Constructing a Coreset

- Claim: For all projection matrices Y=I-X onto (d-k)-dimensional subspaces,

$$\left|\Sigma_m V^T Y\right|_F^2 + c \;= (1 \pm \epsilon)|AY|_F^2,$$

where $c = |A - A_m|_F^2$ does not depend on Y

- Proof: $|AY|_F^2 = \left|U\Sigma_m V^T Y\right|_F^2 + \left|U(\Sigma - \Sigma_m)V^T Y\right|_F^2$

$$\leq \left|\Sigma_m V^T Y\right|_F^2 + |A - A_m|_F^2 = \left|\Sigma_m V^T Y\right|_F^2 + c$$

Also, $\left|\Sigma_m V^T Y\right|_F^2 + |A - A_m|_F^2 \; - |AY|_F^2$

$$= \left|\Sigma_m V^T\right|_F^2 - \left|\Sigma_m V^T X\right|_F^2 + |A - A_m|_F^2 - |A|_F^2 + |AX|_F^2$$

$$= |AX|_F^2 - \left|\Sigma_m V^T X\right|_F^2$$

$$= \left|(\Sigma - \Sigma_m)V^T X\right|_F^2$$

$$\leq \left|(\Sigma - \Sigma_m)V^T\right|_2^2 \cdot |X|_F^2$$

$$\leq \sigma_{m+1}^2\, k \leq \epsilon\, \sigma_{m+1}^2 (m-k) \leq \epsilon \sum_{i\in\{k+1,..,m+1\}} \sigma_i^2 \leq \epsilon |A - A_k|_F^2$$

Unions of Coresets

- Suppose we have matrices $A^1, \ldots, A^s$ and construct $\Sigma_m^1 V^{T,1}, \Sigma_m^2 V^{T,2}, \ldots, \Sigma_m^s V^{T,s}$ as in the previous slide, together with $c_1, \ldots, c_s$

- Then $\sum_i \quad \left|\Sigma_m^i V^{T,i} Y\right|_F^2 + c_i \ = (1 \pm \epsilon)|AY|_F^2$, where A is the matrix formed by concatenating the rows of $A^1, \ldots, A^s$

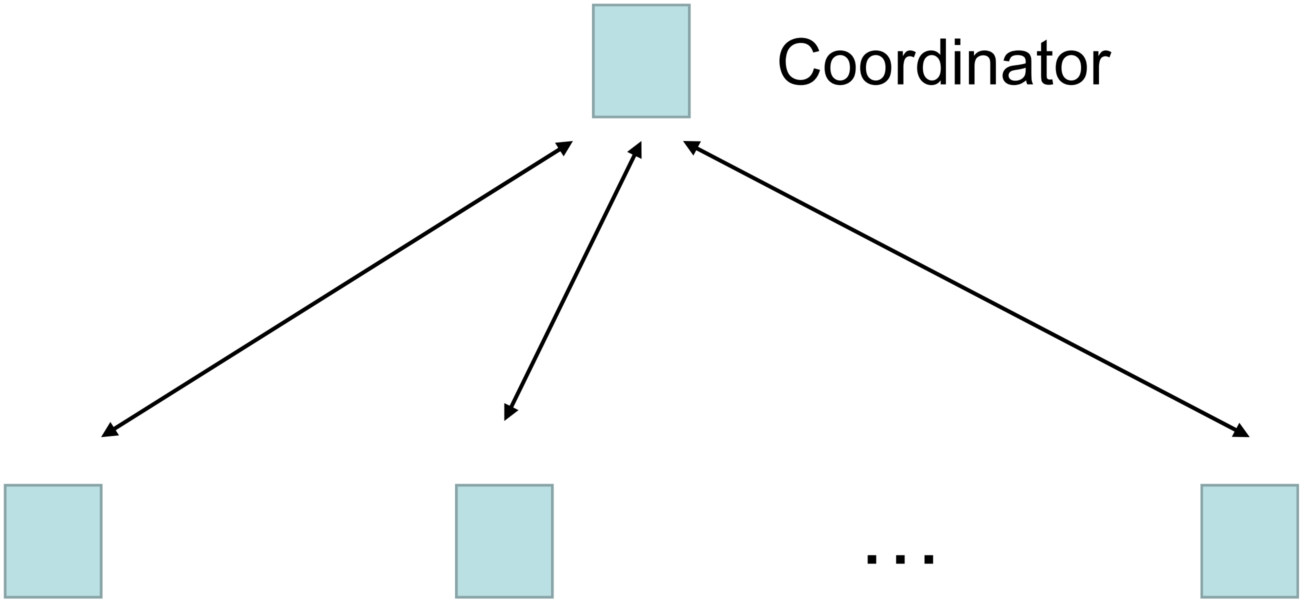
- Let B be the matrix obtained by concatenating the rows of $\Sigma_m^1 V^{T,1}, \Sigma_m^2 V^{T,2}, \ldots, \Sigma_m^s V^{T,s}$

- Suppose we compute $B = U\,\Sigma V^T$ and compute $\Sigma_m V^T$ and $|B - B_m|_F^2$

- Then $\left|\Sigma_m V^T Y\right|_F^2 + c + \sum_i c_i = (1 \pm \epsilon)|BY|_F^2 + \sum_i c_i = (1 \pm O(\epsilon))|AY|_F^2$

- So $\Sigma_m V^T$ and the constant $c + \sum_i c_i$ are a coreset for A

[FSS] Row-Partition Protocol



Coordinator

…

$P^1 \in R^{n_1 \text{ x } d}$ $P^2 \in R^{n_2 \text{ x } d}$ $P^s \in R^{n_s \text{ x } d}$

- Server t sends the top k/ε + k principal components of P^t, scaled by the top k/ε + k singular values Σ^t, together with c^t
- Coordinator returns $c + \sum_i c_i$ and top k principal components of $[\Sigma^1 V^1; \Sigma^2 V^2; \ldots; \Sigma^s V^s]$

[FSS] Row-Partition Protocol

[KVW] protocol will handle 2, 3, and 4

Problems:
1. sdk/ε real numbers of communication
2. bit complexity can be large
3. running time for SVDs
4. doesn’t work in arbitrary partition model

This is an SVD-based protocol. Maybe our random matrix techniques can improve communication just like they improved computation?

[KVW] Arbitrary Partition Model Protocol

- Inspired by the sketching algorithm presented earlier

- Let S be one of the k/ε x n random matrices discussed
    - S can be generated pseudorandomly from small seed
    - Coordinator sends small seed for S to all servers

- Server t computes SA^t and sends it to Coordinator

- Coordinator sends $\Sigma_{t=1}^{s} SA^t = SA$ to all servers

- There is a good k-dimensional subspace inside of SA. If we knew it, t-th server could output projection of A^t onto it

[KVW] Arbitrary Partition Model Protocol

Problems:

- Can’t output projection of A^t onto SA since the rank is too large

- Could communicate this projection to the coordinator who could find a k-dimensional space, but communication depends on n

[KVW] Arbitrary Partition Model Protocol

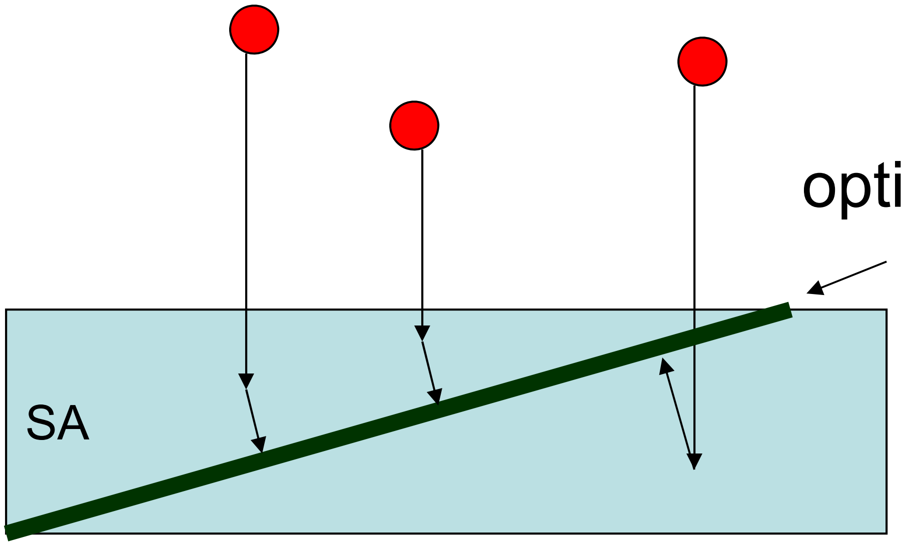
Fix:

- Instead of projecting A onto SA, recall we can solve $\min_{\text{rank}-k\ X} \left|A(SA)^{T}XSA - A\right|_F^2$
- Let T_1, T_2 be affine embeddings, solve $\min_{\text{rank}-k\ X} \left|T_1A(SA)^{T}XSAT_2 - T_1AT_2\right|_F^2$ (optimization problem is small and has a closed form solution)
- Everyone can then compute XSA and then output k directions

[KVW] protocol

- Phase 1:

- Learn the row space of SA



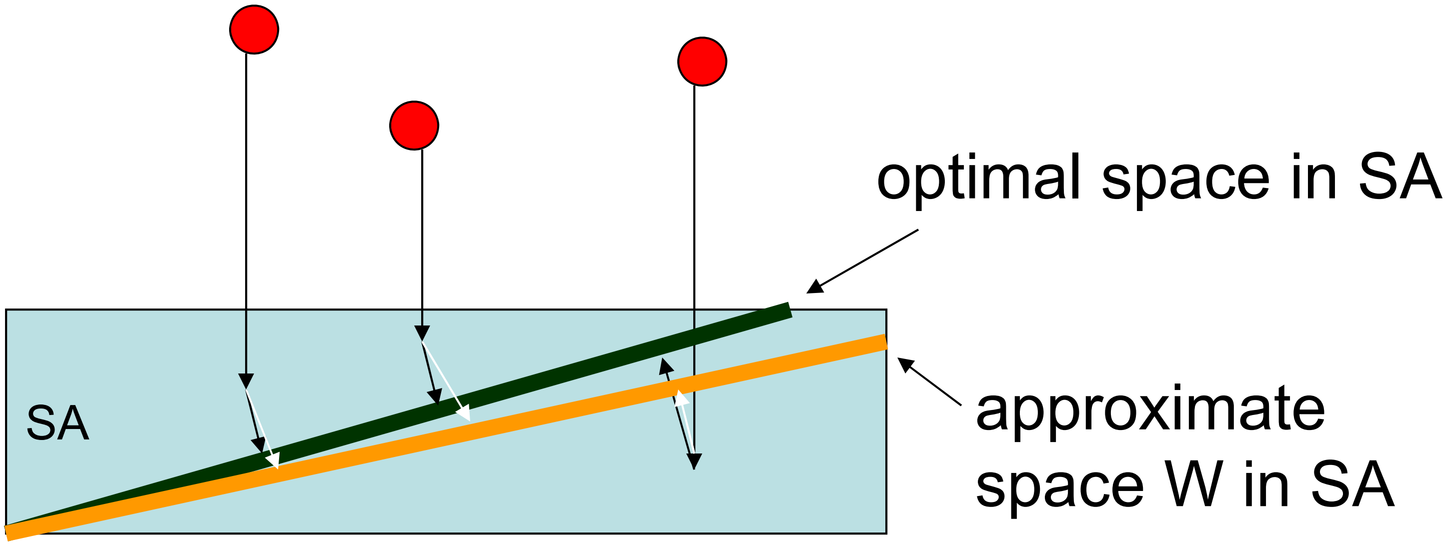
optimal k-dimensional space in SA

cost ‰ $(1+\varepsilon)|A-A_k|_F$

[KVW] protocol

- Phase 2:

- Find an approximately optimal space W inside of SA



cost ‰ $(1+\varepsilon)^2|A-A_k|_F$

[BWZ] Protocol

- Main Problem: communication is O(skd/ε) + poly(sk/ε)
- We want O(skd) + poly(sk/ε) communication!
- Idea: use projection-cost preserving sketches [CEMMP]
- Let A be an n x d matrix
  - If S is a random k/ε^2 x n matrix, then there is a scalar $c \geq 0$ so that for all k-dimensional projection matrices P:

$$|SA(I-P)|_F^2 + c \; = (1 \pm \epsilon)|A(I-P)|_F^2$$

[BWZ] Protocol

Intuitively, U looks like top k left singular vectors of SA

- Let S be a k/ε^2 x n projection-cost preserving sketch
- Let T be a d x k/ε^2 projection-cost preserving sketch
- Server t sends SA^tT to Coordinator
- Coordinator sends back SAT = $\sum_t SA^tT$ to servers
- Each server computes k/ε^2x k matrix U of top k left singular vectors of SAT

Thus, U^TSA looks like top k right singular vectors of SA

- Server t sends U^TSA^t to Coordinator
- Coordinator returns the space $U^TSA = \sum_t U^TSA^t$ to output

Top k right singular vectors of SA work because S is a projection-cost preserving sketch!

[BWZ] Analysis

- Let W be the row span of $\mathrm{U^TSA}$, and P be the projection onto W

- Want to show $|\mathrm{A}-\mathrm{AP}|_\mathrm{F}^2 \leq (1+\epsilon)|\mathrm{A}-\mathrm{A_k}|_\mathrm{F}^2$

- Since T is a projection-cost preserving sketch,

$$(*) \quad |\mathrm{SA}-\mathrm{SAP}|_\mathrm{F}^2 \leq \left|\mathrm{SA} - \mathrm{UU^TSA}\right|_\mathrm{F}^2 + \mathrm{c_1} \leq (1+\epsilon)|\mathrm{SA}-[\mathrm{SA}]_\mathrm{k}|_\mathrm{F}^2$$

- Since S is a projection-cost preserving sketch, there is a scalar c > 0, so that for all k-dimensional projection matrices Q,

$$|\mathrm{SA} - \mathrm{SAQ}|_\mathrm{F}^2 + \mathrm{c} = (1 \pm \epsilon)|\mathrm{A}-\mathrm{AQ}|_\mathrm{F}^2$$

- Add c to both sides of (*) to conclude $|\mathrm{A}-\mathrm{AP}|_\mathrm{F}^2 \leq (1+\epsilon)|\mathrm{A}-\mathrm{A_k}|_\mathrm{F}^2$

Conclusions for Distributed Low Rank Approximation

- [BWZ] Optimal O(sdk) + poly(sk/ε) communication protocol for low rank approximation in arbitrary partition model
  - Handle bit complexity by adding noise (omitted)
  - Input sparsity time
  - 2 rounds, which is optimal [W]

- Communication of other optimization problems?
  - Computing the rank of an n x n matrix over the reals
  - Linear Programming
  - Graph problems: Matching
  - etc.

Course Outline

- Subspace embeddings and least squares regression
  - Gaussian matrices
  - Subsampled Randomized Hadamard Transform
  - CountSketch
- Affine embeddings
  - Application to low rank approximation
- High precision regression
- Leverage score sampling
- Distributed low rank approximation
- L1 Regression
- M-Estimator Regression

Robust Regression

Method of least absolute deviation (l_1 -regression)

- Find x* that minimizes $|Ax-b|_1 = \Sigma\ |b_i - <A_{i*}, x>|$
- Cost is less sensitive to outliers than least squares
- Can solve via linear programming

Solving l_1 -regression via Linear Programming

- Minimize $(1,\ldots,1) \cdot (\alpha^+ + \alpha^-)$
- Subject to:

$$A x + \alpha^+ - \alpha^- = b$$

$$\alpha^+, \alpha^- \geq 0$$

- Generic linear programming gives poly(nd) time
- Want much faster time using sketching!