Accuracy and Stability Issues for Randomized Matrix Algorithms: Sensitivity of Leverage Scores

Ilse Ipsen

Joint work with Thomas Wentworth

North Carolina State University
Raleigh, NC, USA

Research supported by NSF CISE CCF and DARPA XData
Statistical Leverage Scores

Hoaglin & Welsch 1978
Velleman & Welsch 1981
Chatterjee & Hadi 1986

Given: Real \( m \times n \) matrix \( A \), \( m \geq n \), \( \text{rank}(A) = n \)
Want: Potential outliers in \( \min_x \|Ax - b\| \) (two-norm)

- **Hat matrix**: Orthogonal projector onto \( \text{range}(A) \)
  \[
  H = A (A^T A)^{-1} A^T
  \]

- **Leverage scores of** \( A \)
  \[
  \ell_j(A) = H_{jj} \quad 1 \leq j \leq m
  \]

- **Least squares fit**: \( \hat{b} = Hb \)
  - If \( \ell_k(A) = 1 \) then \( b_k \) has maximal leverage: \( \hat{b}_k = b_k \)
  - If \( \ell_k(A) = 0 \) then \( b_k \) has zero leverage over \( \hat{b}_k \)
Overview

1. **Computation**, and use of leverage scores
2. **Coherence**: Largest leverage score
3. **Sensitivity** of leverage scores to **subspace angles**
   
   *Coherence*

   *Large leverage scores*

4. **Sensitivity** of leverage scores to **matrix perturbations**
Computation, and Use of Leverage Scores
Computation of Leverage Scores

Real \( m \times n \) matrix \( A \) with \( \text{rank}(A) = n \)

Hat matrix \( H = A (A^T A)^{-1} A^T \)

Leverage scores \( \ell_j(A) = H_{jj} \quad 1 \leq j \leq m \)

**Singular Value Decomposition**

\[ A = U \Sigma V^T \quad u^T u = I_n \]

Hat matrix \( H = UU^T \)

\[
\ell_j(A) = \| e_j^T U \|^2 \quad 1 \leq j \leq m
\]

**QR decomposition**

\[ A = QR \quad Q^T Q = I_n \]

Hat matrix \( H = QQ^T \)

\[
\ell_j(A) = \| e_j^T Q \|^2 \quad 1 \leq j \leq m
\]
Leverage Scores for Randomized Algorithms

[Drineas, Mahoney et al. 2006-2013]

Use of leverage scores:

As sampling probabilities
To analyze performance of uniform sampling strategies

Randomized subset selection [Boutsidis, Mahoney & Drineas 2010]

Given: $m \times n$ matrix $A$ with $\text{rank}(A) = n$
Want: $k$ most important rows of $A$

Idea: Sample row $j$ of $A$ with probability $p_j = \ell_j(A)/n$
Coherence: Largest Leverage Score
Coherence

Donoho & Huo 2001: Mutual coherence of two bases
Candés, Romberg & Tao 2006
Candés & Recht 2009: Matrix completion

Coherence of $m \times n$ matrix $A$ with $\text{rank}(A) = n$

$$\mu(A) = \max_{1 \leq j \leq m} \ell_j(A)$$

Low coherence $\Rightarrow$ uniform leverage scores

- $n/m \leq \mu(A) \leq 1$
- **Maximal** coherence: $\mu(A) = 1$
  - At least one basis vector for $\text{range}(A)$ is a canonical vector
- **Minimal** coherence: $\mu(A) = n/m$
  - Orthonormal bases for $\text{range}(A)$ are like columns of a Hadamard matrix
- Coherence measures correlation with a standard basis
Sensitivity of Leverage Scores to Subspace Angles
Exact and Perturbed Leverage Scores

- **Exact** matrix: \( A \) is \( m \times n \) with \( \text{rank}(A) = n \)

  Exact leverage scores

  \[
  \ell_j(A) = \| e_j^T A \|^2 \quad 1 \leq j \leq m
  \]

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

- **Perturbed** matrix: \( B \) is \( m \times n \) with \( \text{rank}(B) = n \)

  Perturbed leverage scores

  \[
  \ell_j(B) = \| e_j^T B \|^2 \quad 1 \leq j \leq m
  \]

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

**Question:** How close is \( \ell_j(B) \) to \( \ell_j(A) \)?
Principal Angles between Column Spaces

$A$ and $B$ are $m \times n$ with orthonormal columns, $A^T A = B^T B = I_n$

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

  $$\Sigma = \text{diag} \left( \cos \theta_1, \ldots, \cos \theta_n \right)$$

- **Principal angles** $\theta_j$ between range($A$) and range($B$)

  $$1 \geq \cos \theta_1 \geq \ldots \geq \cos \theta_n \geq 0$$
  $$0 \leq \theta_1 \leq \ldots \leq \theta_n \leq \frac{\pi}{2}$$

- **Special cases**
  
  If $A = B$ then $\Sigma = I_n$ and all $\theta_j = 0$
  
  If $A^T B = 0$ then $\Sigma = 0$ and all $\theta_j = \pi/2$
Sensitivity of Leverage Scores to Angles

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

  \[ 0 \leq \theta_1 \leq \cdots \leq \theta_n \leq \pi/2 \]

- **Leverage score bounds**

  \[
  \ell_j(B) \leq \left( \cos \theta_1 \sqrt{\ell_j(A)} + \sin \theta_n \sqrt{1 - \ell_j(A)} \right)^2 \\
  \ell_j(A) \leq \left( \cos \theta_1 \sqrt{\ell_j(B)} + \sin \theta_n \sqrt{1 - \ell_j(B)} \right)^2 \\
  \text{for } 1 \leq j \leq m
  \]

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

$\mathcal{A}$ is $m \times n$ Hadamard, $m = 1024$, $n = 50$, leverage scores are $n/m$

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

Relative error: $(\ell_j(B) - \ell_j(A))/\ell_j(A)$

Relative bound reflects behaviour of errors
20% Large Leverage Scores

\(A \) is \( m \times n \) \( m = 1000, n = 50 \), and 200 large leverage scores

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

Relative error: \( (\ell_j(B) - \ell_j(A))/\ell_j(A) \) Relative bound

Bound tighter for large leverage scores
**Small Leverage Scores, and Angles**

- **Assume:** Bounded angles
  \[ 0 \leq \theta_1 \leq \cdots \leq \theta_n \leq \pi/4 \]
- **Large leverage score:** \( \ell_k(A) \geq 1/2 \) for some \( k \)
- **Bound for perturbed leverage scores**
  \[ \left(1 - \sqrt{2} \sin \theta_n\right)^2 \ell_k(A) \leq \ell_k(B) \leq (1 + \sin \theta_n)^2 \ell_k(A) \]

Upper and lower bounds for large leverage scores
Coherence and Angles

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

  \[
  0 \leq \theta_1 \leq \cdots \leq \theta_n \leq \pi/2
  \]

- **Coherence**: \( \mu(A) = \max_{1 \leq j \leq m} \ell_j(A) \)

- **Bound for perturbed coherence**

  \[
  \frac{\mu(A)}{\gamma} \leq \mu(B) \leq \gamma \mu(A)
  \]

  where

  \[
  \gamma = \left( \cos \theta_1 + \sin \theta_n \sqrt{\frac{m}{n}} \right)^2
  \]

  Coherence is **sensitive** if

  - Large aspect ratio: \( m \gg n \)
  - Large angles between range($A$) and range($B$)
Sensitivity of Leverage Scores to Matrix Perturbations
Bound for Angles in terms of Perturbations

- \(A\) and \(A + E\) are \(m \times n\) of rank \(n\)
- Condition number and relative perturbation
  \[\kappa = \|A\| \|A^\dagger\| \quad \epsilon = \|E\| / \|A\|\]
- Largest angle between range(\(A\)) and range(\(A + E\)): \(\theta_n\)
- Assume: Perturbation \(\epsilon < 0.5 / \kappa\)
- Bound for largest angle
  \[\sin \theta_n \leq 2 \kappa \epsilon\]

All angles between range(\(A\)) and range(\(A + E\)) are small if \(A\) is well-conditioned with respect to inversion.
Perturbation of Coherence

- \( A \) and \( A + \mathcal{E} \) are \( m \times n \) of rank \( n \)
- Condition number and relative perturbation
  \[
  \kappa = \|A\| \|A^\dagger\| \quad \epsilon = \|\mathcal{E}\| / \|A\|
  \]
- Coherence: \( \mu(A) = \max_{1 \leq j \leq m} \ell_j(A) \)
- Assume: Perturbation \( \epsilon < \frac{.5}{\kappa} \)
- Bound for perturbed coherence
  \[
  \frac{\mu(A)}{\gamma} \leq \mu(B) \leq \gamma \mu(A) \quad \gamma = \left(1 + 2 \sqrt{\frac{m}{n}} \kappa \epsilon \right)^2
  \]

Coherence is sensitive to perturbations if

- Large aspect ratio: \( m \gg n \)
- \( A \) is ill-conditioned with respect to inversion
Perturbation of Large Leverage Scores

- $\mathcal{A}$ and $\mathcal{A} + \mathcal{E}$ are $m \times n$ of rank $n$

- Condition number and relative perturbation

$$\kappa = \|\mathcal{A}\| \|\mathcal{A}^\dagger\| \quad \epsilon = \|\mathcal{E}\|/\|\mathcal{A}\|$$

- Large leverage scores: $\ell_k(\mathcal{A}) \geq 1/2$ for some $k$

- Assume: Perturbation $\epsilon < 0.3/\kappa$

- Relative error for large leverage scores

$$\left| \frac{\ell_k(\mathcal{A} + \mathcal{E}) - \ell_k(\mathcal{A})}{\ell_k(\mathcal{A})} \right| \leq 4\kappa\epsilon (\kappa\epsilon + 1)$$

Large leverage scores are insensitive to perturbations if $\mathcal{A}$ is well-conditioned with respect to inversion.
Well-Conditioned Matrix

\( A \) is \( m \times n \quad m = 1000, \ n = 50 \)

Condition number: \( \kappa \approx 23 \)  \quad Relative perturbation: \( \epsilon \approx 10^{-8} \)

Relative error: \( |\ell_j(B) - \ell_j(A)| / \ell_j(A) \)  \quad Bound

Bound informative even if matrix has no large leverage scores
Moderately Conditioned Matrix

\( A \) is \( m \times n \quad m = 1000, \quad n = 50 \)

Condition number: \( \kappa \approx 10^8 \quad \) Relative perturbation: \( \epsilon \approx 10^{-8} \)

Relative error: \( |\ell_j(B) - \ell_j(A)| / \ell_j(A) \quad \) Bound

Bound informative for all leverage scores (not just large ones)
Summary

Leverage scores

Two-norms of rows of $m \times n$ orthonormal matrices
Sampling probabilities in randomized algorithms

Coherence

Largest leverage score
Performance analysis of sampling strategies

Sensitivity analysis
Relative error bounds for leverage scores of exact and perturbed matrix

Angles between column spaces
Condition number and matrix perturbation

Leverage scores insensitive if

Angles are small
Matrix is well-conditioned

Coherence more sensitive if $m \gg n$
Future Work

- Sampling strategies only need the correct exponent. Are relative error bounds too strong?

- Sampling strategies depend on large leverage scores. Tighter bounds targeted at large leverage scores.

- Extend sensitivity analysis to
  - Rank deficient matrices
  - Low-rank approximations
  - Large perturbations (missing data)
  - Structured perturbations (categorical data)