MANCHESTER
1824

M\cr NA
Manchester Numerical Analysis

The University of Manchester

Random Orthogonal Matrices in High Performance Computing

Nick Higham
Department of Mathematics
The University of Manchester
http://www.maths.manchester.ac.uk/~higham

Slides available at https://bit.ly/rom-hpc

KAUST Extreme Computing Research Center
July 19, 2020

What Is a Random Orthogonal Matrix?

Distributed according to the Haar measure over the group of orthogonal matrices.

- Haar measure provides a uniform distribution over the orthogonal matrices.
- Haar measure is invariant under mult on left and right by orthogonal matrices: if Q is distributed so is UQV for any orthogonal (possibly non-random) U and V.

What Is a Random Orthogonal Matrix?

Distributed according to the Haar measure over the group of orthogonal matrices.

- Haar measure provides a uniform distribution over the orthogonal matrices.
- Haar measure is invariant under mult on left and right by orthogonal matrices: if Q is distributed so is UQV for any orthogonal (possibly non-random) U and V.

These are *not* Haar distributed, where $S = -S^T$:

- random Householder matrix,
- random Cayley transform $(I + S)(I - S)^{-1}$,
- e^S.

Randsvd Matrices

$$A = P\Sigma Q^T \in \mathbb{R}^{m\times n}, \quad P, Q \text{ random orthogonal}$$
$$\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \dots, \sigma_n), \quad \sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_n \ge 0.$$

- **Demmel & McKenney (1989)**: LAPACK's test matrix generation suite.
- **H (1991, 1995)**: MATLAB function `randsvd`. Subsequently incorporated into `gallery('randsvd',...)`.
- **H & Zhang (2016)**: Matrix Depot package for Julia.

Generating Q

Method 1 (inefficient).

```
[Q,R] = qr(randn(n));
Q = Q*diag(sign(diag(R)));
```

Method 2 (efficient, product form, no R).
Stewart (1980): Let $x_k \in \mathbb{R}^{n-k+1}$ be normal (0,1).
H_k Householder matrix that reduces x_k to $r_{kk} e_1$.

$$Q = DH_1' H_2' \dots H_{n-1}',$$

where $H_k' = \text{diag}(I_{k-1}, H_k)$, $D = \text{diag}(\text{sign}(r_{kk}))$, $r_{nn} = x_n$.

- In MATLAB, `Q = gallery('qmult',n)`.
- Halves cost of forming a randsvd matrix: $\approx m^3 + n^3$ flops.

Cheaper Randsvd Matrix

Want to compute A in $O(mn)$ flops.

Cheaper Randsvd Matrix

Want to compute A in $O(mn)$ flops.

- In $A = P\Sigma Q^T$, give up P and Q Haar-distributed, or even random.
- Could use same construction with $k \le \min(m, n)$ Householders.

Cheaper Randsvd Matrix

- Let

$$A = Q\Sigma W^T,$$

where Q has orthonormal cols and W is a random (rectangular) Householder matrix.

- For $m = n$, if only $\kappa_2(A)$ is to be specified, can reduce the cost of formation by setting $\sigma_2 = \cdots = \sigma_{n-1} = 1$.

Properties

- Form in $O(mn)$ flops + cost of Q.
- Little communication required.

M. Fasi and N. J. Higham. **Generating extreme-scale matrices with specified singular values or condition numbers**. MIMS EPrint 2020.8, March 2020.

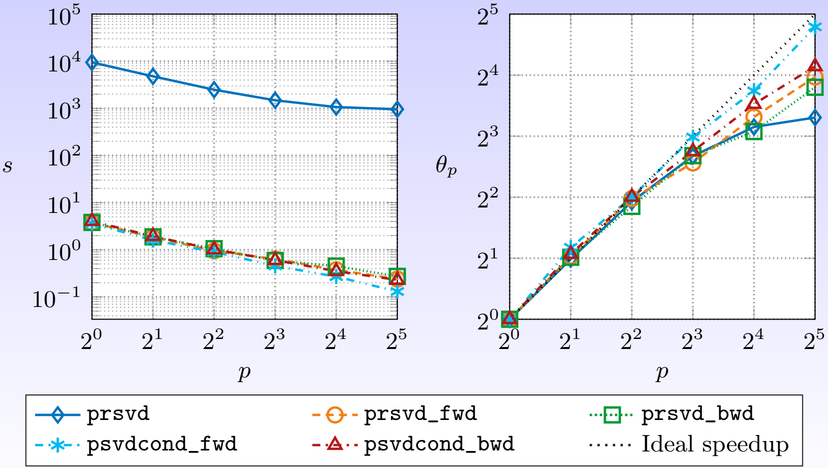
Choice of Q

- Haar distributed *or*
- to reduce communication, $Q = (f(i,j))$, such as

$$q_{ij} = \frac{2}{\sqrt{2n+1}} \sin\left(\frac{2ij\pi}{2n+1}\right).$$

Experiment, $n = 20,000$

- In C using BLACS, PBLAS, ScaLAPACK, Open MPI.
- Nodes have two 16-core Intel Xeon CPUs.
- p processes. Wall clock (left), speedup t_1/t_p (right).



The Growth Factor

Gaussian elimination on $A \in \mathbb{R}^{n \times n}$ produces $A = LU$.

With $A^{(1)} = A$, $A^{(n)} = U$, $A^{(k)} = \left(a_{ij}^{(k)}\right)$ matrix at kth stage of Gaussian elimination,

$$\rho_n(A) = \frac{\max_{i,j,k} |a_{ij}^{(k)}|}{\max_{i,j} |a_{ij}|} \geq 1.$$

The Growth Factor

Gaussian elimination on $A \in \mathbb{R}^{n\times n}$ produces $A = LU$.

With $A^{(1)} = A$, $A^{(n)} = U$, $A^{(k)} = \left(a_{ij}^{(k)}\right)$ matrix at kth stage of Gaussian elimination,

$$\rho_n(A) = \frac{\max_{i,j,k} |a_{ij}^{(k)}|}{\max_{i,j} |a_{ij}|} \geq 1.$$

Theorem (Wilkinson, 1961)

GE produces a computed solution $\widehat{x}$ to $Ax = b$ satisfying

$$(A + \Delta A)\widehat{x} = b, \qquad \|\Delta A\|_\infty \leq p(n)\rho_n u \|A\|_\infty,$$

where u is unit roundoff and p a low degree polynomial.

What We Know About the Growth Factor

- Without pivoting, ρ_n can arbitrarily large.
- With **partial pivoting**, $\rho_n \le 2^{n-1}$, attained for

$$A_4 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ -1 & 1 & 0 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 \end{bmatrix}.$$

- **Wright (1993)** and **Foster (1994)** found applications where partial pivoting suffers exponential growth.
- **Higham & Higham (1989)** found orthogonal matrices with $\rho_n \gtrsim n/2$ for *any* pivoting strategy.
- In practice, ρ_n is **almost always small** for partial pivoting. *Open problem to explain why!*

MATLAB Function

```
function g = gf(A)
%GF     Approximate growth factor.
%   g = GF(A) is an approximation to the
%   growth factor for LU factorization
%   with partial pivoting.
[~,U] = lu(A);
g = max(abs(U),[],'all')/max(abs(A),[],'all');
```

- This is a lower bound on $\rho_n(A)$.
- Can get exact growth factor using `gep.m` from **Matrix Computation Toolbox**.

```
>> rng(1); gf(randn(10))
ans =
   1.5088e+00
>> gf(randn(100))
ans =
   4.4874e+00
>> gf(randn(1000))
ans =
   1.5997e+01
>> gf(randn(10000))
ans =
   5.0505e+01
>> gf(gallery('randsvd',1000,1e8,2,[],[],1))
ans =
   7.5329e+01
```

Does $O(n)$ Growth Matter?

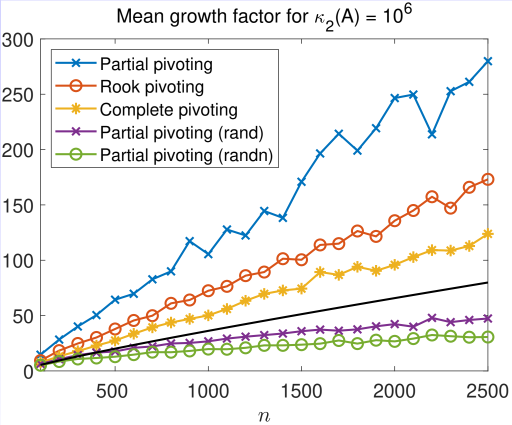
- $n = 10^7$ for today's largest dense $Ax = b$
$\Rightarrow$ problems in single precision.

- For IEEE half precision and $\max_{i,j} |a_{ij}| = 1$, linear growth can cause overflow for $n = 7 \times 10^4$.
(That's how these matrices were spotted.)

D. J. Higham, N. J. Higham, and S. Pranesh. **Random matrices generating large growth in LU factorization with pivoting**. MIMS EPrint 2020.13, May 2020.

Randsvd Matrices (Mode 2)

$$A = P\Sigma Q^T \in \mathbb{R}^{n\times n}, \quad P^TP = Q^TQ = I,$$
$$\Sigma = \mathrm{diag}(1,\dots,1,\sigma_n), \quad 1 \ge \sigma_n \ge 0.$$



Growth Factor Lower Bound

Theorem (H & H, 1989)

Let $A \in \mathbb{C}^{n\times n}$ be nonsingular,

$$\alpha = \max_{i,j} |a_{ij}|, \quad \beta = \max_{i,j} \left|\left(A^{-1}\right)_{ij}\right|, \quad \theta = (\alpha\beta)^{-1}.$$

Then $\theta \le n$, and for any permutation matrices Π_r and Π_c such that $\Pi_r A \Pi_c$ has an LU factorization, the growth factor for GE without pivoting on $\Pi_r A \Pi_c$ satisfies

$$\rho_n(A) \ge \theta.$$

Growth for Random Orthogonal Matrices

Randsvd with $\sigma_n = 1$ gives $A = PQ^T$: **random orthogonal matrix from Haar distribution**.

Jiang (2005) shows that

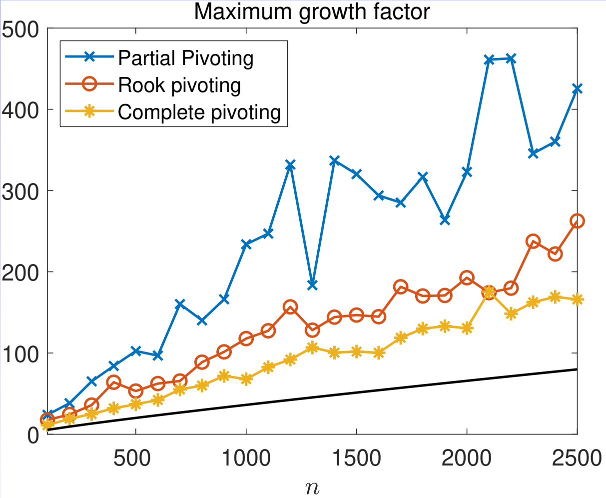
$$\Pr\left(\max_{i,j} |a_{ij}| > 2\sqrt{\frac{\log(n)}{n}}(1+\epsilon)\right) \to 0$$

as $n \to \infty$ for any $\epsilon > 0$. Since $A^{-1} = A^T$, can take $\alpha = \beta = 2\sqrt{\log(n)/n}$ in the theorem and conclude

$$\rho_n(A) \gtrsim \frac{n}{4 \log n}$$

for large n with high probability for *any* pivoting strategy.

Growth Factors for Random Orthogonal



Proof of Large Growth for Randsvd

- The randsvd matrix is

$$A = PQ^T + (\sigma_n - 1)p_n q_n^T,$$

  where p_n and q_n are the last columns of P and Q.

- If W is orthogonal and has large growth then a rank-1 perturbation of norm at most 1 *tends to preserve the large growth*.
- **Not particular** to W being Haar distributed.
- One approach is via **Sherman–Morrison formula**.

Direct Approach

Let W be orthogonal and

$$A = W + xy^T.$$

The U factor of W is given explicitly by

$$u_{ij} = \frac{\det\big(W(1\colon i, [1\colon i-1,\ j])\big)}{\det(W_{i-1})}, \qquad i \le j,$$

where $W_j = W(1\colon j, 1\colon j)$. Find $\widetilde{U}$ factor of A satisfies

$$\frac{\widetilde{u}_{ij}}{u_{ij}} = \frac{1 + y([1\colon i-1,\ j])^T\ W(1\colon i, [1\colon i-1,\ j])^{-1}\ x(1\colon i)}{1 + y(1\colon i-1)^T\ W_{i-1}^{-1}\ x(1\colon i-1)}.$$

Singular Values of Submatrix of W

Lemma

Let $W \in \mathbb{R}^{n\times n}$ be orthogonal and

$$\begin{array}{c c} & \begin{array}{cc} n-k & k \end{array} \\ \begin{array}{c} n-k \\ k \end{array} & \begin{bmatrix} W_{11} & W_{12} \\ W_{12} & W_{22} \end{bmatrix}, \end{array}$$

where $k < n/2$. Then W_{11} has at least $n-2k$ singular values equal to 1 and the remaining k singular values are bounded above by 1.

Singular Values of Submatrix of W

Lemma

Let $W \in \mathbb{R}^{n\times n}$ be orthogonal and

$$\begin{array}{c c} & \begin{array}{cc} n-k & k \end{array} \\ \begin{array}{c} n-k \\ k \end{array} & \begin{bmatrix} W_{11} & W_{12} \\ W_{12} & W_{22} \end{bmatrix}, \end{array}$$

where $k < n/2$. Then W_{11} has at least $n-2k$ singular values equal to 1 and the remaining k singular values are bounded above by 1.

Proof. Use the CS decomposition. □

Iterative Refinement

For $Ax = b$, with precisions **low**, **medium**, **high**.

- Factorize $A = LU$ in **low**.
- Solve $Ax = b$ in **low**.
- Repeat
  - $r = b - Ax$ in **high**.
  - Solve $Ad = r$ in **medium** using LU factors.
    or
  - Solve $U^{-1}L^{-1}Ad = U^{-1}L^{-1}r$ by GMRES in **medium**.
  - $x \leftarrow x + d$ in **medium**.

Large growth does not inhibit convergence of IR.

Summary

New class of **random**, dense $A \in \mathbb{R}^{n\times n}$ (*randsvd mode* 2) for which

- $\rho_n \gtrsim n/(4 \log n)$ for large n with **any form of pivoting**,
- $\kappa_2(A)$ can be **arbitrarily chosen**.

Have been part of MATLAB **gallery** for many years but their growth properties had not been recognized.

- New algorithm forms "randsvd-like" matrices at cost linear in # matrix elements, with little communication.
- Beware mode 2!

Slides available at `https://bit.ly/rom-hpc`

References I

J. W. Demmel and A. McKenney.
A test matrix generation suite.
Preprint MCS-P69-0389, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, USA, Mar. 1989.
16 pp.
LAPACK Working Note 9.

References II

- M. Fasi and N. J. Higham.
  Generating extreme-scale matrices with specified singular values or condition numbers.
  MIMS EPrint 2020.8, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, Mar. 2020.
  21 pp.

- L. V. Foster.
  Gaussian elimination with partial pivoting can fail in practice.
  SIAM J. Matrix Anal. Appl., 15(4):1354–1362, 1994.

References III

D. J. Higham, N. J. Higham, and S. Pranesh.
Random matrices generating large growth in LU factorization with pivoting.
MIMS EPrint 2020.13, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, May 2020.
15 pp.

N. J. Higham.
Algorithm 694: A collection of test matrices in MATLAB.
ACM Trans. Math. Software, 17(3):289–305, Sept. 1991.

References IV

- N. J. Higham.
  The Test Matrix Toolbox for MATLAB (version 3.0).
  Numerical Analysis Report No. 276, Manchester Centre for Computational Mathematics, UK, Sept. 1995.
  70 pp.
- N. J. Higham and D. J. Higham.
  Large growth factors in Gaussian elimination with pivoting.
  SIAM J. Matrix Anal. Appl., 10(2):155–164, Apr. 1989.
- T. Jiang.
  Maxima of entries of Haar distributed matrices.
  Prob. Theory Relat. Fields, 131(1):121–144, 2005.

References V

S. J. Wright.
A collection of problems for which Gaussian elimination with partial pivoting is unstable.
SIAM J. Sci. Statist. Comput., 14(1):231–238, 1993.

W. Zhang and N. J. Higham.
Matrix Depot: An extensible test matrix collection for Julia.
PeerJ Comput. Sci., 2:e58, Apr. 2016.