Leverage Scores: Sensitivity and an App

Ilse Ipsen

Joint work with Thomas Wentworth

North Carolina State University
Raleigh, NC, USA

Research supported by NSF CISE CCF, NSF DMS, DARPA XData

Leverage Scores

- Real $m \times n$ matrix $\mathcal{A}$ with $\text{rank}(\mathcal{A}) = n$
- Columns of Q are orthonormal basis for $\text{range}(\mathcal{A})$

  column i of Q $\perp$ column j of Q, $\|\text{column } j \text{ of } Q\|_2 = 1$
- Leverage scores of $\mathcal{A}$

$$\ell_j(\mathcal{A}) = \|\text{row } j \text{ of } Q\|_2^2 \qquad 1 \leq j \leq m$$

Leverage Scores

- Real $m \times n$ matrix $\mathcal{A}$ with $\text{rank}(\mathcal{A}) = n$
- Columns of Q are orthonormal basis for $\text{range}(\mathcal{A})$

  column i of $Q \perp$ column j of Q, $\|\text{column } j \text{ of } Q\|_2 = 1$

- Leverage scores of $\mathcal{A}$

$$\ell_j(\mathcal{A}) = \|\text{row } j \text{ of } Q\|_2^2 \qquad 1 \leq j \leq m$$

- $0 \leq \ell_j(\mathcal{A}) \leq 1 \qquad \sum_{j=1}^m \ell_j(\mathcal{A}) = n = \|\mathcal{A}\|_F^2$
- Outlier detection in least squares/regression problems

  [Hoaglin & Welsch 1978, Velleman & Welsch 1981, Chatterjee & Hadi 1986]

- Importance sampling in randomized matrix algorithms

  [Avron, Boutsidis, Drineas, Mahoney, Toledo, ...]

Largest Leverage Score

Largest leverage score $=$ coherence of $\mathcal{A}$

$$\mu(\mathcal{A}) \equiv \max_j \ell_j(\mathcal{A})$$

[Donoho & Ho 2001, Candés, Romberg & Tao 2006, Candés & Recht 2009, ...]

Largest Leverage Score

Largest leverage score = coherence of $\mathcal{A}$

$$\mu(\mathcal{A}) \equiv \max_j \ell_j(\mathcal{A})$$

[Donoho & Ho 2001, Candés, Romberg & Tao 2006, Candés & Recht 2009, ...]

- $n/m \leq \mu(A) \leq 1$
- Uniform leverage scores $\ell_j(\mathcal{A}) = n/m$ for all j
  $\Rightarrow$ Minimal coherence $\mu(\mathcal{A}) = n/m$
  Sampling is easy
- Large leverage score $\ell_j(\mathcal{A}) = 1$ for some j
  $\Rightarrow$ Maximal coherence $\mu(\mathcal{A}) = 1$
  Sampling is hard

Why Leverage Scores?

- Analysis of randomized algorithms:
  Quantify the difficulty of sampling
- Probabilities in randomized matrix algorithms
  [Boutsidis, Drineas, Mahoney, ...]
- Least Squares Solver *Blendenpik* [Avron, Maymounkov, Toledo 2010]
  Bounds for condition numbers of sampled matrices

Why Leverage Scores?

- Analysis of randomized algorithms:
  Quantify the difficulty of sampling
- Probabilities in randomized matrix algorithms
  [Boutsidis, Drineas, Mahoney, ...]
- Least Squares Solver *Blendenpik* [Avron, Maymounkov, Toledo 2010]
  Bounds for condition numbers of sampled matrices

Condition number (sensitivity of basis to perturbations)
If $\mathcal{A}$ has linearly independent columns then

$$\kappa(\mathcal{A}) \equiv \|\mathcal{A}\|_2 \, \|\mathcal{A}^\dagger\|_2$$

If Q has orthonormal columns then $\kappa(Q) = 1$

Overview

1. Motivation
2. The App
3. Sensitivity of leverage scores, to:

    Rotation of subspace
    Matrix perturbations

Motivation

The Problem

Given

Real $m \times n$ matrix Q with orthonormal columns

coherence μ and leverage scores ℓ_j

Real $c \times m$ "sampling" matrix S with $n \leq c \ll m$

Condition number of sampled matrix

$$\kappa(SQ) \equiv \|SQ\|_2 \, \|(SQ)^\dagger\|_2$$

Sensitivity of SQ to perturbations

Given η and δ, for which values of c (in terms of μ and ℓ_j) is

$$\kappa(SQ) \leq 1 + \eta$$

with probability at least $1 - \delta$?

Example: A Bound [Ipsen & Wentworth 2012]

Label leverage scores so that $\mu \equiv \ell_{[1]} \geq \cdots \geq \ell_{[m]}$

$$\tau \equiv \sum_{j=1}^{t} \ell_{[j]} + \left(\frac{1}{\mu} - t\right) \ell_{[t+1]} \qquad \text{where} \quad t \equiv \lfloor 1/\mu \rfloor$$

SQ : c rows of Q sampled uniformly with replacement

Example: A Bound [Ipsen & Wentworth 2012]

Label leverage scores so that $\mu \equiv \ell_{[1]} \geq \cdots \geq \ell_{[m]}$

$$\tau \equiv \sum_{j=1}^{t} \ell_{[j]} + \left(\frac{1}{\mu} - t\right) \ell_{[t+1]} \qquad \text{where} \quad t \equiv \lfloor 1/\mu \rfloor$$

SQ : c rows of Q sampled uniformly with replacement

Given ϵ, with probability at least $1 - \delta$

$$\kappa(SQ) \leq \sqrt{\frac{1+\epsilon}{1-\epsilon}}$$

provided $c \geq m\,\mu\,(2\,\tau + \frac{2}{3}\,\epsilon)\,\ln(2n/\delta)/\epsilon^2$

Example: A Bound [Ipsen & Wentworth 2012]

Label leverage scores so that $\mu \equiv \ell_{[1]} \geq \cdots \geq \ell_{[m]}$

$$\tau \equiv \sum_{j=1}^{t} \ell_{[j]} + \left(\frac{1}{\mu} - t\right) \ell_{[t+1]} \qquad \text{where} \quad t \equiv \lfloor 1/\mu \rfloor$$

SQ : c rows of Q sampled uniformly with replacement

Given ϵ, with probability at least $1 - \delta$

$$\kappa(SQ) \leq \sqrt{\frac{1+\epsilon}{1-\epsilon}}$$

provided $c \geq m\,\mu\,(2\,\tau + \frac{2}{3}\,\epsilon)\ \ln(2n/\delta)/\epsilon^2$

In practice: $\kappa(SQ) \leq 10$ with at least 99% probability
provided $c \geq m\,\mu\,(2.1\,\tau + .7)\,(\ln(2n) + 4.7)$

Practical Questions

For various values of m, n, coherence, leverage scores, and c:

- How does this bound compare to existing bounds?
- How does this bound compare to the condition numbers of sampled matrices?
- What are the smallest values of c for which the bound becomes informative?
- What are the smallest values of c for which the sampled matrices have full rank?
- How large are the condition numbers of the sampled matrices?
- When is sampling 10% of the rows sufficient?
- For a given coherence, are there leverage score distributions that make sampling harder?
- How large can n/m be, before sampling becomes inefficient?

The App

Matlab App with GUI (arXiv:1402.0642)

kappaSQ $= \kappa(SQ)$

1. Plot *probabilistic bounds* for $\kappa(SQ)$
2. Run *experiments*, and plot *actual values* of $\kappa(SQ)$

Matlab App with GUI (arXiv:1402.0642)

kappaSQ $= \kappa(SQ)$

1. Plot *probabilistic bounds* for $\kappa(SQ)$
2. Run *experiments*, and plot *actual values* of $\kappa(SQ)$

Features of kappaSQ

- Four randomized sampling methods

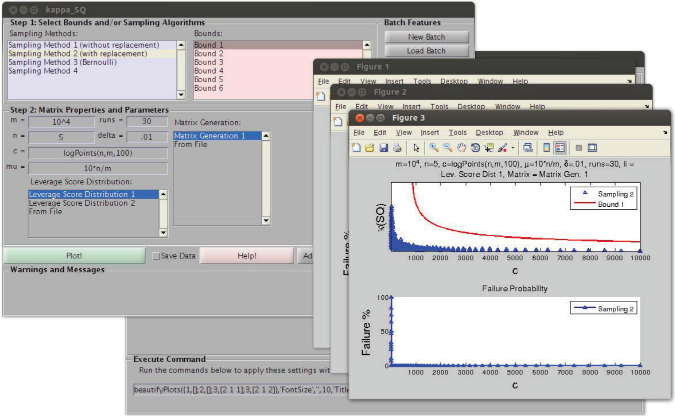
  Uniform & leverage score sampling with replacement
  Uniform sampling without replacement
  Bernoulli sampling

- Six probabilistic bounds
- Test matrix generation (for given m, n and leverage scores)
- Leverage score distributions: "adversarial" or not

  (for given m, n and coherence)

- "Publication-ready" plots
- Easy incorporation of user's own codes

Screen Shot of kappaSQ



Advantages of kappaSQ

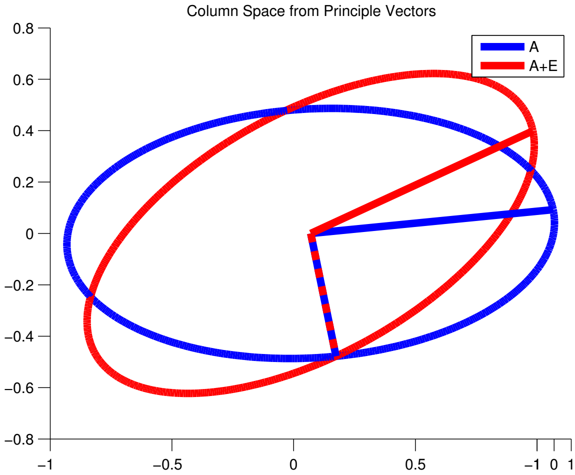
- Insight into behavior of sampling methods & bounds
  in a practical, non-asymptotic context

  Compare different bounds
  Compare bounds to experiments
  Explore the limits of randomized sampling

- Intuitive user interface (plot button)
- Visually appealing plots
- Sensible default values
- Extensive facilities for customizing plots (beautify)
- Very little familiarity with Matlab required

Sensitivity of Leverage Scores to Rotation of Subspace

Rotation of Subspace



Exact and Perturbed Leverage Scores

- Exact subspace: range($\mathcal{A}$), $\mathcal{A}$ is $m \times n$ with rank($\mathcal{A}$) $= n$

  Exact leverage scores

  $$\ell_j(\mathcal{A}) = \|e_j^T A\|_2^2 \qquad 1 \leq j \leq m$$

  where A is orthonormal basis for range($\mathcal{A}$)

Exact and Perturbed Leverage Scores

- Exact subspace: range($\mathcal{A}$), $\mathcal{A}$ is $m \times n$ with rank($\mathcal{A}$) $= n$

  Exact leverage scores

  $$\ell_j(\mathcal{A}) = \|e_j^T A\|_2^2 \qquad 1 \leq j \leq m$$

  where A is orthonormal basis for range($\mathcal{A}$)

- Rotated subspace: range($\mathcal{B}$), $\mathcal{B}$ is $m \times n$ with rank($\mathcal{B}$) $= n$

  Perturbed leverage scores

  $$\ell_j(\mathcal{B}) = \|e_j^T B\|_2^2 \qquad 1 \leq j \leq m$$

  where B is orthonormal basis for range($\mathcal{B}$)

Question: How close are $\ell_j(B)$ to $\ell_j(A)$?

Principal Angles between Column Spaces

A and B are $m \times n$ with orthonormal columns

- SVD of $n \times n$ matrix $A^T B = U \Sigma V^T$

$$\Sigma = \operatorname{diag}\begin{pmatrix} \cos\theta_1 & \cdots & \cos\theta_n \end{pmatrix}$$

- Principal angles θ_j between range(A) and range(B)

$$1 \geq \cos\theta_1 \geq \ldots \geq \cos\theta_n \geq 0$$

$$0 \leq \theta_1 \leq \cdots \leq \theta_n \leq \pi/2$$

- Special cases

  If range$(A) =$ range(B) then $\Sigma = I_n$ and all $\theta_j = 0$

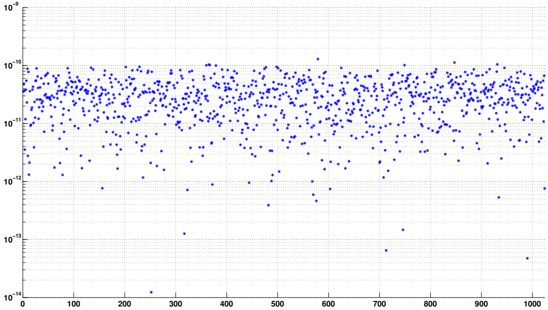
  If $A^T B = 0$ then $\Sigma = 0$ and all $\theta_j = \pi/2$

Uniform Leverage Scores

A is $m \times n$ Hadamard $\quad m = 1024,\ n = 50, \quad$ all leverage scores $\ell_j(A) = n/m$

Angles: $\quad \cos\theta_1 = 1 \quad \sin\theta_n \approx 10^{-8}$

Absolute errors $|\ell_j(B) - \ell_j(A)|$ vs index j



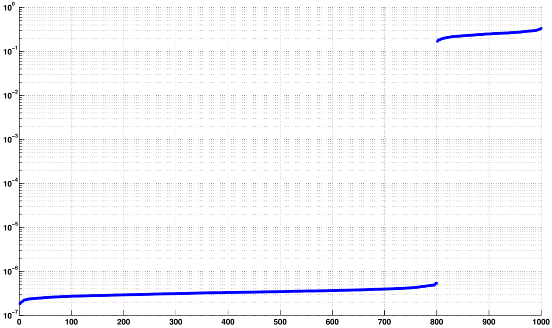
Absolute errors $\leq 10^{-10}$

20% Large Leverage Scores

A is $m \times n$ $\quad m = 1000, \; n = 50$

800 small $\ell_j(A) \leq 10^{-6}$ $\quad$ 200 large $\ell_j(A) \approx .3$

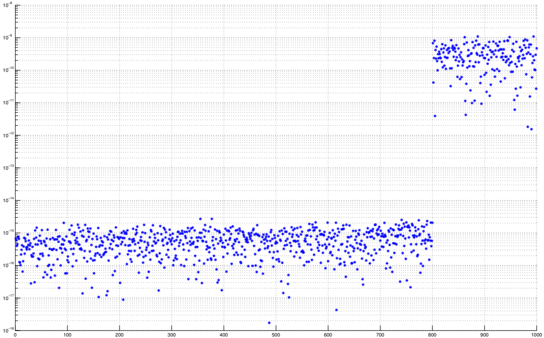
Leverage scores $\ell_j(A)$ vs index j



20% Large Leverage Scores: Absolute Errors

Angles: $\cos\theta_1 = 1$ $\sin\theta_n \approx 10^{-8}$

Absolute errors $|\ell_j(B) - \ell_j(A))|$ vs index j

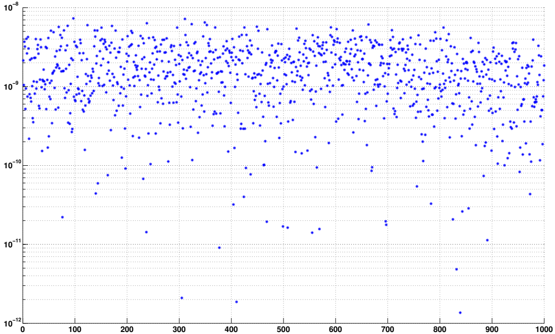


Small leverage scores: absolute errors $\leq 10^{-15}$
Large leverage scores: absolute errors $\leq 10^{-9}$

20% Large Leverage Scores: Relative Errors

Angles: $\cos\theta_1 = 1$ $\sin\theta_n \approx 10^{-8}$

Relative errors $|\ell_j(B) - \ell_j(A))|/|\ell_j(A)|$ vs index j



All leverage scores have relative errors $\leq 10^{-8}$

Sensitivity of Leverage Scores to Subspace Rotation

A and B are $m \times n$ with orthonormal columns

- Angles between range(A) and range(B)

$$0 \leq \theta_1 \leq \cdots \leq \theta_n \leq \pi/2$$

- Perturbation bounds

$$\begin{aligned} \ell_j(B) &\leq \left(\cos\theta_1 \sqrt{\ell_j(A)} + \sin\theta_n \sqrt{1-\ell_j(A)}\right)^2 \\ \ell_j(A) &\leq \left(\cos\theta_1 \sqrt{\ell_j(B)} + \sin\theta_n \sqrt{1-\ell_j(B)}\right)^2 \qquad 1 \leq j \leq m \end{aligned}$$

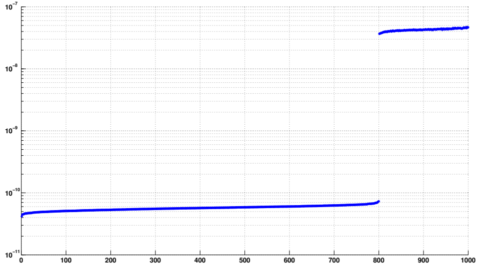
Leverage scores of A and B are close,
if all angles between range(A) and range(B) are small

Tightness of Bound

A is $m \times n$ $\quad m = 1000,\ n = 50 \quad \sin\theta \approx 10^{-8}$

800 small $\ell_j(A) \leq 10^{-6} \quad$ 200 large $\ell_j(A) \approx .3$

Bound - $\ell_j(B)$ vs index j



Bound tight in relative sense: $\quad \text{Bound} - \ell_j(B) \lesssim 10^{-8}\,\ell_j(B)$

Large Leverage Scores

A and B are $m \times n$ with orthonormal columns

- Perturbed leverage score is large: $\ell_k(B) \geq 1/2$ for some k

$$\frac{\ell_k(A)}{(\cos\theta_1 + \sin\theta_n)^2} \leq \ell_k(B) \leq \left(\cos\theta_1 \sqrt{\ell_k(A)} + \sin\theta_n \sqrt{1 - \ell_k(A)}\right)^2$$

- Exact leverage score is also large: $\ell_k(A) \geq 1/2$

$$\frac{\ell_k(A)}{(\cos\theta_1 + \sin\theta_n)^2} \leq \ell_k(B) \leq (\cos\theta_1 + \sin\theta_n)^2 \, \ell_k(A)$$

Upper and lower bounds for large leverage scores

Summary: Sensitivity to Subspace Rotation

Leverage scores of matrices with orthonormal columns

- Leverage scores are close, if subspace rotation small
- Small leverage scores as insensitive as large ones
- Our perturbation bounds are qualitatively informative
- Simpler bounds for special cases:
  Large leverage scores, $m = 2n$

Sensitivity of Leverage Scores to Matrix Perturbations

Norm-wise Relative Matrix Perturbations

- $\mathcal{A}$ and $\mathcal{A} + \mathcal{E}$ are $m \times n$ of rank n
- Two-norm condition number and relative perturbation

$$\kappa \equiv \|\mathcal{A}\|_2 \, \|\mathcal{A}^\dagger\|_2 \qquad \epsilon \equiv \|\mathcal{E}\|_2 / \|\mathcal{A}\|_2$$

- First-order bounds

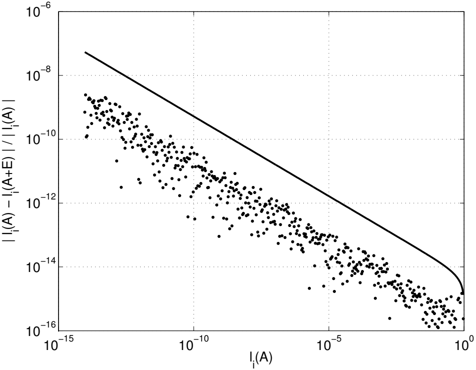
$$\left| \frac{\ell_j(\mathcal{A}+\mathcal{E}) - \ell_j(\mathcal{A})}{\ell_j(\mathcal{A})} \right| \;\leq\; 2\sqrt{\tfrac{1-\ell_j(\mathcal{A})}{\ell_j(\mathcal{A})}}\; \kappa\, \epsilon + \mathcal{O}(\epsilon^2)$$

- Sensitivity proportional to condition number of $\mathcal{A}$
  Leverage scores sensitive if $\kappa \gg 1$
- Small leverage scores more sensitive than large ones

Small Perturbations

$\mathcal{A}$ is $m \times n$ $\quad m = 500,\ n = 15,\ \ \kappa = 1,\ \ \epsilon \approx 10^{-15}$

Relative error $|\ell_j(\mathcal{A}+\mathcal{E}) - \ell_j(\mathcal{A})|/\ell_j(\mathcal{A})$ vs $\ell_j(\mathcal{A})$

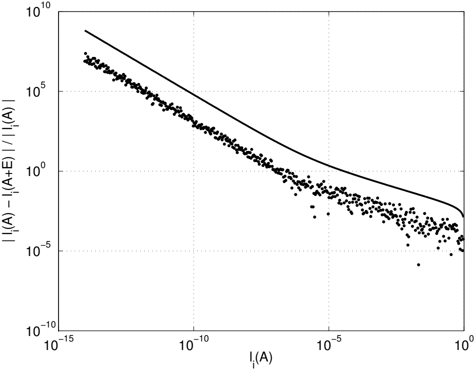


Bound reflects behavior of error

Large Perturbations

$\mathcal{A}$ is $m \times n$ $\quad m = 500,\ n = 15,\ \kappa = 1,\ \epsilon \approx 10^{-3}$

Relative error $|\ell_j(\mathcal{A}+\mathcal{E}) - \ell_j(\mathcal{A})|/\ell_j(\mathcal{A})$ vs $\ell_j(\mathcal{A})$



Bound reflects behavior of error

Summary

Motivation

Sampling rows from matrices Q with orthonormal columns
Want: Condition number of sampled matrix $\kappa(SQ)$
Condition number depends on leverage scores
Largest leverage scores = coherence

Matlab App kappaSQ (arXiv:1402.0642)

Sampling & bounds in non-asymptotic context
Compare different probabilistic bounds
Run experiments and test tightness of bounds
"Publication-ready" plots

Sensitivity of Leverage Scores, to:

- Subspace rotations

  Leverage scores close if subspace rotation small
  Small leverage scores as insensitive as large ones

- Relative matrix perturbations

  Sensitivity depends on condition number of matrix
  Small leverage scores more sensitive than large ones