IIT KHARAGPUR

IIT GANDHINAGAR

NPTEL

NPTEL ONLINE
CERTIFICATION COURSES

Scalable Data Science

Lecture 16b: Leverage Score & Applications

Anirban Dasgupta
Computer Science and Engineering
IIT GANDHINAGAR

Until now…

- Studied applications of random projection techniques on linear algebraic problems
  - random projection to approximate PCA
  - QB decomposition
  - random projection for efficient L2 regression

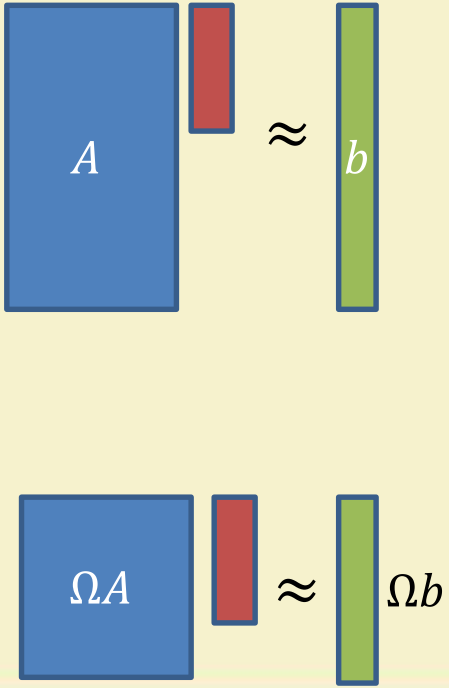
- Pros: numerically stable, computationally efficient
- Cons: interpretability, sparsity

Projection vs Sampling

- Projection ≡ linear combinations
- Wish to select actual data points instead

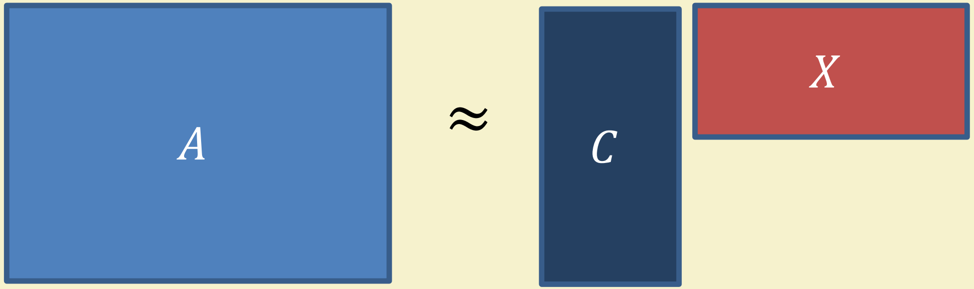
- Broader question: Given an optimization problem, can we create a "smaller", weighted, dataset such that solving the optimization on the smaller problem gives a good approximation
  - much literature in "coresets", techniques mostly orthogonal to RandNLA, but starting to converge
  - we look at settings of linear regression and matrix factorization

Linear regression



- Does there exist subset of (scaled) rows that forms the smaller problem, i.e. can Ω be a purely sampling matrix?
  - Still need to maintain guarantees
  - Choosing rows here will help preserve "memory footprint" of problem
  - recall that our proof was based on the fact that ΩA has same rank as A

CX decomposition



Can we find C, a subset of columns of A, such that

$$|A - CX|_F \approx |A - A_k|_F$$

and $|C| = O(k)$? It is not clear that this even exist !! Useful in ML applications when actual columns, and not their linear combinations, are needed.

Length-squared sampling

$p_i = \frac{|A_{*i}|^2}{|A|_F^2}$, each row sampled proportional to squared length

Gave us approximation

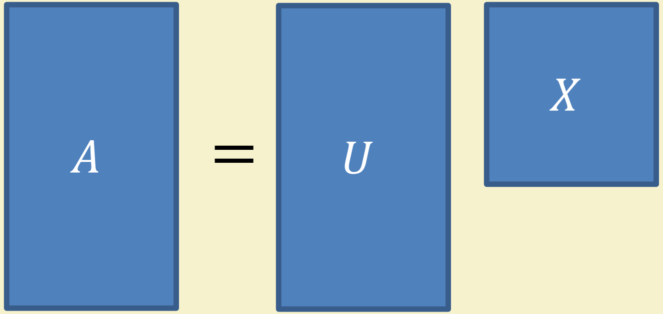
$$|AA^t \; - CC^t|_F \leq \frac{1}{\sqrt{c}} |A|_F^2$$

Similar additive approximation can be obtained

- Low rank approximation
- CX decomposition

Leverage scores

$A \in \Re^{n \times d} \quad n > d$



U orthonormal

$\ell_i = \frac{|U_{i*}|^2}{|U|_F^2}$, leverage score of the i^{th} row

$\sum_i \ell_i = 1$

Create a sample of rows with probabilities $\{\ell_i\}$, with replacement

Leverage scores: few properties

Row leverage scores make sense only if #rows > #columns

- else we should be focussing on columns instead

Leverage scores are independent of the particular orthogonal basis considered

Leverage scores: few properties

Row leverage scores make sense only if #rows > #columns

- else we should be focussing on columns instead

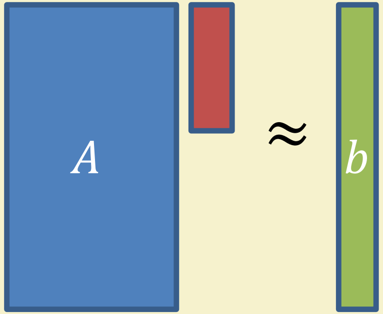
Leverage scores are independent of the particular orthogonal basis considered

Q and U be two orthogonal basis of columns of A

$\Rightarrow Q = UR$, for some $d \times d$ rotation R

$\Rightarrow |Q_{*i}| = |U_{*i}R|$

Linear regression

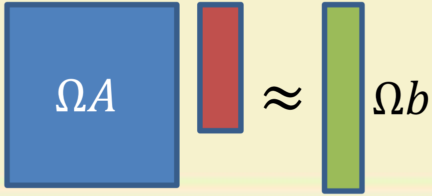


$\Omega \in \Re^{s \times n}$

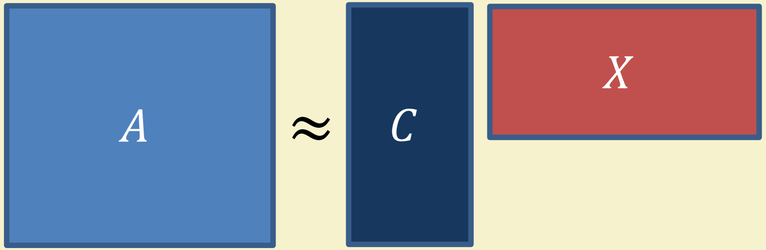
For each $t \in [1, s]$

- choose i with prob ℓ_i with replacement
- $\Omega_{ti} = \frac{1}{\sqrt{c\ell_i}}$

Solve $|\Omega A x \; - \Omega b|$



CX decomposition



$A = U\Sigma V^t$ and V_k = top k right singular vectors

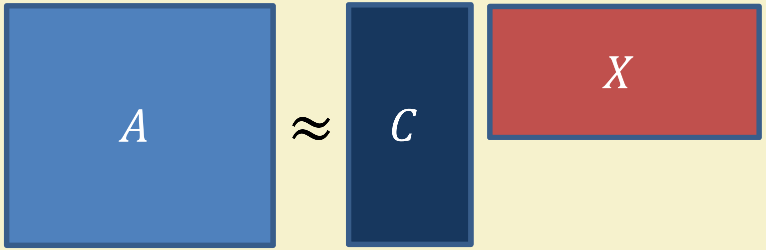
$\ell_i \propto$ row squared length of V_k

For $t = 1..c$

- pick column i with prob ℓ_i , normalize by $1/\sqrt{\ell_i}$ and add to C

$X = C^+ A$

CX decomposition



$$|A - CX|_F \leq (1+\epsilon)|A - A_k|_F$$

$A = U\Sigma V^t$ and V_k = top k right singular vectors

$\ell_i \propto$ row squared length of V_k

For $t = 1..c$

- pick column i with prob ℓ_i , normalize by $1/\sqrt{\ell_i}$ and add to C

$X = C^+ A$

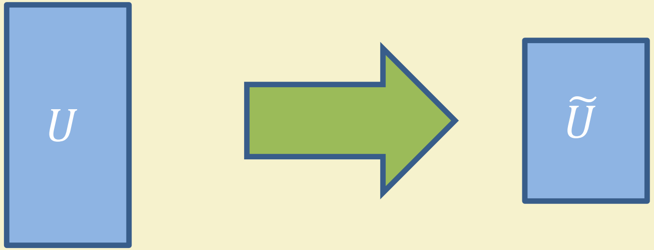
Why do leverage scores work?

All proofs that show leverage scores work have the following form

U orthogonal

$$U^t U = I_d$$

$$\ell_i = \frac{|U_{*i}|^2}{|U|_F^2} = \frac{|U_{*i}|^2}{d}$$



$n \gg d$

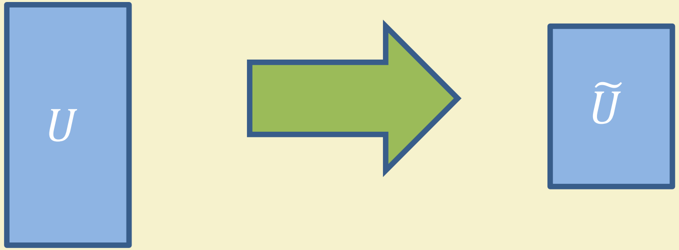
[slides from Petros Drineas]

Why do leverage scores work?

All proofs that show leverage scores work have the following form

U orthogonal
$U^t U = I_d$

$$\ell_i = \frac{|U_{*i}|^2}{|U|_F^2} = \frac{|U_{*i}|^2}{d}$$



$n \gg d$

$r \times d$

$$r = O\left(\frac{d \log(d/\delta)}{\epsilon^2}\right)$$

Then, with prob at least $1 - \delta$

$$\left|U^t U - \widetilde{U}^t \widetilde{U}\right| = \left| I_d - \widetilde{U}^t \widetilde{U}\right| \le \epsilon$$

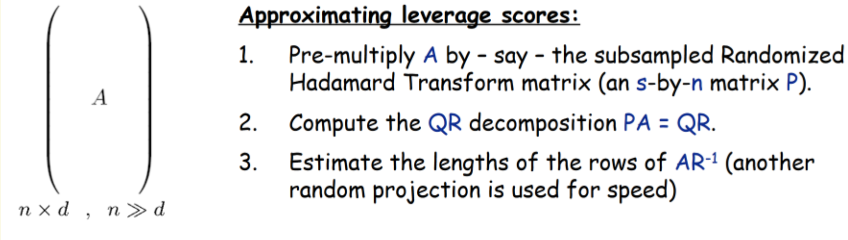
Leverage score intuition

- $\left| I_d - \widetilde{U^t}\widetilde{U} \right| \leq \epsilon$ implies all singular values are in a small interval
  - $\widetilde{U}$ is full rank
  - This allows us to bound the norm of pseudo-inverse of A

Estimating leverage score

- Getting the exact leverage score is expensive
  - All the analysis work with an upper bound

- Possible to get an approximation of the left/right singular vectors



$$\begin{pmatrix} \\ \\ A \\ \\ \\ \end{pmatrix}$$

$n \times d \quad , \quad n \gg d$

Approximating leverage scores:

1. Pre-multiply A by – say – the subsampled Randomized Hadamard Transform matrix (an s-by-n matrix P).
2. Compute the QR decomposition PA = QR.
3. Estimate the lengths of the rows of AR^{-1} (another random projection is used for speed)

$$s = O\left(\frac{d}{\epsilon} polylog(n)\right)$$

Runtime = $O\left(\frac{nd}{\epsilon} polylog(n)\right)$

[slide from Petros Drineas]

Summary

- Sampling is often preferable to projection if we are more interested in preserving space (sparsity) + interpretability + downstream ML applications

- Interesting ways to sample non-uniformly : leverage score
  - has extensions to other norms

References:

- Primary reference
  - Lecture notes on Randomized Numerical Linear Algebra by Petros Drineas and Michael Mahoney, https://arxiv.org/abs/1712.08880
  - Google "Simon's Institute Bootcamp on Randomized Numerical Linear Algebra"

Thank You!!