Randomly Sampling from Orthonormal Matrices: Coherence and Leverage Scores

Ilse Ipsen

Joint work with Thomas Wentworth

(thanks to Petros & Joel)

North Carolina State University
Raleigh, NC, USA

Research supported in part by NSF CISE CCF

This Talk

Given:

- Real $m \times n$ matrix Q with orthonormal columns, $Q^T Q = I_n$
- Real $c \times m$ "sampling" matrix S with $c \ll m$

Want: Probability that

1. SQ has full column rank $\quad (\mathrm{rank}(SQ) = n)$
2. Condition number: Given η

$$\kappa(SQ) = \|SQ\|_2 \, \|(SQ)^\dagger\|_2 \leq 1 + \eta$$

Motivation: `Blendenpik` [Avron, Maymounkov & Toledo 2010]

$\kappa(SQ) =$ condition number of preconditioned matrix

Outline

1. Three different uniform sampling strategies
2. Bounds for condition numbers of sampled matrices

   Based on coherence

   Lower bounds on number of samples}
3. Improving on coherence: Leverage scores
4. Summary

Three Different Uniform Sampling Strategies

Uniform sampling c rows out of m

1. Sampling without replacement
2. Sampling with replacement (Exactly(c))
3. Bernoulli sampling

Uniform Sampling without Replacement

[Gittens & Tropp 2011, Gross & Nesme 2010]

Choose random permutation $k_1, \ldots, k_m$ of $1, \ldots, m$

Sampling matrix $\quad S = \sqrt{\frac{m}{c}} \begin{pmatrix} e_{k_1}^T \\ \vdots \\ e_{k_c}^T \end{pmatrix}$

- S is $c \times m$, and samples exactly c rows
- Each row sampled at most once
- Expected value $\mathbf{E}(S^T S) = I_m$

Uniform Sampling with Replacement (Exactly(c))

[Drineas, Kannan & Mahoney 2006]

for $t = 1 : c$ **do**
  Sample k_t from $\{1, \ldots, m\}$ with probability $1/m$
    independently and with replacement
end for

Sampling matrix $$S = \sqrt{\frac{m}{c}} \begin{pmatrix} e_{k_1}^T \\ \vdots \\ e_{k_c}^T \end{pmatrix}$$

- S is $c \times m$, and samples exactly c rows
- A row can be sampled more than once
- Expected value $\mathbf{E}(S^T S) = I_m$

Bernoulli Sampling

[Avron, Maymounkov & Toledo 2010, Gittens & Tropp 2011]

$S = 0_{m\times m}$
for $j = 1 : m$ **do**

$$S_{jj} = \sqrt{\frac{m}{c}} \begin{cases} 1 & \text{with probability } \frac{c}{m} \\ 0 & \text{with probability } 1 - \frac{c}{m} \end{cases}$$

end for

- S is $m \times m$, and samples each row at most once
- Expected number of sampled (non zero) rows: c
- Expected value $\mathbf{E}(S^T S) = I_m$

Sampling Rows from Orthonormal Matrices

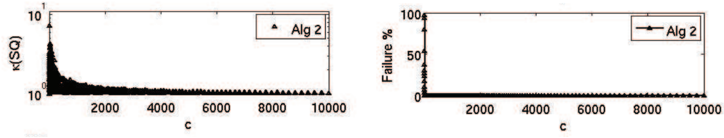
Sampling c rows from $m \times n$ matrix Q with $Q^T Q = I_n$
$m = 10^4$, $n = 5$ (30 runs for each value of c)

For each sampling strategy:

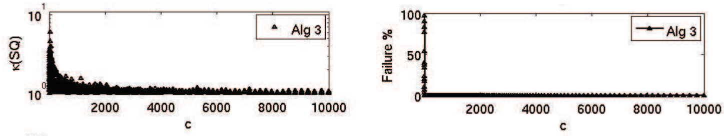
1. Two-norm condition number of SQ
   $\kappa(SQ) = \|SQ\|_2 \, \|(SQ)^\dagger\|_2$ (if SQ has full column rank)
2. Percentage of matrices SQ that are rank deficient

First Comparison

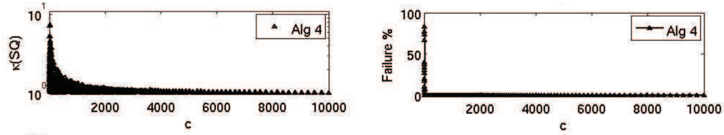
Sampling without replacement



Sampling with replacement (Exactly(c))

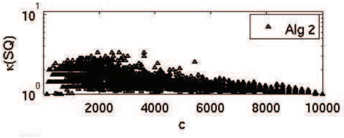


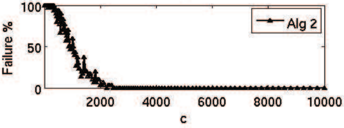
Bernoulli sampling



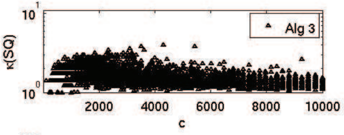
Second Comparison

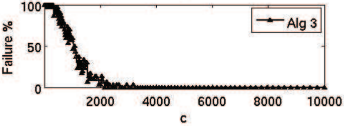
Sampling without replacement



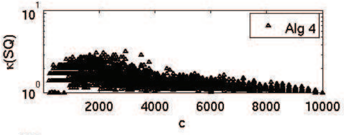


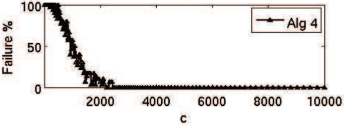
Sampling with replacement (Exactly(c))





Bernoulli sampling





Comparison of Sampling Strategies

Three strategies for sampling SQ:

- *Sampling without replacement*
- *Sampling with replacement (Exactly(c))*
- *Bernoulli sampling*

Summary

- *Little difference among the sampling strategies*
- *If SQ has full rank then $\kappa(SQ) \leq 10$*

Recommendation: Use sampling with replacement

- *Fast: Need to generate/inspect only c values*
- *Easy to implement*
- *Replacement does not affect accuracy*

(for small amounts of sampling)

Condition Number Bound

Same bound for all three sampling strategies

- $m \times n$ matrix Q with $Q^T Q = I_n$

  coherence $\mu \equiv \max_{1 \leq j \leq m} \|e_j^T Q\|_2^2$
- Number of samples $c \geq n$, $0 < \epsilon < 1$
- Failure probability

$$\delta \equiv n \left(f(-\epsilon)^{c/(m\,\mu)} + f(\epsilon)^{c/(m\,\mu)} \right)$$

where $f(x) = e^x (1+x)^{-(1+x)}$

With probability at least $1 - \delta$

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

Proofs

1. Sampling without replacement: [Tropp: *SRHT*, 2011]
2. Sampling with replacement:
   Use matrix Chernoff bound from [Tropp: *Userfriendly*, 2011]
3. Bernoulli sampling:
   Use matrix Chernoff bound from [Tropp: *Userfriendly*, 2011]
   See also [Gittens & Tropp, 2011]

Lower Bounds on Number of Samples

Failure probability

$$\delta \le n \left(f(-\epsilon)^{c/(m\,\mu)} + f(\epsilon)^{c/(m\,\mu)} \right)$$

where $f(x) = e^x (1+x)^{-(1+x)}$

Number of samples

$$c \ge \mu\, m \ln(2n/\delta) / |\ln(f(\epsilon))|$$

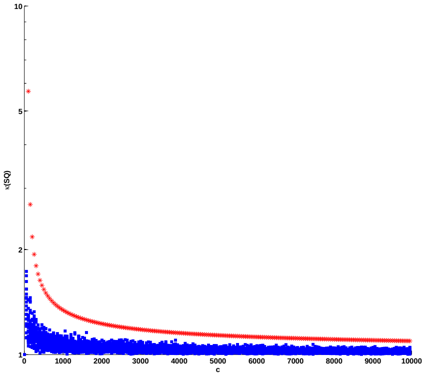
- Condition number: For $\kappa(SQ) \le 10$ need $\epsilon = 99/101$

$$c \ge 2.7\, \mu\, m \ln(2n/\delta)$$

- If Q has minimal coherence $\mu = n/m$

$$c \ge 2.7\, n \ln(2n/\delta)$$

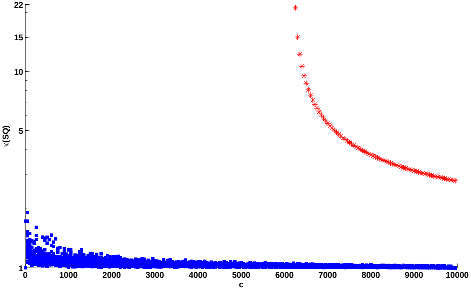
Tightness of Bound: Low Coherence ($\mu = 1.5n/m$)



Numerical bound for $c \geq 104$, lower bound $c \geq 108$

($m = 10^4$, $n = 4$, $\delta = .01$)

Tightness of Bound: Higher Coherence ($\mu = 100n/m$)



Numerical bound $c \geq 6,254$, lower bound $c \geq 7,219$
$(m = 10^4,\ n = 4,\ \delta = .01)$

Conclusions from the Bound

Input: $m \times n$ matrix Q with orthonormal columns

- Bound is based on coherence $\mu = \max_{1 \leq j \leq m} \|e_j^T Q\|_2^2$
- Predicts correct magnitude for condition number of sampled matrix, even for small matrix dimensions
- Required number of samples $c \geq m \, \mu \ln n$

  But too pessimistic for matrices with higher coherence
- Informative only for matrices

  that are very tall and skinny, $m \gg n$

  with almost minimal coherence $\mu \approx n/m$

Improving on Coherence: Leverage Scores

Idea: Use all row norms

- Q is $m \times n$ with orthonormal columns
- Leverage scores = row norms^2

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

[Hoaglin & Welsch 1978], [Chatterjee & Hadi 1986]
[Drineas, Mahoney & al], [Avron & Toledo]

- Coherence $\mu = \max_j \ell_j$
- Low coherence $\approx$ uniform leverage scores

Leverage Score Bound for Sampling with Replacement (Exactly(c))

- $m \times n$ matrix Q with orthonormal columns
- Leverage scores $\ell_j = \|e_j^T Q\|_2^2, \quad \mu = \max_{1 \leq j \leq m} \ell_j$

$$L = \mathrm{diag}\begin{pmatrix} \ell_1 & \ldots & \ell_m \end{pmatrix}$$

- $0 < \epsilon < 1$

Failure probability

$$\delta = 2n \exp\left(-\frac{3}{2}\,\frac{c\,\epsilon^2}{m\,(3\,\|Q^T L Q\|_2 + \mu\,\epsilon)}\right)$$

With probability at least $1 - \delta$: $\quad \kappa(SQ) \leq \sqrt{\frac{1+\epsilon}{1-\epsilon}}$

Leverage Scores vs. Coherence

- Failure probability

$$\delta = 2n \exp\left(-\frac{3}{2}\,\frac{c\,\epsilon^2}{m\,(3\,\|Q^T L Q\|_2 + \mu\,\epsilon)}\right)$$

- Bounds in terms of coherence:

$$\mu^2 \leq \|Q^T L Q\|_2 \leq \mu$$

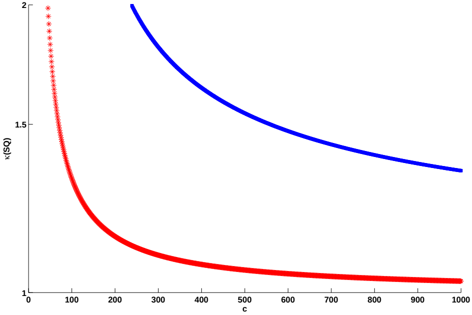
- Estimation in terms of largest leverage scores
  If $k = 1/\mu$ is an integer then

$$\|Q^T L Q\|_2 \leq \mu \sum_{j=1}^{k} \ell_{[j]}$$

where $\ell_{[1]} \geq \cdots \geq \ell_{[m]}$

Improvement with Leverage Score Bound

Low coherence: $\mu = 1.5n/m$, small amounts of sampling

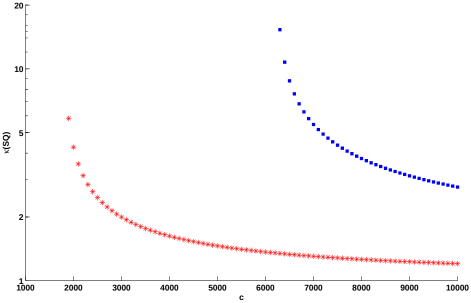


Leverage score bound vs. Coherence bound

$(m = 10^4,\ n = 4,\ \delta = .01)$

Improvement with Leverage Score Bound

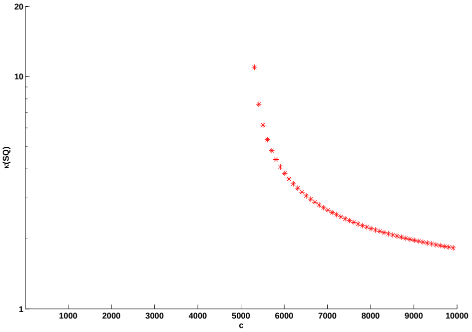
Higher coherence: $\mu = 100n/m$, large amounts of sampling



Leverage score bound vs. Coherence bound
($m = 10^4$, $n = 4$, $\delta = .01$)

Improvement with Leverage Score Bound

More columns, higher coherence: $\mu = 100n/m$



Coherence bound does not hold

$(m = 10^4,\ n = 10,\ \delta = .01)$

Summary

- Sampling rows from $m \times n$ matrices with orthonormal columns
- Want: Bounds for condition number of sampled matrices
- Three different sampling strategies:
  Without replacement, with replacement, Bernoulli
- Numerical experiments:
  All strategies behave the same for $c < m$
- Recommendation: Sampling with replacement

  Fast: Need to generate/inspect only c values
  Easy to implement
  Replacement does not affect accuracy
  (for small amounts of sampling)

Summary (cdt)

- Bounds for condition number of sampled matrices

  Explicit and non-asymptotic
  Predictive even for small matrix dimensions

- Bound based on coherence μ:

  Same bound for all strategies
  Number of samples $c \geq \mu\, m \ln n$

- Bound based on leverage scores:

  Tighter than coherence-based bound
  Holds for smaller number of samples c,
  higher coherence and fatter matrices