[BWZ] Protocol

- Main Problem: communication is O(skd/ε) + poly(sk/ε), but 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:
$$|A(I-P)|_F^2 \leq |SA(I-P)|_F^2 + c \leq (1+\epsilon)|A(I-P)|_F^2$$

- Implication: If I-$\widetilde{P}$ is the minimizer of $|SA(I-P)|_F^2$, and $I - P^*$ is the minimizer of $|A(I-P)|_F^2$, then $\left|A(I-\widetilde{P})\right|_F^2 \leq (1+\epsilon)|A-A_k|_F^2$
- So $|SA-[SA]_k|_F^2 + c \leq (1+\epsilon)|A(I-\tilde{P})|_F^2 \leq (1+O(\epsilon))|A-A_k|_F^2$

[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 scaled 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 $U^T SA$, and P be the projection onto W
- Want to show $|A - AP|_F^2 \leq (1+\epsilon)|A - A_k|_F^2$
- Since T is a projection-cost preserving sketch,

$$(*) \quad |SA - SAP|_F^2 \leq \left|SA - UU^T SA\right|_F^2 \leq (1+\epsilon)|SA - [SA]_k|_F^2$$

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

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

- Add c to both sides of (*) to conclude $|A - AP|_F^2 \leq (1 + O(\epsilon))|A - A_k|_F^2$

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 $|Ax-b|_1 = \Sigma\ |b_i - \langle A_{i*}, x\rangle|$
- 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!

Well-Conditioned Bases

- For an n x d matrix A, can choose an n x d matrix U with orthonormal columns for which A = UW, and $|\mathrm{Ux}|_2 = |\mathrm{x}|_2$ for all x
- Can we find a U for which A = UW and $|\mathrm{Ux}|_1 \approx |\mathrm{x}|_1$ for all x?
- Let A = QW where Q has full column rank, and define $|\mathrm{z}|_{\mathrm{Q},1} = |\mathrm{Qz}|_1$
  - $|\mathrm{z}|_{\mathrm{Q},1}$ is a norm
- Let C = $\{\mathrm{z} \in \mathrm{R}^{\mathrm{d}} : |\mathrm{z}|_{\mathrm{Q},1} \leq 1\}$ be the unit ball of $|.|_{\mathrm{Q},1}$
- C is a convex set which is symmetric about the origin
  - Lowner-John Theorem: can find an ellipsoid E such that: $\mathrm{E} \subseteq \mathrm{C} \subseteq \sqrt{\mathrm{d}}\mathrm{E}$, where E = $\{\mathrm{z} \in \mathrm{R}^{\mathrm{d}} : \mathrm{z}^{\mathrm{T}}\mathrm{Fz} \leq 1\}$
  - $(\mathrm{z}^{\mathrm{T}}\mathrm{Fz})^{.5} \leq |\mathrm{z}|_{\mathrm{Q},1} \leq \sqrt{\mathrm{d}}(\mathrm{z}^{\mathrm{T}}\mathrm{Fz})^{.5}$
  - $\mathrm{F} = \mathrm{G}^{\mathrm{T}}\mathrm{G}$ since F defines an ellipsoid
- Define $\mathrm{U} = \mathrm{QG}^{-1}$

Well-Conditioned Bases

- Recall $U = QG^{-1}$ where
  $(z^T F z)^{.5} \leq |z|_{Q,1} \leq \sqrt{d}(z^T F z)^{.5}$ and $F = G^T G$
- $|Ux|_1 = |QG^{-1}x|_1 = |Qz|_1 = |z|_{Q,1}$ where $z = G^{-1}x$
- $z^T F z \quad = (x^T (G^{-1})^T G^T G\,(G^{-1})x) = x^T x = |x|_2^2$
- So $|x|_2 \leq |Ux|_1 \leq \sqrt{d}|x|_2$
- So $\frac{|x|_1}{\sqrt{d}} \leq |x|_2 \leq |Ux|_1 \leq \sqrt{d}|x|_2 \leq \sqrt{d}|x|_1$

Net for ℓ_1 – Ball

- Consider the unit ℓ_1-ball B = $\{x \in R^d : |x|_1 = 1\}$
- Subset N is a γ-net if for all $x \in B$, there is a $y \in N$, such that $|x - y|_1 \leq \gamma$
- Greedy construction of N
  - While there is a point $x \in B$ of distance larger than γ from every point in N, include x in N
- The ℓ_1-ball of radius γ/2 around every point in N is contained in the ℓ_1-ball of radius 1+ γ/2 around 0^d
- Further, all such ball are disjoint
- Ratio of volume of d-dimensional similar polytopes of radius 1+ γ/2 to radius $\gamma/2$ is $(1 + \gamma/2)^d/(\gamma/2)^d$, so $|N| \leq (1 + \gamma/2)^d/(\gamma/2)^d$

Net for $\ell_1 - \mathrm{Subspace}$

- Let A = UW for a well-conditioned basis U
  - $|x|_1 \leq |Ux|_1 \leq d|x|_1$ for all x

- Let N be a (γ/d) −net for the unit ℓ_1-ball B

- Let M = {Ux | x in N}, so $|M| \leq (1 + \gamma/(2d))^d/(\gamma/(2d))^d$

- Claim: For every x in B, there is a y in M for which $|Ux - y|_1 \leq \gamma$

- Proof: Let x’ in N be such that $|x - x'|_1 \leq \gamma/d$

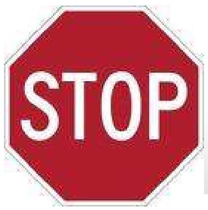
  Then $|Ux - Ux'|_1 \leq d|x - x'|_1 \leq \gamma$, using the well-conditioned basis property. Set y = Ux’

- $|M| \leq \left(\frac{d}{\gamma}\right)^{O(d)}$

Rough Algorithm Overview

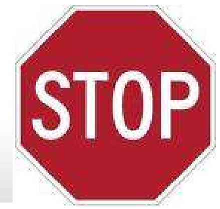
$\min_{x \in R^d} |Ax-b|_1 = \min_{x \in R^d} |Ux - b'|_1$

Sample poly(d/ε) rows of U∘b' proportional to their l_1-norm.



Compute poly(d)-approximation

Compute well-conditioned basis



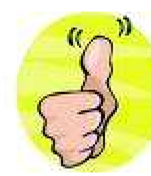
Find x' such that $|Ax'-b|_1 \leq \text{poly}(d) \min_{x \in R^d} \ldots$
Let b' = b-Ax' be the residual

Find a basis A=UW so that for all x in R^d,
$|x|_1/\text{poly}(d) \leq |Ux|_1 \leq \text{poly}(d)\,|x|_1$

Takes nnz(A) ...

Now generic linear programming is efficient

Solve l_1-regression on the sample, obtaining vector x, and output x



Will focus on showing how to quickly compute

1. A poly(d)-approximation

2. A well-conditioned basis

Sketching Theorem

Theorem

- There is a probability space over $(d \log d) \times n$ matrices R such that for any $n \times d$ matrix A, with probability at least 99/100 we have for all x:

$$|Ax|_1 \leq |RAx|_1 \leq d \log d \cdot |Ax|_1$$

Embedding

- is linear
- is independent of A
- preserves lengths of an infinite number of vectors

Application of Sketching Theorem

Computing a d(log d)-approximation

- Compute RA and Rb
- Solve $x' = \text{argmin}_x |RAx-Rb|_1$
- Main theorem applied to A∘b implies x' is a d log d – approximation
- RA, Rb have d log d rows, so can solve l_1-regression efficiently

Application of Sketching Theorem

Computing a well-conditioned basis

1. Compute RA
2. Compute W so that RAW is orthonormal (in the l_2-sense)
3. Output U = AW

U = AW is well-conditioned because

$|AWx|_1 \leq |RAWx|_1 \leq (d \log d)^{1/2} |RAWx|_2 = (d \log d)^{1/2} |x|_2 \leq (d \log d)^{1/2} |x|_1$

and

$|AWx|_1 \geq |RAWx|_1/(d \log d) \geq |RAWx|_2/(d \log d) = |x|_2/(d \log d) \geq |x|_1 /(d^{3/2} \log d)$

Sketching Theorem

Theorem:

- There is a probability space over $(d \log d) \times n$ matrices R such that for any $n \times d$ matrix A, with probability at least 99/100 we have for all x:

$$|Ax|_1 \leq |RAx|_1 \leq d \log d \cdot |Ax|_1$$

A dense R that works:

The entries of R are i.i.d. Cauchy random variables, scaled by $1/(d \log d)$