A Matrix-Free, Transpose-Free Norm Estimator

Ilse Ipsen

Department of Mathematics

North Carolina State University

Raleigh, USA

Joint work with Alina Chertock, Pierre Gremaud, Tim Kelley

Motivation

Flow in unsaturated porous media

[Kelley, Kees, Miller, Tocci, 1998]

- Solution of DAE $F(x) = 0$ by Newton iterative method
- To decide whether to update Jacobian need to estimate $||A||$, where $A = [F'(x)]^{-1}$
- A is sparse
- A only accessible as matvec Ax
- Transpose of A not available

Conventional Norm Estimation

Given: real square matrix A
Want: estimate of $||A||$

- p-norm power method estimator
  Generalizes power method on $A^T A$
- LAPACK 1-norm estimator
  Improvement of 1-norm power method estimator
- LINPACK-type estimators
  Require factorization of A

Statistical Norm Estimation

A is real $n \times n$ matrix
z is uniformly distributed on unit n-sphere

Dixon 1983: A symmetric positive definite

$$\mathrm{Prob}\left(z^T A z \geq \epsilon \|A\|_2\right) \geq 1 - .8\sqrt{n\epsilon}$$

Kuczyński & Woźniakowski 1992

Gudmundsson, Kenney & Laub 1995:

$$E\left(n\|Az\|_2^2\right) = \|A\|_F^2$$

Kenney, Laub & Reese 1998: $\|A^{-1}\|_F$

Generating Random Vectors

$$y = \left(y_1 \ \ \ldots \ \ y_n\right)^T$$

y_i are independent normal $(0, 1)$

$$\|y\|_2 = \left(\sum_{i=1}^n |y_i|^2\right)^{1/2}$$

$y/\|y\|_2$ uniformly distributed on unit n-sphere

Matlab: `y=randn(n,1); z=y/norm(y)`

Devroye 1986

Calafiore, Dabbene & Tempo 1999

Main Idea

$$z = \begin{pmatrix} z_1 & \ldots & z_n \end{pmatrix}^T$$

uniformly distributed on unit n-sphere

$$\text{Prob}\left(\sum_{i=1}^{k} z_i^2 \leq \epsilon^2\right) = \frac{B(a, b, \epsilon^2)}{B(a, b, 1)}$$

where $a = \frac{k}{2}$, $b = \frac{n-k}{2}$

$$B(a, b, x) = \int_0^x t^{a-1}(1-t)^{b-1} dt$$

Jessup & Ipsen 1992, Kenney & Laub 1994

Vector Norm Estimation

Kenney & Laub: $\|v\|_2 = E(|v^T z|)/E_n$

$$E_n = \begin{cases} \frac{(n-2)!!}{(n-1)!!} & \text{if } n \text{ odd} \\ \frac{2}{\pi}\frac{(n-2)!!}{(n-1)!!} & \text{if } n \text{ even} \end{cases}$$

$$\mathrm{Prob}\left(\epsilon\|v\|_2 \le \frac{|v^T z|}{E_n} \le \frac{\|v\|_2}{\epsilon}\right)$$

$$\ge 1 - \epsilon\left(\frac{2}{\pi} + 4e^{-1/(4\epsilon^2)}\right)$$

Our Bounds

$$P = \text{Prob}\left(|v^T z| \leq \epsilon \|v\|_2\right)$$

$$\epsilon_1 \sqrt{\frac{n-2}{n-3}}\,\text{erf}\left(\frac{\epsilon n_1}{\epsilon_1}\right) \leq P \leq \sqrt{\frac{n-1}{n-3}}\,\text{erf}\left(\epsilon n_1\right)$$

where

$$\text{erf}(x) = \tfrac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

$$n_1 = \sqrt{\tfrac{n-3}{2}},\ \epsilon_1 = \sqrt{1-\epsilon^2}$$

Probabilities for larger n

If $n \geq 1000$ then

$$\mathbf{Prob}\left(\sqrt{n}|v^T z| \leq 3\|v\|_2\right) \approx .99$$

$$\mathbf{Prob}\left(\sqrt{n}|v^T z| \geq \frac{\|v\|_2}{100}\right) \approx .92$$

With high probability

$$\sqrt{n}|v^T z| \approx \mu\|v\|_2, \qquad 10^{-2} \leq \mu \leq 3$$

Matrix Frobenius Norm Estimation

A is $n \times n$ real matrix

$$\|A\|_F = \left(\sum_{i,j=1}^{n} |a_{ij}|^2 \right)^{1/2}$$

Two approaches:

- Treat A as a vector: $\|A\|_F = \|\operatorname{vec}(A)\|_2$
- Approximate $\|A\|_F$ by matrix vector products

First Approach

$$\|A\|_F = \|\operatorname{vec}(A)\|_2$$

$\operatorname{vec}(A)$ is $1 \times n^2$

z is uniformly distributed on unit n^2-sphere

$$\text{Prob}\left(|\operatorname{vec}(A) z| \leq \epsilon \|A\|_F\right) = \frac{B(a, b, \epsilon^2)}{B(a, b, 1)}$$

where $a = \frac{1}{2}$, $b = \frac{n^2-1}{2}$

$$B(a, b, x) = \int_0^x t^{a-1}(1-t)^{b-1} dt$$

Probabilities for larger n

If $n \geq 1000$ then

$$\mathrm{Prob}\,(n|\,\mathrm{vec}(A)z| \leq 3\|A\|_F) \approx .99$$

$$\mathrm{Prob}\left(n|\,\mathrm{vec}(A)z| \geq \frac{\|A\|_F}{100}\right) \approx .92$$

With high probability

$$n|\,\mathrm{vec}(A)z| \approx \mu\|A\|_F, \qquad 10^{-2} \leq \mu \leq 3$$

Second Approach

Use matrix vector product $\|Az\|_2$
z is uniformly distributed on unit n-sphere

$$\text{Prob}\left(\|Az\|_2 \le \epsilon\|A\|_F\right) \ge \frac{B(a,b,\epsilon^2)}{B(a,b,1)}$$

$$\text{Prob}\left(\|Az\|_2 \ge \epsilon\|A\|_F\right) \ge 1 - \frac{B(a,b,\epsilon^2)}{B(a,b,1)}$$

where $a = \frac{1}{2}$, $b = \frac{n-1}{2}$

Probabilities for larger n

If $n \geq 1000$ then

$$\text{Prob}\left(\sqrt{n}\|Az\|_2 \leq 3\|A\|_F\right) \geq .99$$

$$\text{Prob}\left(\sqrt{n}\|Az\|_2 \geq \frac{\|A\|_F}{100}\right) \geq .92$$

With high probability

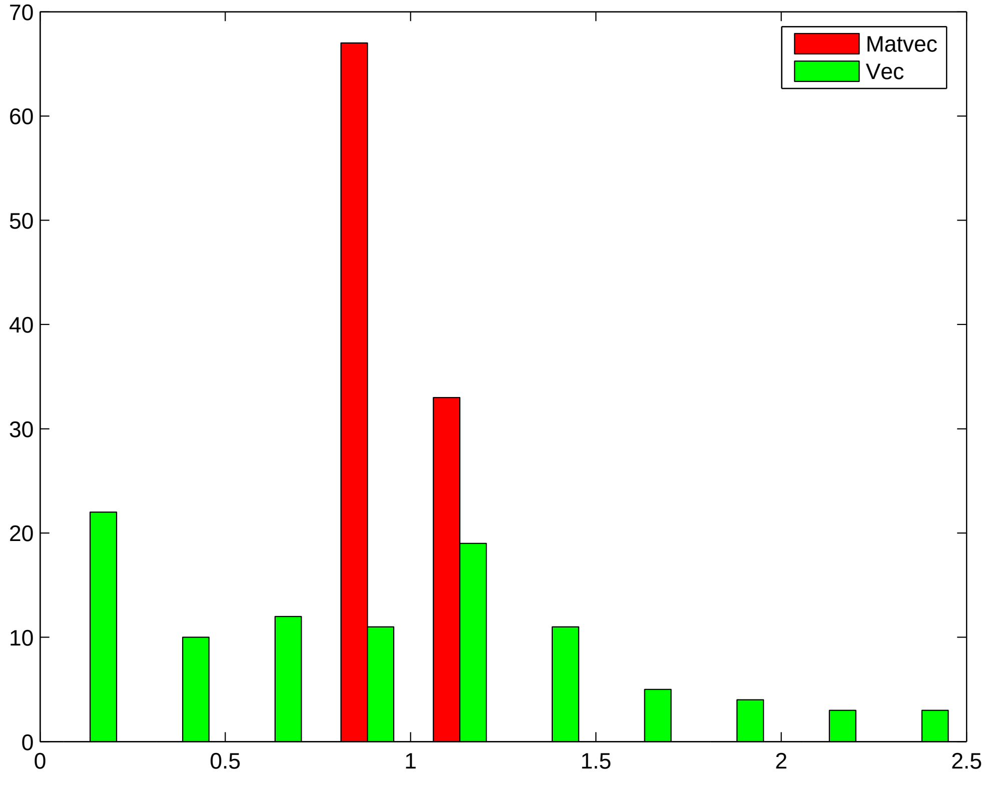
$$\sqrt{n}\|Az\|_2 \approx \mu\|A\|_F, \qquad 10^{-2} \leq \mu \leq 3$$

Experiments

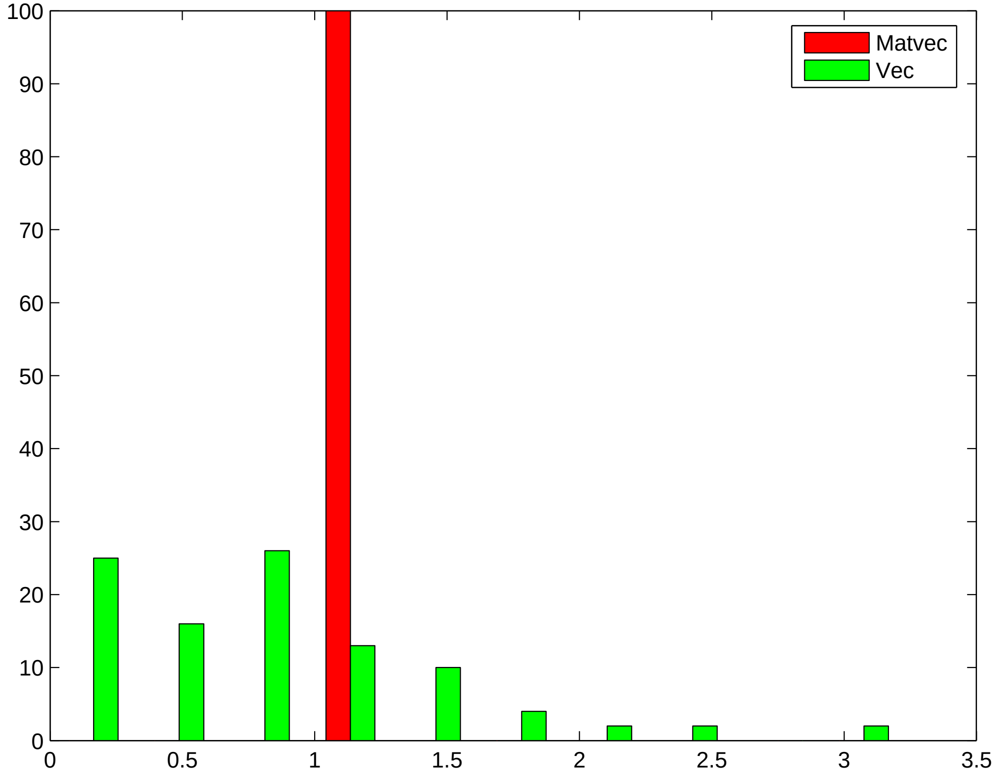
- Matrix dimension $n = 1000$
- Matvec:
  $\|A\|_F \approx \sqrt{n}\|Az\|_2$ where z is $n \times 1$
- Vec:
  $\|A\|_F \approx n|\,\mathrm{vec}(A)z|$ where z is $n^2 \times 1$
- 100 tries for each matrix
- Display:

$$\frac{\sqrt{n}\|Az\|_2}{\|A\|_F} \qquad \frac{n|\,\mathrm{vec}(A)z|}{\|A\|_F}$$

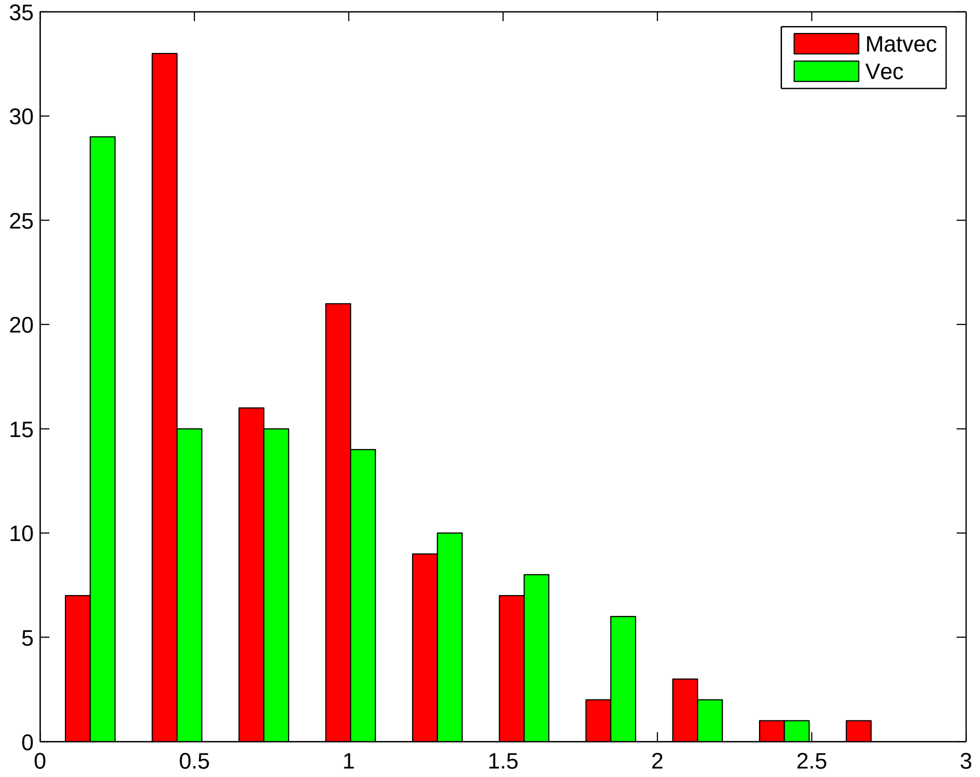
Grcar Matrix



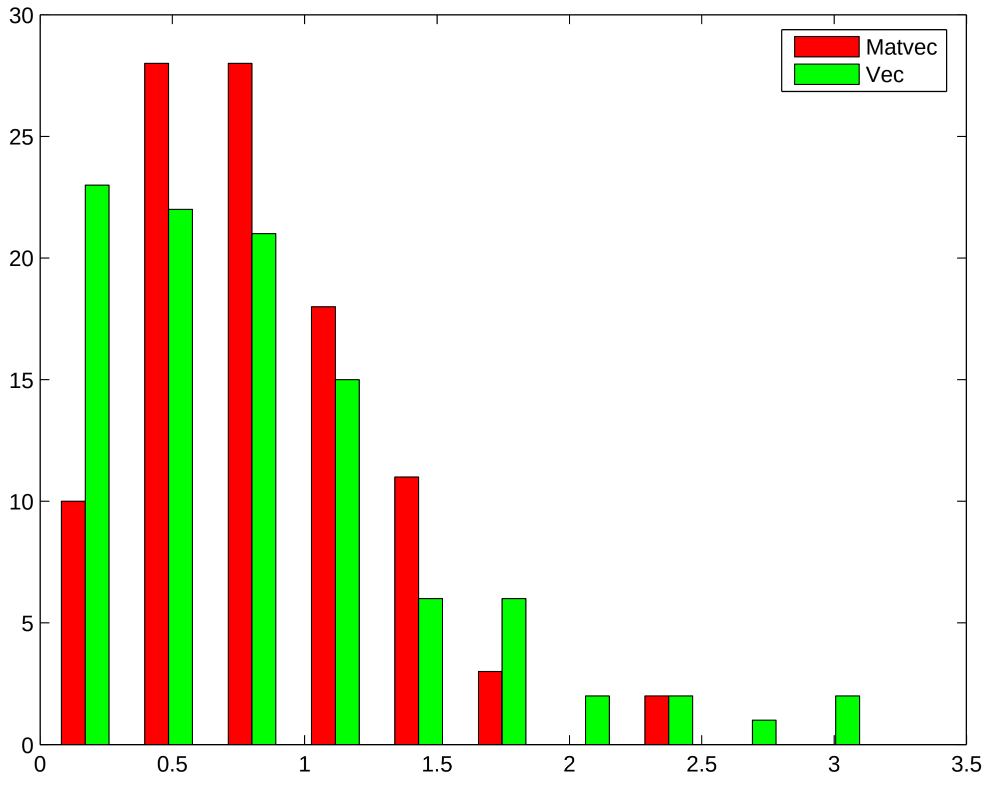
Orthogonal Matrix



Cauchy Matrix



Hilbert Matrix



Experiments

Matrix	Vec	Matvec
Grcar	$[.01, 2.3]$	$[.96, 1.04]$
Randn	$[.004, 3.2]$	$[.94, 1.1]$
Orthog	$[.005, 2.7]$	$[1, 1]$
Chebvand	$[.01, 3.1]$	$[.94, 1.06]$
Hilbert	$[.001, 2.1]$	$[.21, 2.2]$
Cauchy	$[.01, 2.9]$	$[.18, 2.3]$
Rand	$[.002, 2.9]$	$[.49, 1.9]$
Clement	$[.003, 2.7]$	$[.98, 1.02]$
Chebspec	$[.01, 2.5]$	$[.29, 2.5]$

Matvec vs. Vec

Experiments suggest:

- $\|A\|_F \approx \sqrt{n}\|Az\|_2$
- more accurate
- easier to compute (1 matvec)

- Vec: $\|A\|_F \approx n|\,\mathrm{vec}(A)z|$
- less accurate
- tends to underestimate more

Several Samples

A is a $n \times n$ real matrix

- z uniformly distributed on unit n-sphere

$$\|A\|_F \approx \sqrt{n}\|Az\|_F$$

- Z is $n \times m$ “random” matrix with orthonormal columns

$$\|A\|_F \approx \sqrt{\frac{n}{m}}\|AZ\|_F$$

“Projection” on a subspace

Orthogonal Random Matrices

Birkhoff & Gulati 1979
Stewart 1980
Anderson, Olkin & Underhill 1987

- X is $n \times m$, x_{ij} independent normal (0,1)
- Columns of X lin. indep. (with probability 1)
- QR decomposition

$$X = ZR, \quad Z^*Z = I_m, \quad R_{ii} > 0$$

- Z is invariantly distributed

Experiments

- Tridiagonal matrices of order $n = 1000$
- Matlab `randsvd`
  5 different singular value distributions
- Condition numbers: $\|A\|_2 \|A^{-1}\|_2 = 10^{16}$
- # samples: $m = 1, 2, 3$
- 100 tries for each matrix and sample
- Display:

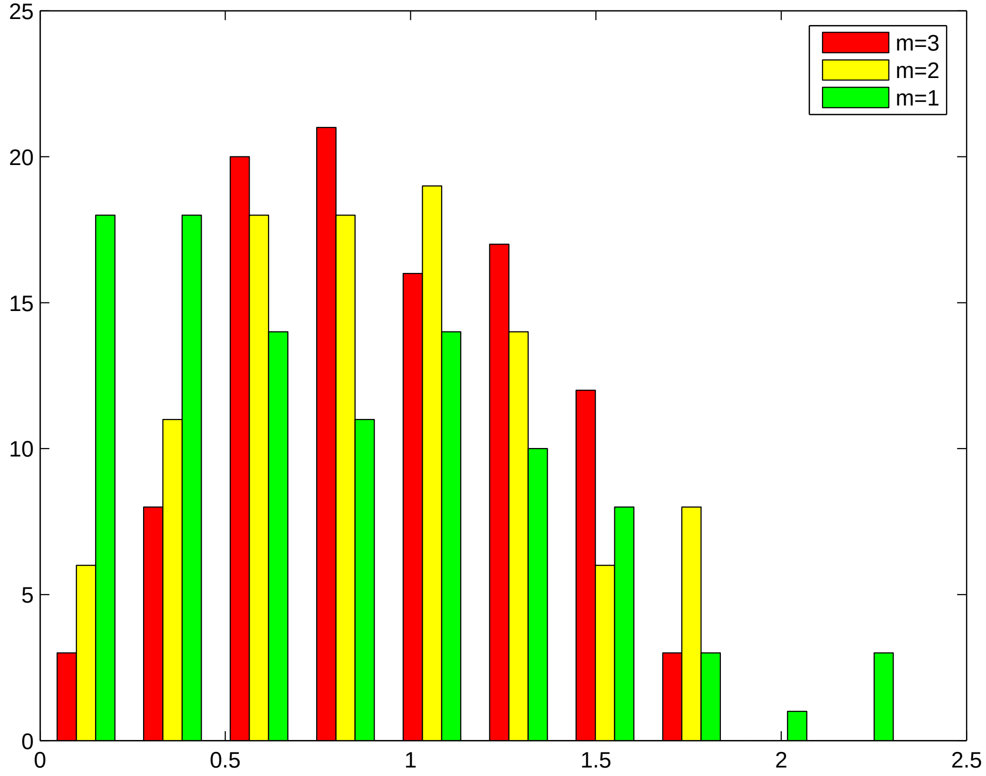
$$\sqrt{\frac{n}{m}} \frac{\|AZ\|_F}{\|A\|_F} \qquad m = 1, 2, 3$$

Singular Value Distributions

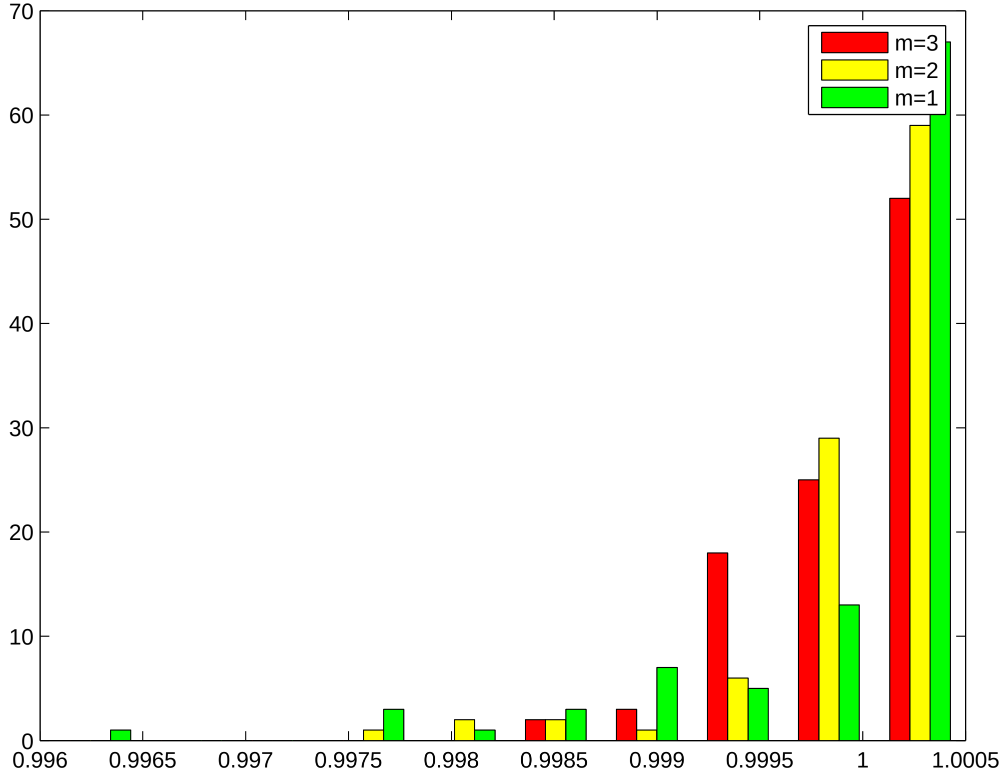
Singular values $\sigma_1 \geq \ldots \geq \sigma_n$
Condition number $\kappa = \sigma_1/\sigma_n$

- One large singular value:
  $\sigma_1 = 1,$ all other $\sigma_i = 1/\kappa$
- One small singular value:
  $\sigma_n = 1/\kappa,$ all other $\sigma_i = 1$
- Geometrically distributed singular values:
  $\sigma_i = \kappa^{-(i-1)/(n-1)}$

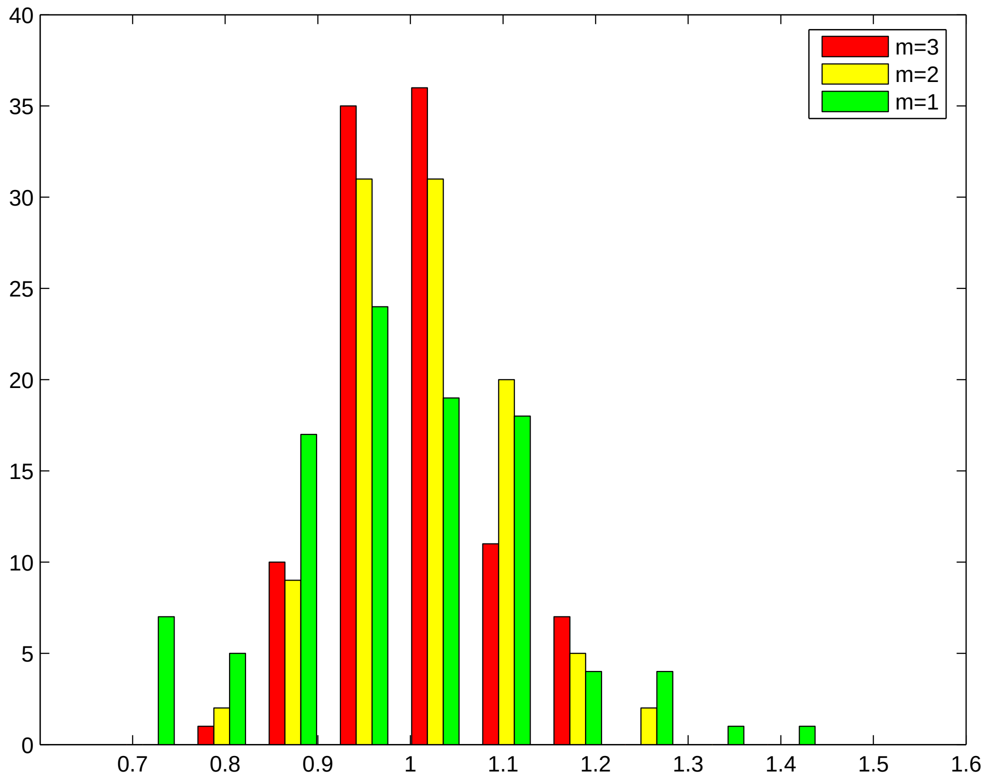
One Large Singular Value



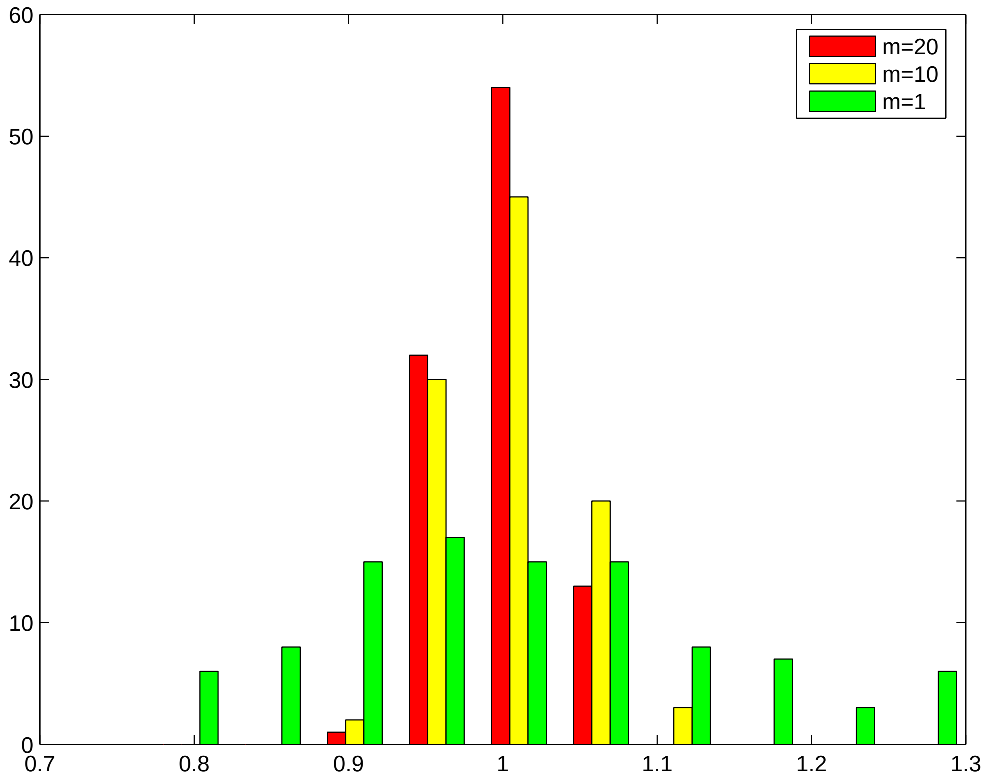
One Small Singular Value



Geometrically Distributed SValues



Many Samples: $m = 1, 10, 20$



Performance of $\sqrt{\frac{n}{m}}\|AZ\|_F$

Experiments suggest:

- A few samples ($m \leq 5$) do not help much
- Many samples ($m \geq 10$) can help more
- Sensitive to singular value distribution
- If A has many large singular values of similar magnitude then for $m = 1$

$$\sqrt{n}\|Az\|_2 \approx (1 \pm 10^{-2})\|A\|_F$$

Summary

- Statistical norm estimation
- $\|A\|_F \approx \sqrt{n}\|Az\|_2$
  A is $n \times n$
  z uniformly distributed on unit n-sphere
- With high probability for $n \geq 1000$

$$\sqrt{n}\|Az\|_2 \approx \mu\|A\|_F, \qquad 10^{-2} \leq \mu \leq 3$$

- Works well if A has many large singular values
- m samples: $\|A\|_F \approx \sqrt{\frac{n}{m}}\|AZ\|_F$
  Significant improvements only for large m