Sensitivity of Leverage Scores and Coherence for Randomized Matrix Algorithms

Ilse Ipsen

Joint work with Thomas Wentworth

North Carolina State University
Raleigh, NC, USA

Research supported by NSF CISE CCF and DARPA XData

Randomized Algorithms Collaborators

John Holodnak (NCState)

Michael Mahoney (Stanford)
Petros Drineas (RPI)

Costas Bekas (IBM Zürich)
Haim Avron (IBM Watson)
Christos Boutsidis (IBM Watson)
Vikas Sindhwani (IBM Watson)
Anju Kambadur (IBM Watson)
Ken Clarkson (IBM Almaden)
David Woodruf (IBM Almaden)

Tamás Budavári (Johns Hopkins)
Ching-Wa Yip (Johns Hopkins)

What are Leverage Scores and Coherence?

Diagonal elements of (orthogonal) projector

- Orthogonal projector
  Square matrix P with $P^2 = P$ and $P^T = P$
- Leverage scores
  $$\ell_j = P_{jj}$$
- Coherence = largest leverage score
  $$\mu = \max \ell_j$$
- Leverage scores of $m \times n$ matrix A with $m \geq n$
  $$\ell_j(A) = P_{jj} \qquad 1 \leq j \leq m$$
  where P orthogonal projector onto range(A)

Overview

1. Applications
2. Computation
3. Sensitivity to subspace angles
4. Sensitivity to matrix perturbations

Applications

Statistics
Astronomy
Physics
Graph theory
Compressed sensing, matrix completion
Randomized algorithms

Statistical Leverage Scores

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

Purpose: Regression diagnostics

- $\min_x \|Ax - b\|_2$ where A is $m \times n$ with $\text{rank}(A) = n$
- Hat matrix: Orthogonal projector onto $\text{range}(A)$

$$H = A\,(A^T A)^{-1}\,A^T$$

- Leverage scores of A

$$\ell_j(A) = H_{jj} \qquad 1 \leq j \leq m$$

- 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 $\hat{b}_k$

Leverage scores determine potential outliers

Astronomy

[Yip, Mahoney, Szalay, Csabai, Budavári, Wyse & Dobos 2013]

Purpose: Evolution of galaxies

- Informative wave length regions
- Matrix A

  Row i: Wavelength λ_i
  Column j: Spectrum of jth Simple Stellar Population
  Given rank parameter k: Truncated SVD A_k

- Importance of wavelength λ_i: $\ell_i(A_k)$
- Importance of wavelength region $[\lambda_s, \lambda_e]$

$$\text{Lick index} \quad \sum_{i=s}^{e} \ell_i(A_k)$$

Leverage scores determine important wave lengths

Physics

[Bekas, Kokiopoulou & Saad 2008]

Purpose: Physical properties of complex materials

- Electronic structure calculations (DFT)
- Discretization of Hamiltonian: Symmetric matrix A
- Invariant subspace V associated with n smallest eigenvalues
  V is $m \times n$ with $V^T V = I_n$
- Functional density matrix $P = VV^T$
- Charge density at jth point $\ell_j = P_{jj}$

Leverage scores yield charge densities

Graph Theory

[Drineas & Mahoney 2010]

Purpose: Analyse diffusion processes and random walks on graphs
Detect clusters and community structure in large networks

- Undirected unweighted graph with n vertices and m edges
- $m \times n$ edge incidence matrix B
  If edge between vertices i and j then $b_{ij} = 1$ and $b_{ji} = -1$
- Graph Laplacian $L = B^T B$
- Effective resistance matrix $R = BL^\dagger B^T$
- Effective resistance of edge j $\ell_j(B) = R_{jj}$

Leverage scores yield effective resistance

Compressed Sensing and Matrix Completion

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

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

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

Low coherence $\Rightarrow$ uniform leverage scores

- $n/m \le \mu(A) \le 1$
- 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 columns of a Hadamard matrix

Coherence reflects difficulty of recovering matrix from sampling

Randomized Algorithms

- Leverage scores for analysis

  Randomized preconditioners for least squares

  [Avron, Maymounkov & Toledo 2010]

  Sampling rows from from orthonormal matrices

  [Ipsen & Wentworth 2012]

- Leverage scores for importance sampling

  Randomized subset selection

  [Boutsidis, Mahoney & Drineas 2010]

  Low rank approximation

  [Drineas, Mahoney & Muthukrishnan 2008, Mahoney & Drineas 2009]

  Randomized least squares

  [Drineas, Mahoney & Muthukrishnan 2007, 2008]

  [Drineas, Mahoney, Muthukrishnan & Sarlós 2007]

Computation of Leverage Scores

Dense matrices
Other approaches

Computation: Dense Matrices

Given: *Real* $m \times n$ *matrix* A *with* $\text{rank}(A) = n$

Want: *Leverage scores* $\ell_j(A) = P_{jj} \qquad 1 \leq j \leq m$

Projector $P = A\,(A^T A)^{-1}\,A^T$

- Singular Value Decomposition $A = U\,\Sigma V^T \qquad U^T U = I_n$
  Projector $P = UU^T$

$$\ell_j(A) = \|e_j^T\, U\|_2^2 \qquad 1 \leq j \leq m$$

- QR decomposition $A = Q\,R \qquad Q^T Q = I_n$
  Projector $P = QQ^T$

$$\ell_j(A) = \|e_j^T\, Q\|_2^2 \qquad 1 \leq j \leq m$$

Computation: Other Approaches

- Sparse matrices: Estimating diagonal elements of P by probing
  [Bekas, Kokiopoulou & Saad 2007, Coleman & Moré 1983]
- Fast randomized approximations
  [Mori & Talwalkar 2010, 2011]
  [Drineas, Magdon-Ismail, Mahoney & Woodruff 2011]

Connections to

- Eigenvector and invariant subspace localization
  [Vömel & Parlett 2011, Slanina & Konopásek 2010]
- Diagonal elements of matrix functions
- CS decomposition

Sensitivity of Leverage Scores to Subspace Angles

Exact and Perturbed Leverage Scores

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

  Exact leverage scores

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

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

Exact and Perturbed Leverage Scores

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

  Exact leverage scores

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

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

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

  Perturbed leverage scores

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

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

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, $A^TA = B^TB = I_n$

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

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

- 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 $A = B$ then $\Sigma = I_n$ and all $\theta_j = 0$

  If $A^TB = 0$ 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

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

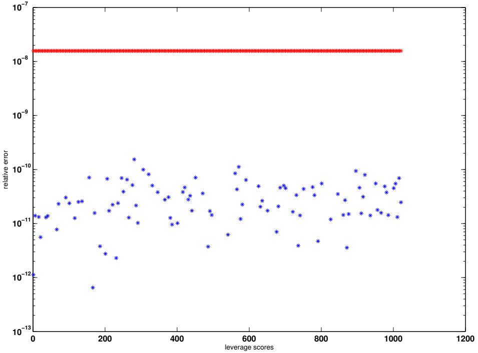
Leverage scores of A and B are close,
if all angles between range(A) and range(B) are small

Uniform Leverage Scores

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

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

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



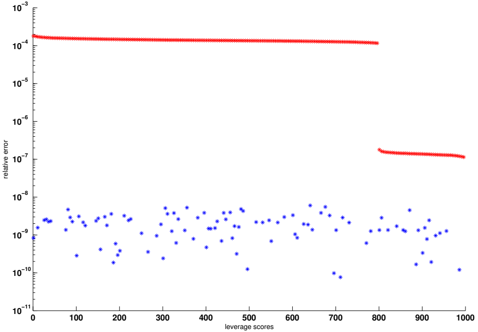
Bound reflects behaviour of errors

20% Large Leverage Scores

$\mathcal{A}$ is $m \times n$ $\quad m = 1000,\ n = 50,$ $\quad$ and 200 large leverage scores

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

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



Bound tighter for large leverage scores

Large Leverage Scores, and Angles

- Assume: Bounded angles

$$0 \le \theta_1 \le \cdots \le \theta_n \le \pi/4$$

- Large leverage score: $\ell_k(A) \ge 1/2$ for some k
- Bound for perturbed leverage scores

$$\left(1 - \sqrt{2}\sin\theta_n\right)^2 \ell_k(A) \le \ell_k(B) \le (1 + \sin\theta_n)^2 \ell_k(A)$$

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

$$\mu(A)/\gamma \leq \mu(B) \leq \gamma\, \mu(A)$$

where

$$\gamma = \left(\cos\theta_1 + \sin\theta_n \sqrt{\tfrac{m}{n}}\right)^2$$

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

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

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

- Largest angle between range($\mathcal{A}$) and range($\mathcal{A}+\mathcal{E}$): θ_n
- Assume: Perturbation $\epsilon < .5/\kappa$
- Bound for largest angle

$$\sin\theta_n \le 2\,\kappa\,\epsilon$$

All angles between range($\mathcal{A}$) and range($\mathcal{A}+\mathcal{E}$) are small
if $\mathcal{A}$ is well-conditioned with respect to inversion

Perturbation of Coherence

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

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

- Coherence: $\mu(\mathcal{A}) = \max_{1 \leq j \leq m} \ell_j(\mathcal{A})$
- Assume: Perturbation $\epsilon < .5/\kappa$
- Bound for perturbed coherence

$$\mu(\mathcal{A})/\gamma \leq \mu(\mathcal{A} + \mathcal{E}) \leq \gamma \, \mu(\mathcal{A}) \qquad \gamma = \left(1 + 2\sqrt{\tfrac{m}{n}}\, \kappa \, \epsilon\right)^2$$

Coherence is sensitive to perturbations if

Large aspect ratio: $m \gg n$

$\mathcal{A}$ *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

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

- Large leverage scores: $\ell_k(\mathcal{A}) \geq 1/2$ for some k
- Assume: Perturbation $\epsilon < .3/\kappa$
- Relative error for large leverage scores

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

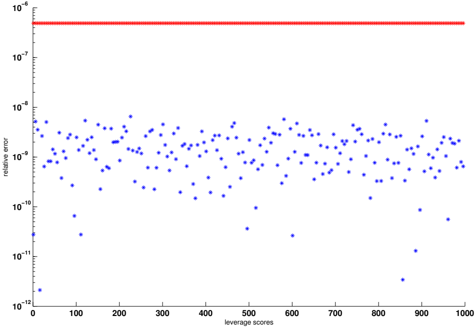
Large leverage scores are insensitive to perturbations
if $\mathcal{A}$ is well-conditioned with respect to inversion

Well-Conditioned Matrix

$\mathcal{A}$ 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



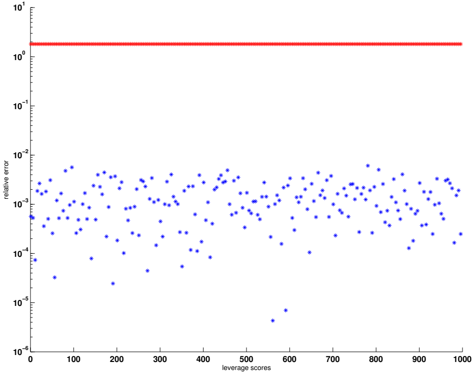
Bound informative even if matrix has no large leverage scores

Moderately Conditioned Matrix

$\mathcal{A}$ is $m \times n$ $\quad m = 1000,\ n = 50$

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

Diagonal elements of $m \times m$ orthogonal projectors
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
Underlying $m \times n$ matrix 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)