\ P} \|(I-P)SAT\|_F^2$. We know from equation (2) that $\|(I-\tilde{P})SA\|_F^2 \leq (1+\epsilon)\|SA - [SA]_k\|_F^2$. We know from above that the optimal $\tilde{P}$ for SAT is UU^T, so we have the inequality $\|SA - UU^T SA\|_F^2 \leq (1+\epsilon)\|SA - [SA]_k\|_F^2$.

Now we want to show $\|SA - SAP\|_F^2 \leq \|SA - UU^T SA\|_F^2$. Note that the rows of $UU^T SA$ live in W since $W = U^T SA$, and the rows of SAP also live in W because P projects onto W. Since SAP projects rows of SA onto W and $UU^T SA$ only puts them somewhere in W, we know the rows of

SAP are closer to SA than the rows of UU^TSA, so $\|SA - SAP\|_F^2 \leq \|SA - UU^TSA\|_F^2$. Putting these two inequalities together, we get

$$\|SA - SAP\|_F^2 \leq (1+\epsilon)\|SA - [SA]_k\|_F^2 \tag{3}$$

Recall that S is a PCP for A. Then by equation (1), there is a scalar $c \geq 0$ such that for all k-dimensional projection matrices Q,

$$\|SA - SAQ\|_F^2 + c \leq (1+\epsilon)\|A - AQ\|_F^2$$

To complete the proof, add c to both sides of equation (3) to get

$$\|SA - SAP\|_F^2 + c \leq (1+\epsilon)\|SA - [SA]_k\|_F^2 + c$$

Since S is a PCP for A, $\|A(I-P)\|_F^2 \leq \|SA - SAP\|_F^2 + c$ and $\|(1+\epsilon)\|SA - [SA]_k\|_F^2 + c \leq (1+\epsilon)(\|SA - [SA]_k\|_F^2 + c) \leq (1+\epsilon)(1+\epsilon)\|A - A_k\|_F^2$. Putting these together, we get

$$\|A - AP\|_F^2 \leq (1 + O(\epsilon))\|A - A_k\|_F^2$$

which is what we wanted to show.

1.3 Conclusion

To summarize, the BWZ protocol achieves the optimal $O(sdk) + \text{poly}(sk/\epsilon)$ communication cost for low rank approximation in the arbitrary partition model. We did not go into resolving bit complexity: in our protocol, the servers send matrices U^TSA^t. If A has bounded bit complexity, then SA^t also has bounded bit complexity, but U^T may not. It turns out that we can reduce bit complexity of U by adding noise, though we will not go into detail. The computations done by the coordinator and the servers can be done in input sparsity time. Our protocol runs in 2 rounds, which is optimal for the $O(sdk)$ bound, since with 1 round there is a lower bound of $\Omega(sdk/\epsilon)$ for communication.

There are many open problems regarding the communication complexity of various optimization problems.

- There are two servers with $n \times n$ matrices A and B, and they want to find the rank of $A + B$. The only known bounds are an upper bound of $O(n^2)$ and a lower bound of $O(n)$.
- For linear programming, each server has a set of constraints $\langle A_i, x\rangle \leq b_i$, and they want to find $\min_x(c^Tx)$ subject to the constraints $Ax \leq b$ and $x \geq 0$. If A is an $n \times d$ matrix, meaning there are n constraints and d variables, the only known bounds are $O(sd^3)$ and $\Omega(d^2 + sd)$.
- There are many servers, each with some subset of the edges of a graph. Can we answer graph questions like finding a matching or vertex cover without sending all the edges to the coordinator?
- etc...

References

[CEM+14] Michael B. Cohen, Sam Elder, Cameron Musco, Christopher Musco, and Madalina Persu. Dimensionality reduction for k-means clustering and low rank approximation. *CoRR*, abs/1410.6801, 2014.