SIAM J. MATRIX ANAL. APPL.
Vol. 42, No. 1, pp. 185–201

© 2021 SIAM. Published by SIAM under the terms
of the Creative Commons 4.0 license

RANDOM MATRICES GENERATING LARGE GROWTH IN LU FACTORIZATION WITH PIVOTING*

DESMOND J. HIGHAM†, NICHOLAS J. HIGHAM‡, AND SRIKARA PRANESH‡

Abstract. We identify a class of random, dense, $n \times n$ matrices for which LU factorization with any form of pivoting produces a growth factor typically of size at least $n/(4 \log n)$ for large n. The condition number of the matrices can be arbitrarily chosen, and large growth also happens for the transpose. Previously, no matrices with all these properties were known. The matrices can be generated by the MATLAB function `gallery('randsvd',...)`, and they are formed as the product of two random orthogonal matrices from the Haar distribution with a diagonal matrix having only one diagonal entry different from 1, which lies between 0 and 1 (the "one small singular value" case). Our explanation for the large growth uses the fact that the maximum absolute value of any element of a Haar distributed orthogonal matrix tends to be relatively small for large n. We verify the behavior numerically and find that for partial pivoting the actual growth is significantly larger than the lower bound and much larger than the growth observed for random matrices with elements from the uniform $[0, 1]$ or standard normal distributions. We show more generally that a rank-1 perturbation to an orthogonal matrix producing large growth for any form of pivoting also generates large growth under reasonable assumptions. Finally, we demonstrate that GMRES-based iterative refinement can provide stable solutions to $Ax = b$ when large growth occurs in low precision LU factors, even when standard iterative refinement cannot.

Key words. LU factorization, Gaussian elimination, large growth factor, partial pivoting, rook pivoting, complete pivoting, random orthogonal matrix, Haar distribution, MATLAB, randsvd, GMRES-based iterative refinement

AMS subject classification. 65F05

DOI. 10.1137/20M1338149

1. Introduction.

The MATLAB code

```
rng(1), n = 750; kappa = 1e8; mode = 2;
A = gallery('randsvd',n,kappa,mode,[],[],1);
[L,U,P,~,growth] = gep(A,'p'); growth % Partial pivoting
```

produces the output

```
growth =
  103.7971
```

The code uses the function `gep` from the Matrix Computation Toolbox [18] to compute the growth factor for LU factorization with partial pivoting on a random $n \times n$ matrix A with $n = 750$. The growth factor is defined by

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

*Received by the editors May 14, 2020; accepted for publication (in revised form) November 17, 2020; published electronically February 2, 2021.
https://doi.org/10.1137/20M1338149
Funding: The work of the authors was supported by the Engineering and Physical Sciences Research Council grant EP/P020720/1, and by the Royal Society.

†School of Mathematics, University of Edinburgh, James Clerk Maxwell Building, Peter Guthrie Tait Road, Edinburgh, EH9 3FD, UK (d.j.higham@ed.ac.uk).
‡Department of Mathematics, University of Manchester, Manchester, M13 9PL, UK (nick.higham@manchester.ac.uk, srikara.pranesh@manchester.ac.uk).

Downloaded 06/18/21 to 192.41.114.224. Redistribution subject to CCBY license

© 2021 SIAM. Published by SIAM under the terms of the Creative Commons 4.0 license

where $a_{ij}^{(k)}$ $(k = 1\colon n)$ are the elements at the kth stage of the factorization [19, sect. 9.3], [39]. Growth of over 100 for a matrix of this size with partial pivoting is very unusual. Unusually large growth is also obtained for the same matrix with rook pivoting and complete pivoting:

```
>> [L,U,P,Q,growth] = gep(A,'r'); growth % Rook pivoting
growth =
   57.1362
>> [L,U,P,Q,growth] = gep(A,'c'); growth % Complete pivoting
growth =
   43.2643
```

(See [19, sect. 9.1], [34], [39] for details of all these pivoting strategies.) Large growth factors are undesirable because they are a warning that numerical instability is likely in the LU factorization, as originally shown by Wilkinson [39].

Several classes of matrices generating large growth factors for partial pivoting are known. Wilkinson [39, p. 327], [40, p. 212] showed that the $n \times n$ matrix of the form illustrated for $n = 4$ by

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

gives $\rho_n = 2^{n-1}$, which is the worst case for partial pivoting. Higham and Higham [20] give examples of practically occurring $n \times n$ matrices for which $\rho(A) \gtrsim n/2$ for any pivoting strategy; they are all orthogonal matrices or well conditioned diagonal scalings of orthogonal matrices. Wright [41] describes a class of two-point boundary value problems for which the multiple shooting method leads to a linear system on which partial pivoting suffers exponential growth. The matrix is block lower bidiagonal, except for a nonzero block in the top right-hand corner. Foster [10] shows that a quadrature method for solving a practically occurring Volterra integral equation gives rise to dense linear systems for which partial pivoting again gives growth factors exponential in the dimension. In all these examples the matrices are well conditioned.

The matrix in our example has 2-norm condition number $\kappa_2(A) = \sigma_1/\sigma_n = 10^8$ and a singular value decomposition (SVD) of the form

$$A = P\Sigma Q^T \in \mathbb{R}^{n\times n}, \quad P^TP = Q^TQ = I, \tag{1.1a}$$

$$\Sigma = \mathrm{diag}(1, \dots, 1, \sigma_n), \quad 1 \geq \sigma_n \geq 0. \tag{1.1b}$$

Here, $n-1$ of the singular values of A are equal to 1, and the last one is less than or equal to 1. The matrices P and Q are orthogonal matrices from the Haar distribution, that is, they are distributed according to the Haar measure, which is the unique measure on the orthogonal matrices that is invariant under multiplication on the left and right by orthogonal matrices [31]. A Haar distributed random orthogonal matrix can be obtained as the orthogonal QR factor of a matrix with elements from the normal (0,1) distribution, provided that the factorization is normalized so that the diagonal elements of R are nonnegative [3], [36].

Matrices of the form (1.1) are generated by a MATLAB function call of the form `gallery('randsvd',n,kappa,mode)` with `kappa` $= \sigma_n^{-1} \geq 1$ and `mode = 2` (the default value of `mode` is 3, which produces geometrically distributed singular values). Figure 1.1 shows the results of an experiment in which we generated matrices this way for dimensions $n = 100\colon 100\colon 2500$ and computed the growth factors for partial

Downloaded 06/18/21 to 192.41.114.224. Redistribution subject to CCBY license

© 2021 SIAM. Published by SIAM under the terms of the Creative Commons 4.0 license

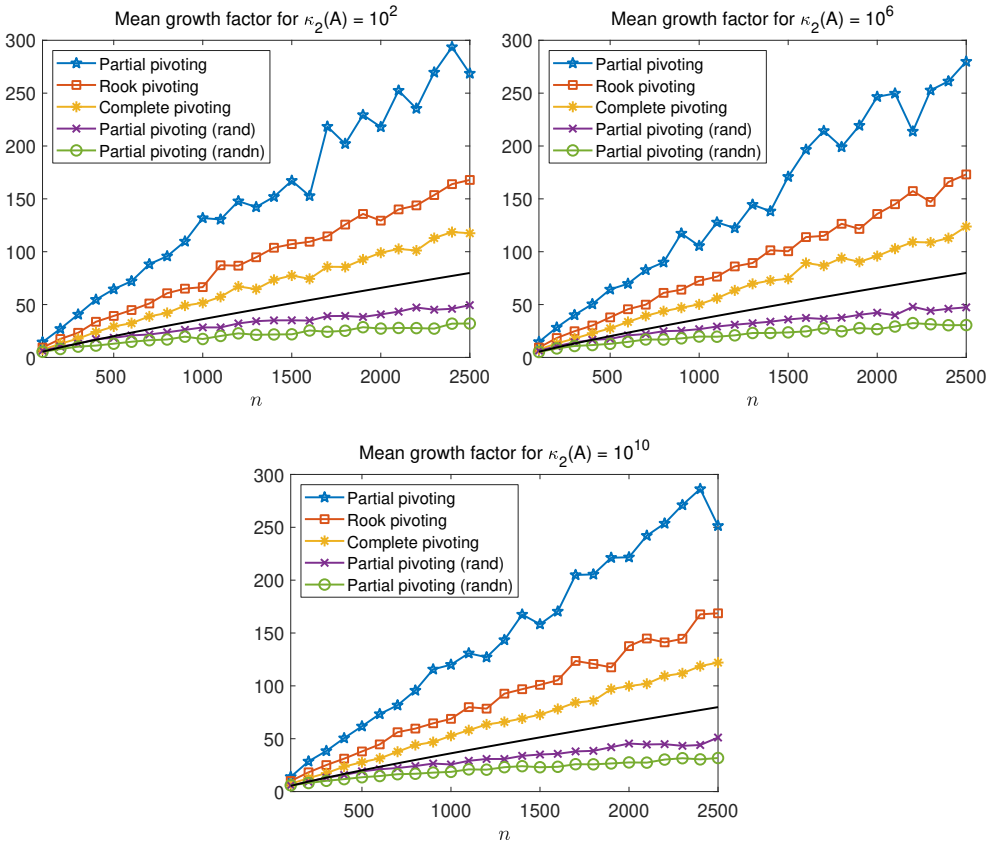


FIG. 1.1. *Mean growth factors for matrices* (1.1) *with* $\kappa(A) = 10^2, 10^6, 10^{10}$ *and for* `rand` *and* `randn` *matrices, with* 12 *samples for each* n. *The black curve is* $n/(4\log n)$.

pivoting, rook pivoting, and complete pivoting. For each dimension, we generated 12 matrices and took the mean growth factor. The figure illustrates the results for $\kappa_2(A) = 10^2, 10^6, 10^{10}$. As above, we used the `gep` function, which computes the exact growth factor (as opposed to the lower bound $\max_{i,j} |u_{ij}| / \max_{i,j} |a_{ij}|$ that must be used if we have access to the LU factors but not the intermediate quantities). We used the Parallel Computing Toolbox [33] to speed up the computations. We see that *irrespective of the condition number*, the growth factor increases with n at a rate roughly proportional to n for all three pivoting strategies. Experiments with other condition numbers confirm that the condition number has little effect on the growth factor. The largest growth factor observed in this experiment was 497. By contrast, for random matrices with elements from the uniform $[0, 1]$ distribution (`rand` in MATLAB) or the normal $(0, 1)$ distribution (`randn` in MATLAB) the figure shows that the growth factor for partial pivoting grows more slowly than linearly in n (as previously observed in [38]). The curves for the minimum and maximum growth factors are broadly similar to those for the means shown in Figure 1.1, and indeed every growth factor for the matrices (1.1) lies above the black curve, whose significance is explained in the following sections.

The significance of the matrices (1.1) is that they provide a new class of dense matrices A for which

Downloaded 06/18/21 to 192.41.114.224. Redistribution subject to CCBY license

© 2021 SIAM. Published by SIAM under the terms of the Creative Commons 4.0 license

- A generates large growth for any pivoting strategy,
- A^T also generates large growth for any pivoting strategy, and
- $\kappa_2(A)$ is arbitrary and easily assigned by choosing σ_n in (1.1).

The existing examples of large growth mentioned above are all well conditioned, some produce large growth only for partial pivoting, and not all of them produce large growth for A^T.

A growth factor of order αn for some constant $\alpha < 1$ with $\alpha > 1/10$ (say) may not seem to be a serious problem, given that the worst-case growth for partial pivoting is 2^{n-1}. But matrix dimensions in practical problems are increasing, with dense linear systems of order 10^7 being solved on today's largest machines [4]. The backward error bound for solution of a linear system $Ax = b$ by LU factorization is proportional to $\rho_n u$, where u is the unit roundoff [19, Thm. 9.5], so growth of order n can be problematic. Matters are exacerbated by the increasing use of low precision arithmetics such as IEEE half precision ($u \approx 5 \times 10^{-4}$) [24] and bfloat16 ($u \approx 4 \times 10^{-3}$) [25]. Low precision LU factorizations are being combined with iterative refinement to achieve faster solution times [14], [15], [16], [23], and the new HPL-AI benchmark uses this approach [8]. In low precision arithmetic large growth can even cause overflow. Indeed, we spotted the large growth factor for randsvd matrices with mode 2 because it led to overflow in LU factorization on these matrices in IEEE half precision arithmetic, for which the largest finite number is of order 6×10^4 [15].

In the next section we prove that for large n, large growth typically occurs for the matrices (1.1) with $\kappa_2(A) = 1$. This is the case of Haar distributed orthogonal matrices. In section 3 we show that if an orthogonal matrix generates large growth for any pivoting strategy then large growth persists after a random rank-1 perturbation, under reasonable assumptions. We specialize the results to a rank-1 perturbation of a Haar distributed orthogonal matrix, that is, matrices of the form (1.1) with an arbitrary $\kappa_2(A)$. In section 4 we provide an alternative analysis for the growth factor of a rank-1 perturbation of an orthogonal matrix based on the Sherman–Morrison formula. In section 5 we investigate the ability of mixed precision iterative refinement to overcome the instability in LU factorization caused by large growth factors.

2. Orthogonal matrices from the Haar distribution. We first consider the case where $\sigma_n = 1$ in (1.1), so that $A = PQ^T$ with P and Q orthogonal matrices from the Haar distribution. Since the Haar distribution is invariant under left or right multiplication by an orthogonal matrix, A is also Haar distributed, so we are effectively taking a single sample from the Haar distribution.

We need the following result from [20].

THEOREM 2.1. *Let $A \in \mathbb{R}^{n\times n}$ be nonsingular and set $\alpha = \max_{i,j} |a_{ij}|$, $\beta = \max_{i,j} \left|\left(A^{-1}\right)_{ij}\right|$, and $\theta = (\alpha\beta)^{-1}$. 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 LU factorization without pivoting on $\Pi_r A \Pi_c$ satisfies $\rho(A) \ge \theta$.*

Theorem 2.1 is used in [20] to show that for certain specific matrices that are orthogonal, or are well conditioned diagonal scalings of orthogonal matrices, the inequality $\rho_n(A) \gtrsim n/2$ holds for any pivoting strategy.

Donoho and Huo [9, Thm. VIII.1] show that for $n \times n$ matrices A drawn from the Haar distribution, $\Pr\big(\max_{i,j} |a_{ij}| > 2\sqrt{\log(n)/n}(1+\epsilon)\big) \to 0$ as $n \to \infty$ for any $\epsilon > 0$. Jiang [27, Prop. 1] proves the stronger result that $\sqrt{n/\log n}\max_{i,j} |a_{ij}|$ converges in probability to 2 as $n \to \infty$. We can say, then, that $\max_{i,j} |a_{ij}|$ is typically not larger than $2\sqrt{\log(n)/n}$ for large n, which we write as $\max_{i,j} |a_{ij}| \lesssim 2\sqrt{\log(n)/n}$ for large

Downloaded 06/18/21 to 192.41.114.224. Redistribution subject to CCBY license

© 2021 SIAM. Published by SIAM under the terms of the Creative Commons 4.0 license

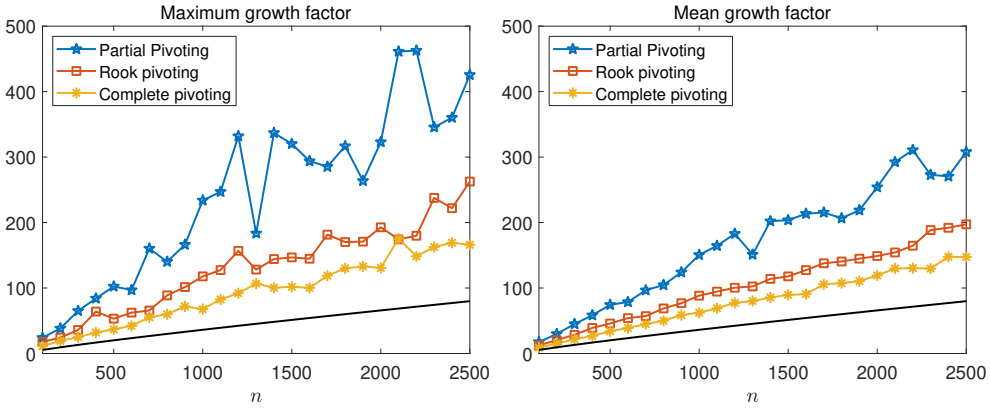


FIG. 2.1. *Growth factors for orthogonal matrices from the Haar distribution: maximum growth factor* (*left*) *and mean growth factor* (*right*) *over* 12 *samples for each* n. *The black curve is* $n/(4\log n)$.

n. Since $A^{-1} = A^T$, we can take $\alpha = \beta = 2\sqrt{\log(n)/n}$ in Theorem 2.1 and conclude that

$$\rho_n(A) \gtrsim \frac{n}{4\log n} \tag{2.1}$$

for large n for *any* pivoting strategy.

The lower bound in (2.1) is not as large as those for the orthogonal matrices and well conditioned diagonal scalings of orthogonal matrices in [20], but those matrices are nonrandom. Orthogonal matrices from the Haar distribution are the first class of random orthogonal matrices to be shown to give large growth.

Figure 2.1 shows the results of an experiment in which we generated Haar distributed orthogonal matrices of dimensions $n = 100\colon 100\colon 2500$ and computed the growth factors for partial pivoting, rook pivoting, and complete pivoting. For each dimension we generated 12 matrices and we show the maximum and average growth factors. All the growth factors in this experiment exceed $n/(4\log n)$ by a factor of more than 5, and they increase with n a little more rapidly than this approximate lower bound. As expected, the growth factor for partial pivoting exceeds that for rook pivoting, which in turn exceeds that for complete pivoting.

3. Rank-1 perturbations of orthogonal matrices. Matrices A of the form (1.1) can be written as

$$A = PQ^T + (\sigma_n - 1)p_n q_n^T, \tag{3.1}$$

where P and Q are orthogonal matrices from the Haar distribution and p_n and q_n are the last columns of P and Q, respectively. So A is a rank-1 perturbation of PQ^T, which is a Haar distributed orthogonal matrix and hence tends to give large growth. Preservation of large growth under rank-1 perturbations is not limited to Haar distributed orthogonal matrices or to the special form of the vectors making up the rank-1 perturbation in (3.1), as we now show with an experiment.

We consider four different nonrandom orthogonal matrices W identified in [20] that have a maximum element of at most $(2/n)^{1/2}$ and produce growth factors of at least $n/2$ for any pivoting strategy. Specifically, we take for W the MATLAB

Downloaded 06/18/21 to 192.41.114.224. Redistribution subject to CCBY license

© 2021 SIAM. Published by SIAM under the terms of the Creative Commons 4.0 license

TABLE 3.1
Growth factors for LU factorization with partial pivoting for four orthogonal matrices $W =$ `gallery('orthog',1000,j)` *and 20 rank-1 perturbations* $A = W + xy^T$ *with random* x *and* y *sampled from the uniform* $(0,1)$ *or normal* $(0,1)$ *distributions and scaled so that* $\|x\|_2 = \|y\|_2 = 1$.

		Random uniform			Random normal		
j	$\rho_n(W)$	$\min \rho_n(A)$	$\text{mean}\, \rho_n(A)$	$\max \rho_n(A)$	$\min \rho_n(A)$	$\text{mean}\, \rho_n(A)$	$\max \rho_n(A)$
1	5.24e+02	4.09e+02	4.79e+02	6.46e+02	4.70e+02	5.49e+02	6.80e+02
2	5.00e+02	3.85e+02	4.10e+02	4.55e+02	4.64e+02	4.71e+02	5.03e+02
5	5.63e+02	3.77e+02	4.96e+02	5.77e+02	4.70e+02	5.52e+02	8.81e+02
6	5.00e+02	3.68e+02	4.16e+02	4.68e+02	4.67e+02	4.74e+02	5.27e+02

matrices `gallery('orthog',n,j)` with $j = 1, 2, 5, 6$ and dimension $n = 1000$. We use LU factorization with partial pivoting and show in Table 3.1 the growth factor for W and the minimum, mean, and maximum growth factors for $A = W + xy^T$ with 20 vectors x and y having elements sampled from the uniform $(0,1)$ or normal $(0,1)$ distributions and then scaled to have unit 2-norm. The results show at worst a minor attenuation of the growth factor, with at least one perturbation giving an increased growth factor for every matrix type. Another notable feature of this experiment is that the largest element in magnitude of the upper triangular U factor was in the (n,n) position in every case for both W and A.

3.1. Analysis for general orthogonal matrices. We wish to analyze how element growth changes under a rank-1 update. We will initially derive a formula that holds for a rank-1 update

$$A = B + xy^T \tag{3.2}$$

of a general nonsingular matrix $B \in \mathbb{R}^{n\times n}$, with $x, y \in \mathbb{R}^n$. We assume that B has an LU factorization. We know that the U factor in the LU factorization of $B \in \mathbb{R}^{n\times n}$ is given explicitly by [19, sect. 9.2][1]

$$u_{ij} = \frac{\det\big(B(1\colon i, [1\colon i-1,\ j])\big)}{\det(B_{i-1})}, \quad 1 \le i \le j \le n, \tag{3.3}$$

where $B_j = B(1\colon j, 1\colon j)$. Suppose A in (3.2) has the LU factorization $A = \widetilde{L}\widetilde{U}$. It is easy to show that

$$\det(A) = \det(B)(1 + y^T B^{-1} x). \tag{3.4}$$

Now

$$A(1\colon i, [1\colon i-1,\ j]) = B(1\colon i, [1\colon i-1,\ j]) + x(1\colon i)y([1\colon i-1,\ j])^T,$$

and hence, analogously to (3.4), assuming $B(1\colon i, [1\colon i-1,\ j])$ is nonsingular,

$$\det(A(1\colon i, [1\colon i-1,\ j]) = \det\big(B(1\colon i, [1\colon i-1,\ j])\big) \times \big(1 + y([1\colon i-1,\ j])^T B(1\colon i, [1\colon i-1,\ j])^{-1} x(1\colon i)\big).$$

[1] We write $A(u,v)$, where u and v are vectors of subscripts, to denote the submatrix formed from the intersection of the rows indexed by u and the columns indexed by v.

Downloaded 06/18/21 to 192.41.114.224. Redistribution subject to CCBY license

© 2021 SIAM. Published by SIAM under the terms of the Creative Commons 4.0 license

Similarly,

$$\det(A_{i-1}) = \det\big(B_{i-1}\big)\big(1 + y(1\colon i-1)^T B_{i-1}^{-1}\, x(1\colon i-1)\big),$$

so using (3.3) with B replaced by A we have

$$\widetilde{u}_{ij} = \frac{\det\big(B(1\colon i, [1\colon i-1,\, j])\big)\big(1 + y([1\colon i-1,\, j])^T B(1\colon i, [1\colon i-1,\, j])^{-1} x(1\colon i)\big)}{\det\big(B_{i-1}\big)\big(1 + y(1\colon i-1)^T B_{i-1}^{-1}\, x(1\colon i-1)\big)}.$$

Combining this equation with (3.3) we obtain

$$\frac{\widetilde{u}_{ij}}{u_{ij}} = \frac{1 + y([1\colon i-1,\, j])^T B(1\colon i, [1\colon i-1,\, j])^{-1} x(1\colon i)}{1 + y(1\colon i-1)^T B_{i-1}^{-1}\, x(1\colon i-1)}. \tag{3.5}$$

Now we specialize to the situation of interest,

$$A = W + xy^T, \quad \|x\|_2 \le 1, \quad \|y\|_2 \le 1, \tag{3.6}$$

where W is an orthogonal matrix. Since we found in our experiments at the start of this section that the largest element of U in magnitude was always the (n, n) element, we take $i = j = n$. Now (3.5) becomes, using $W^{-1} = W^T$,

$$\frac{\widetilde{u}_{nn}}{u_{nn}} = \frac{1 + y^T W^T x}{1 + y(1\colon n-1)^T W_{n-1}^{-1}\, x(1\colon n-1)}. \tag{3.7}$$

We will argue that this ratio is likely to be of order 1, so that large growth for W manifested in a large $|u_{nn}|$ translates into a large $|\widetilde{u}_{nn}|$ and hence large growth for A. This is certainly not guaranteed; indeed, if $y = -W^T x$ then the numerator is zero, but such x, y, and W are specially correlated. Indeed, for (3.1) we have $y^T W^T x = \sigma_n - 1$, so for small σ_n, (3.7) need not be of order 1. We will adapt our analysis for this case in the next subsection.

To analyze the more typical behavior, we will assume that x is constructed as follows, and likewise for y: let $z_1, \ldots, z_n$ be independent random variables from the normal $(0, 1)$ distribution, and set $x = z/\|z\|_2$. Then x and y are uniformly distributed over the n-dimensional unit sphere [29]. We will assume that W is a fixed matrix and will bound the expectations of the numerator and denominator in (3.7).

We need the following lemmas, in which

$$\mu_n = \left(\frac{2}{\pi\big(n - \frac{1}{2}\big)}\right)^{1/2} + O(n^{-3/2}), \quad \mu_n < n^{-1/2}.$$

LEMMA 3.1 (Kenney and Laub [29, Thm. 2.1, Lem. 6.1]). *Let $z \in \mathbb{R}^n$ be uniformly distributed over the n-dimensional unit sphere and let $g \in \mathbb{R}^n$ be a constant vector. Then $\mathbb{E}(|z^T g|) = \mu_n \|g\|_2$.*

LEMMA 3.2. *Let $x, y \in \mathbb{R}^n$ be independent vectors uniformly distributed over the n-dimensional unit sphere and let $B \in \mathbb{R}^{n\times n}$ be a constant matrix. Then*

$$\mathbb{E}(|y^T B x|) \le \frac{\mu_n}{n^{1/2}} \|B\|_F. \tag{3.8}$$

Proof. By Lemma 3.1, since x and y are independent,

$$\mathbb{E}(|y^T B x|) = \mathbb{E}_{x,y}(|y^T B x|) = \mathbb{E}_x\left(\mathbb{E}_y(|y^T B x|)\right) = \mu_n\, \mathbb{E}_x\left(\|Bx\|_2\right).$$

Downloaded 06/18/21 to 192.41.114.224. Redistribution subject to CCBY license

© 2021 SIAM. Published by SIAM under the terms of the Creative Commons 4.0 license

As a special case of a result of Gudmundsson, Kenney, and Laub [13, Lem. 2.2] we have

$$\mathbb{E}(\|Bx\|_2^2) = \frac{\|B\|_F^2}{n}.$$

Hence, by Jensen's inequality [2, p. 80],

$$\mathbb{E}(\|Bx\|_2) \le \frac{\|B\|_F}{n^{1/2}}$$

and the result follows. ∎

First, we consider the numerator of (3.7). By Lemma 3.2, $\mathbb{E}(|y^T W^T x)|) \le \mu_n < n^{-1/2}$. Hence the expected value of the numerator of (3.7) is of order 1.

Now we turn to the denominator of (3.7). Using Lemma 3.2, we have

$$\begin{aligned}\mathbb{E}\big(|y(1\colon n-1)^T W_{n-1}^{-1} x(1\colon n-1)|\big) &= \mathbb{E}\big(\big|y^T \operatorname{diag}(W_{n-1}^{-1}, 0)x\big|\big)\\ &\le \frac{\mu_n}{n^{1/2}}\|W_{n-1}^{-1}\|_F.\end{aligned}$$

By the CS decomposition (see Theorem A.1), W_{n-1} has $n-2$ singular values 1 and one singular value $c \le 1$, and $|w_{nn}| = |c|$. Hence $\|W_{n-1}^{-1}\|_F^2 = n-2+c^{-2} = n-2+|w_{nn}|^{-2}$. From $W^T = W^{-1} = U^{-1}L^{-1}$ we have $|w_{nn}| = |u_{nn}|^{-1}$. Hence, using $\mu_n < n^{-1/2}$, we obtain

$$\begin{aligned}\mathbb{E}\big(|y(1\colon n-1)^T W_{n-1}^{-1} x(1\colon n-1)|\big)^2 &\le \frac{\mu_n^2}{n}(n-2+|u_{nn}|^2)\\ &< \frac{1}{n} + \frac{|u_{nn}|^2}{n^2}\\ &\le \frac{1}{n} + \frac{\rho_n^2 \max_{i,j}|w_{ij}|^2}{n^2}. \qquad (3.9)\end{aligned}$$

For the matrices in the example at the start of this section, this bound is approximately

$$\frac{1}{n} + \frac{(n^2/4)(2/n)}{n^2} = \frac{3}{2n}.$$

For Haar distributed orthogonal matrices the upper bound (3.9) is approximately

$$\frac{1}{n} + \frac{(n/4\log n)^2(4\log(n)/n)}{n^2} = \frac{1}{n} + \frac{1}{4n\log n}.$$

In both cases, these quantities are much less than 1 for large n, so the expected value of the denominator of (3.7) will be of order 1.

We have focused on the (n,n) element of U and argued that when x and y are uniformly distributed on the unit sphere and $|u_{nn}| = \max_{i,j}|u_{ij}|$ we can expect $|\widetilde{u}_{nn}|$ to be of a similar order of magnitude to $|u_{nn}|$. This is what we observed in the experiment at the start of this section, where the (n,n) element was always the largest for both U and $\widetilde{U}$. If large growth is not reflected in the (n,n) element one can consider a pair (i,j) in (3.5) for which $|u_{ij}|$ is large. As long as $i = n-k$ and $j \ge n-k$ with k small, the analysis generalizes. In particular, we can apply a similar argument to the $(n-k)\times(n-k)$ submatrices appearing in the numerator (after suitable column permutations) and denominator of (3.5). A case in point is the randsvd matrices, which we consider next.

Downloaded 06/18/21 to 192.41.114.224. Redistribution subject to CCBY license

© 2021 SIAM. Published by SIAM under the terms of the Creative Commons 4.0 license

This analysis is for LU factorization without pivoting. If the pivoting strategy produces an LU factorization $\Pi_1 A \Pi_2 = \widehat{L}\widehat{U}$, with Π_1 and Π_2 permutation matrices, then we can rewrite (3.2) as $\Pi_1 A \Pi_2 = \Pi_1 B \Pi_2 + (\Pi_1 x)(\Pi_2 y)^T$ and apply the analysis with $A \leftarrow \Pi_1 A \Pi_2$, $B \leftarrow \Pi_1 W \Pi_2$, $x \leftarrow \Pi_1 x$, and $y \leftarrow \Pi_2 y$. Since we are interested in orthogonal W for which large growth is obtained for any pivot sequence, our conclusions are unaffected.

3.2. Analysis for randsvd matrices. We now consider matrices A of the form (1.1) with P and Q from the Haar distribution, which we recall from (3.1) can be written as

$$A = PQ^T + (\sigma_n - 1) p_n q_n^T, \tag{3.10}$$

where p_n and q_n are the last columns of P and Q, respectively. Since $W = PQ^T$ is Haar distributed, it typically gives a large growth factor for large n, as shown in section 2. However, this large growth is not usually reflected in u_{nn} when σ_n is small. Indeed, A has just one nonunit singular value, σ_n, and $PA = LU$ implies $\pm\sigma_n = \det(A) = \det(U) = u_{11} \dots u_{nn}$; for $\sigma_n \ll 1$, the pivoting strategy will tend to produce a rank revealing factorization, which in this context means one with a well conditioned leading principal $(n-1) \times (n-1)$ submatrix and hence a small $|u_{nn}|$.

However, in the experiment with randsvd matrices reported in section 1, large growth was always observed in the $(n-1, n-1)$ element of U; indeed, the ratio $|u_{n-1,n-1}| / \max_{i,j} |u_{ij}|$ exceeded 0.5 and 0.1 in 82 percent and 98 percent of the cases, respectively. We therefore set $i = j = n-1$ in (3.5) to obtain

$$\frac{\widetilde{u}_{n-1,n-1}}{u_{n-1,n-1}} = \frac{1 + y(1\colon n-1)^T W_{n-1}^{-1} x(1\colon n-1)}{1 + y(1\colon n-2)^T W_{n-2}^{-1} x(1\colon n-2)}. \tag{3.11}$$

The numerator is the same as the denominator in (3.7), and the denominator has an analogous form. Therefore in (3.10) we take $x = (\sigma_n - 1) p_n$ and $y = q_n$. For large n, the vectors p_n and q_n have components that are approximately normally distributed random variables with mean 0 and standard deviation $n^{-1/2}$ [12], [27, Cor. 1], so they are approximately uniformly distributed on the unit sphere. The analysis of section 3.1 therefore gives insight into why the ratio (3.11) is typically of order 1 and hence why A inherits a large growth factor from PQ^T.

Experiments show that Haar distributed orthogonal matrices maintain large growth under a wider class of rank-1 perturbations than (3.10). Figure 3.1 plots growth factors for partial pivoting for $A = W + xy^T$, with W an orthogonal matrix from the Haar distribution and x and y generated with elements from the uniform distribution on $[0, 1]$ and then scaled so that $\|x\|_2 = \|y\|_2 = 1$. For each $n = 100\colon 100\colon 2500$ we generated 12 random A and took the mean growth factor. We see that the growth factors for A are very similar to those for W.

It is interesting to note that, unlike for (3.10) with small σ_n, $A = W + uv^T$ is very well conditioned when u and v are random unit 2-norm vectors with independent entries from the same distribution. Indeed, Benaych-Georges and Nadakuditi [1, sect. 3.2] show that, almost surely

$$\sigma_1(A) \to \frac{1+\sqrt{5}}{2}, \quad \sigma_n(A) \to \frac{-1+\sqrt{5}}{2} \quad \text{as } n \to \infty,$$

and we know that the other $n-2$ singular values remain at 1 (because the singular values are the square roots of the eigenvalues of $A^T A$, which is the identity plus a

Downloaded 06/18/21 to 192.41.114.224. Redistribution subject to CCBY license

© 2021 SIAM. Published by SIAM under the terms of the Creative Commons 4.0 license

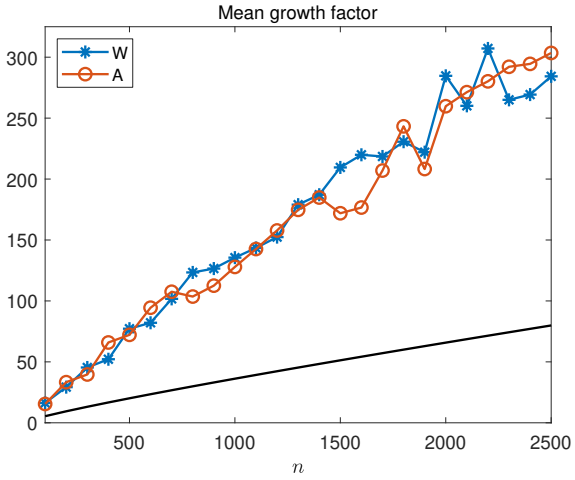


FIG. 3.1. *Mean growth factors for partial pivoting on orthogonal matrices W from the Haar distribution and on $A = W + xy^T$, where x and y are generated with elements from the uniform distribution on $[0, 1]$ and then scaled so that $\|x\|_2 = \|y\|_2 = 1$. The mean is over 12 matrices A and W for each n. The black curve is $n/(4 \log n)$.*

rank-2 matrix). Hence $\kappa_2(A) \approx (1 + \sqrt{5})/(-1 + \sqrt{5}) \approx 2.8$ for large n. In the experiment just mentioned the values of $\kappa_2(A)$ were all on the interval $[2.49, 2.93]$.

We mention some further matrices that generate large growth and are related to those we have considered. Matrices of the form (1.1a) with arithmetically distributed singular values are found to produce large growth in [15], though these are not low rank updates of orthogonal matrices. A referee reported experimental evidence that matrices of the form (1.1a) whose first $n - 1$ singular values are exponentially distributed on $[0, 1/2]$ with $\sigma_n \ll 1/2$ give large growth. We have also observed that large growth is preserved under rank-k perturbations of Haar distributed orthogonal matrices for $k \geq 1$, with the growth factor decreasing as k increases.

4. Analysis via the Sherman–Morrison formula. In section 3 we used the explicit characterization (3.3) of U in order to study growth factors for rank-1 perturbations xy^T of orthogonal matrices, focusing on the case where $\|x\|_2 \leq 1$ and $\|y\|_2 \leq 1$. In this section we look at rank-1 perturbations of orthogonal matrices from a different perspective, applying the Sherman–Morrison formula and then making use of the indirect bound from Theorem 2.1. We will show that growth of order $n/(4 \log(n))$ typically arises for large n for any rank-1 perturbation xy^T of a Haar distributed orthogonal matrix whenever the vectors x and y have 1-norm bounded by 1 and have elements of roughly uniform magnitude. We need the following general result.

THEOREM 4.1. *Let*

$$A = W + txy^T, \tag{4.1}$$

where $W \in \mathbb{R}^{n \times n}$ is orthogonal, $t \in [0, 1]$, and $x, y \in \mathbb{R}^n$ satisfy $\|x\|_1 \leq 1$ and $\|y\|_1 \leq 1$. Let

$$\alpha_w = \max_{i,j} |w_{ij}|,$$

and suppose that $\alpha_w < 1$. Then A is nonsingular and, for any pivoting strategy

Downloaded 06/18/21 to 192.41.114.224. Redistribution subject to CCBY license

© 2021 SIAM. Published by SIAM under the terms of the Creative Commons 4.0 license

producing an LU factorization for A, the growth factor satisfies

$$\rho_n(A) \geq \frac{1 - t\alpha_w}{\alpha_w \left(\alpha_w + t\|x\|_\infty \|y\|_\infty\right)}. \tag{4.2}$$

Proof. Since W is orthogonal, the Sherman–Morrison formula [17] gives

$$A^{-1} = W^T - \frac{tW^T x y^T W^T}{1 + t y^T W^T x}. \tag{4.3}$$

Using the Hölder inequality ($|f^T g| \leq \|f\|_\infty \|g\|_1$), we have

$$\max_k |W^T x|_k \leq \alpha_w, \quad \max_k |Wy|_k \leq \alpha_w.$$

Also, $|y^T W^T x| \leq \alpha_w < 1$, which confirms that the denominator in (4.3) is nonzero and hence that A is nonsingular, and indeed

$$\left| \frac{tW^T x y^T W^T}{1 + t y^T W^T x} \right|_{ij} \leq \frac{t\alpha_w^2}{1 - t\alpha_w}.$$

Hence, in (4.3),

$$\max_{i,j} |A^{-1}|_{ij} \leq \alpha_w + \frac{t\alpha_w^2}{1 - t\alpha_w} = \frac{\alpha_w}{1 - t\alpha_w}.$$

Using this bound in Theorem 2.1, along with $\max_{i,j} |a_{ij}| \leq \alpha_w + t\|x\|_\infty \|y\|_\infty$, we arrive at (4.2). ∎

In the case where W is an orthogonal matrix from the Haar distribution, we have $\alpha_w \lesssim 2\sqrt{\log(n)/n}$ for large n, as noted in section 2. In this case, Theorem 4.1 gives

$$\rho(A) \gtrsim \frac{1}{4\log(n)/n + 2t\sqrt{\log(n)/n}\,\|x\|_\infty \|y\|_\infty}. \tag{4.4}$$

So if

$$t\|x\|_\infty \|y\|_\infty = o(\sqrt{\log(n)/n}) \tag{4.5}$$

we obtain

$$\rho_n(A) \gtrsim \frac{n}{4\log n}, \tag{4.6}$$

which matches the bound (2.1) for the unperturbed case. Under the constraints $\|x\|_1 \leq 1$ and $\|y\|_1 \leq 1$, the additional requirement (4.5) will hold for any $0 \leq t \leq 1$ when the vectors x and y have elements of roughly equal magnitude, because then $\|x\|_\infty \approx \|y\|_\infty \approx 1/n$.

Now we consider two particular random choices of x and y.

If vectors u and v are constructed by drawing elements independently from the uniform [0,1] distribution then each element has mean 1/2, so $\|u\|_1 \approx n/2$, and likewise for v. Let $x = u/\|u\|_1$, $y = v/\|v\|_1$, and $t = 1$. Then $\|x\|_1 = \|y\|_1 = 1$ and $\|x\|_\infty \approx \|y\|_\infty \approx 2/n$, so (4.5) is satisfied and hence (4.6) holds.

Now let $x = u/\|u\|_1$ and $y = v/\|v\|_1$, where u and v are columns of Haar distributed orthogonal matrices. For large n, the vectors u and v have components that are approximately independent normally distributed random variables with mean 0 and

Downloaded 06/18/21 to 192.41.114.224. Redistribution subject to CCBY license

© 2021 SIAM. Published by SIAM under the terms of the Creative Commons 4.0 license

standard deviation $n^{-1/2}$ [27, Cor. 1], [28, Thm. 3]. Since the mean of the absolute value of a standard normal random variable is $(2/\pi)^{1/2}$ [30, eq. (3)], the 1-norms of u and v have means approximately $(2/\pi)^{1/2}n$. Moreover, the ∞-norm of a random vector $z \in \mathbb{R}^n$ with independent standard normal components has mean and variance bounded above by terms of order $\sqrt{\log n}$ and $\log n$, respectively [7, Appendix A]; an application of the Chebyshev inequality [2, p. 80] then allows us to bound $\|z\|_\infty$ by order $\sqrt{n}\log n$ with high probability. Identifying u and v with $z/\sqrt{n}$, we find that x and y both have ∞-norms bounded above by order $\log n/n$ with high probability whence, with $t = 1$, (4.5) and (4.6) follow.

Theorem 4.1 also shows that existing growth factor bounds obtained for orthogonal matrices, such as those in [20], are essentially unchanged under appropriate rank-1 perturbations.

We note that Theorem 4.1 constrains the 1-norms of x and y. Since the 1-norm is generally larger than the 2-norm, this Sherman–Morrison-based analysis complements rather than replaces that in section 3.

5. Curing instability with mixed precision iterative refinement. When an LU factorization of A suffers large growth and we use the factorization to solve $Ax = b$, the solution usually (but not always [20]) has a correspondingly large backward error. Suppose A is one of the types of matrix identified in this paper that has an LU factorization with a large growth factor; how can we obtain a backward stable solution to $Ax = b$ using this factorization? The natural answer is to apply iterative refinement. Indeed, it has been known since the 1970s that iterative refinement can cure instability in LU factorization [26], [35].

A recent usage of iterative refinement is with the LU factorization computed at a lower precision than the working precision, with residuals possibly computed in extra precision, and with the refinement equation solved either by substitution using the LU factors (denoted LU-IR) or by GMRES using the LU factors as preconditioners (known as GMRES-IR). GMRES-IR was proposed by Carson and Higham in [5], [6], and the analysis therein (notably [6, Thm. 4.1]) implies that it can tolerate instability in the factorization provided that the convergence of GMRES is not hindered by a lower quality preconditioner. Element growth is likely to reduce the quality of the preconditioner, so it is of interest to test experimentally the effect of a large growth factor on the convergence of GMRES.

We present an experiment in which we used mode 2 `gallery('randsvd')` matrices (that is, matrices of the form (1.1)) of varying dimensions, and $\kappa_2(A) = 10^2$ and $\kappa_2(A) = 10^7$. The iterative refinement algorithms that we use are characterized by a triple of precisions (p_1, p_2, p_3), where p_1 is the precision at which the LU factorization is computed, p_2 is the working precision, and p_3 is the precision at which the residual is computed. We consider three precision combinations (H, S, D), (H, D, D), and (S, D, D), where H, S, and D denote half precision ($u \approx 4.88 \times 10^{-4}$), single precision ($u \approx 5.96 \times 10^{-8}$), and double precision ($u \approx 1.11 \times 10^{-16}$), respectively. Half precision computations are performed using the `chop` function[2] of Higham and Pranesh [22]. The right-hand side vector is generated using `randn`. Iterative refinement is terminated when

$$\frac{\|b - A\widehat{x}\|_\infty}{\|b\|_\infty + \|A\|_\infty \|\widehat{x}\|_\infty} \leq nu,$$

where the left-hand side is the relative backward error, and u is the unit roundoff

[2] https://github.com/higham/chop

Downloaded 06/18/21 to 192.41.114.224. Redistribution subject to CCBY license

© 2021 SIAM. Published by SIAM under the terms of the Creative Commons 4.0 license

TABLE 5.1

Total number of iterative refinement steps in standard iterative refinement (LU-IR) and in GMRES-IR for different precision combinations for $\kappa_2(A) = 10^2$. Numbers in parentheses denote the total number of GMRES iterations.

	(H, S, D)		(H, D, D)		(S, D, D)	
n	LU-IR	GMRES-IR	LU-IR	GMRES-IR	LU-IR	GMRES-IR
500	1	2 (2)	5	3 (6)	2	2 (2)
750	1	1 (1)	5	3 (6)	2	2 (2)
1000	1	1 (1)	6	3 (6)	2	2 (2)
1250	1	2 (2)	6	3 (6)	2	2 (2)
1500	1	1 (1)	5	3 (6)	2	2 (2)
1750	1	1 (2)	5	3 (6)	2	2 (2)
2000	1	1 (1)	6	3 (6)	2	2 (2)
2250	1	1 (2)	6	3 (7)	2	2 (2)
2500	1	1 (2)	6	2 (6)	2	2 (2)

TABLE 5.2

Growth factors for partial pivoting and condition number of left preconditioned matrix for $\kappa_2(A) = 10^2$.

		$\kappa_\infty(\widehat{U}^{-1}\widehat{L}^{-1}A)$		
n	ρ_n	(H, S, D)	(H, D, D)	(S, D, D)
500	44.61	6.44e+00	6.44e+00	1.00
750	92.63	8.43e+00	8.43e+00	1.00
1000	221.61	1.59e+01	1.59e+01	1.00
1250	125.77	2.13e+01	2.13e+01	1.00
1500	167.26	2.09e+01	2.09e+01	1.00
1750	349.38	3.31e+01	3.31e+01	1.00
2000	170.52	3.91e+01	3.91e+01	1.01
2250	256.11	5.41e+01	5.41e+01	1.01
2500	248.20	6.45e+01	6.45e+01	1.01

of the working precision. The inner GMRES iterations are terminated based on a backward error criterion for the preconditioned system with tolerances 10^{-2} and 10^{-4} for working precisions of single and double, respectively, and a maximum of 20 iterative refinement steps are performed. In practice, we hope for convergence in a handful of iterative refinement steps, but we allow more in order to explore the speed of convergence for different problems and the two methods.

Table 5.1 shows the convergence for $\kappa_2(A) = 10^2$ and Table 5.2 shows the growth factors and condition numbers. Tables 5.3 and 5.4 give the corresponding information for $\kappa_2(A) = 10^7$. We need $\kappa_\infty(A)u$ sufficiently less than 1 to guarantee convergence of LU-IR and $\kappa_\infty(\widehat{U}^{-1}\widehat{L}^{-1}A)u$ sufficiently less than 1 to guarantee convergence of GMRES-IR [6].

Both LU-IR and GMRES-IR successfully solve the problems with $\kappa_2(A) = 10^2$. For $\kappa_2(A) = 10^7$, LU-IR fails to converge in several instances, whereas GMRES-IR always converges within three iterative refinement steps, even though the condition guaranteeing convergence is not satisfied for (H, S, D). This behavior is consistent with the theory [6]. The important finding is that the inner GMRES solves converge in a modest number iterations, which shows that the large growth does not inhibit the ability of the computed low precision LU factors to act as effective preconditioners for GMRES.

We note that the convergence of the refinement could be enhanced by improving

Downloaded 06/18/21 to 192.41.114.224. Redistribution subject to CCBY license

© 2021 SIAM. Published by SIAM under the terms of the Creative Commons 4.0 license

TABLE 5.3
Total number of iterative refinement steps in standard iterative refinement (LU-IR) and in GMRES-IR for different precision combinations for $\kappa_2(A) = 10^7$. Numbers in parentheses denote the total number of GMRES iterations. "–" denotes that iterative refinement failed to converge.

	(H, S, D)		(H, D, D)		(S, D, D)	
n	LU-IR	GMRES-IR	LU-IR	GMRES-IR	LU-IR	GMRES-IR
500	–	2 (3)	–	3 (10)	12	3 (5)
750	4	1 (2)	–	3 (10)	–	3 (5)
1000	6	3 (7)	17	3 (13)	–	2 (4)
1250	16	2 (3)	–	4 (16)	16	3 (5)
1500	2	1 (2)	19	3 (12)	12	3 (5)
1750	2	1 (2)	–	3 (12)	19	3 (5)
2000	2	1 (2)	18	3 (12)	19	3 (5)
2250	3	1 (2)	–	2 (8)	–	3 (5)
2500	–	3 (9)	–	3 (13)	–	2 (4)

TABLE 5.4
Growth factors for partial pivoting and condition number of left preconditioned matrix for $\kappa_2(A) = 10^7$.

		$\kappa_\infty(\widehat{U}^{-1}\widehat{L}^{-1}A)$		
n	ρ_n	(H, S, D)	(H, D, D)	(S, D, D)
500	53.29	3.31e+10	3.30e+10	2.32e+03
750	103.80	3.60e+10	3.62e+10	3.33e+03
1000	90.27	8.96e+10	9.07e+10	4.84e+03
1250	102.03	1.58e+11	1.57e+11	1.72e+04
1500	178.48	1.24e+11	1.23e+11	1.51e+04
1750	186.22	2.10e+11	2.11e+11	3.31e+04
2000	321.61	2.49e+11	2.50e+11	1.48e+04
2250	349.27	3.85e+11	3.84e+11	3.28e+04
2500	188.25	3.95e+11	3.97e+11	1.34e+05

the preconditioner using a correction term based on an inexpensive estimate of the error in the factorization, as proposed by Higham and Mary [21].

6. Conclusions. The matrices (1.1) tend to produce growth factors in LU factorization of order $n/\log n$ for any pivoting strategy. Although these matrices are readily generated by the MATLAB `randsvd` function (albeit not with the default value of the `mode` parameter), this property appears to have gone unnoticed. The large growth stems from two properties. First, a random orthogonal matrix from the Haar distribution has relatively small elements with high probability for large n, which implies that the growth factor must be large for any pivoting strategy by a result from [20]. Second, if W is an orthogonal matrix that gives large growth for any pivoting strategy then a rank-1 perturbation of norm at most 1 to W tends to preserve large growth. We have given two explanations for this second property, one based on a determinantal formula for the elements of U and the other based on the Sherman–Morrison formula. The rank-1 perturbation allows the matrix to be given any 2-norm condition number, resulting in the class (1.1) of matrices with large growth and an arbitrary condition number.

With matrix dimensions in practical problems growing ever larger, and low precision arithmetic becoming increasingly prevalent, growth of order $n/\log n$ in LU factorization can render the solution to a linear system unstable. Fortunately, iter-

Downloaded 06/18/21 to 192.41.114.224. Redistribution subject to CCBY license

© 2021 SIAM. Published by SIAM under the terms of the Creative Commons 4.0 license

ative refinement is able to cure the instability, and we found that the performance of GMRES-IR, which uses the low precision computed LU factors as preconditioners for a GMRES-based solution to the correction equations, is unaffected by the lower quality computed LU factors.

Appendix A. The CS decomposition. We state here the CS decomposition with square diagonal blocks. For more details (including the most general form of the CS decomposition) see, e.g., Golub and Van Loan [11, p. 85], Paige and Wei [32], or Stewart and Sun [37, sect. 5.1].

THEOREM A.1. *Let $W \in \mathbb{R}^{n\times n}$ be orthogonal, and let $k \le n/2$. There exist orthogonal matrices $U_1, V_1 \in \mathbb{R}^{(n-k)\times(n-k)}$ and $U_2, V_2 \in \mathbb{R}^{k\times k}$ such that*

$$\begin{array}{c} \\ n-k \\ k \end{array}\!\!\begin{array}{c} \begin{array}{cc} n-k & k \end{array} \\ \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \end{array} = \begin{bmatrix} V_1 & 0 \\ 0 & U_2 \end{bmatrix} \left[\begin{array}{ccc} I_{n-2k} & 0 & 0 \\ 0 & C & S \\ \hline 0 & S & -C \end{array}\right] \begin{bmatrix} U_1 & 0 \\ 0 & V_2 \end{bmatrix}^T,$$

where $C = \mathrm{diag}(c_1, \ldots, c_k)$ and $S = \mathrm{diag}(s_1, \ldots, s_k)$ with $c_i \ge 0$, $s_i \ge 0$, and $c_i^2 + s_i^2 = 1$ for all i.

Acknowledgments. We thank Tim Davis, Cleve Moler, Rob Schreiber, and Nick Trefethen for their comments on a draft manuscript.

REFERENCES

[1] F. BENAYCH-GEORGES AND R. R. NADAKUDITI, *The singular values and vectors of low rank perturbations of large rectangular random matrices*, J. Multivariate Anal., 111 (2012), pp. 120–135, https://doi.org/10.1016/j.jmva.2012.04.019.
[2] P. BILLINGSLEY, *Probability and Measure*, 3rd ed., Wiley, New York, 1995.
[3] G. BIRKHOFF AND S. GULATI, *Isotropic distributions of test matrices*, Z. Angew. Math. Phys., 30 (1979), pp. 148–158, https://doi.org/10.1007/BF01601929.
[4] I. BUCK, *World's Fastest Supercomputer Triples Its Performance Record*, https://blogs.nvidia.com/blog/2019/06/17/hpc-ai-performance-record-summit/, June 17, 2019 (accessed June 24, 2019).
[5] E. CARSON AND N. J. HIGHAM, *A new analysis of iterative refinement and its application to accurate solution of ill-conditioned sparse linear systems*, SIAM J. Sci. Comput., 39 (2017), pp. A2834–A2856, https://doi.org/10.1137/17M1122918.
[6] E. CARSON AND N. J. HIGHAM, *Accelerating the solution of linear systems by iterative refinement in three precisions*, SIAM J. Sci. Comput., 40 (2018), pp. A817–A847, https://doi.org/10.1137/17M1140819.
[7] S. CHATTERJEE, *Superconcentration and Related Topics*, Springer-Verlag, Cham, Switzerland, 2014, https://doi.org/10.1007/978-3-319-03886-5.
[8] J. J. DONGARRA, P. LUSZCZEK, AND Y. M. TSAI, *HPL-AI Mixed-Precision Benchmark*, https://icl.bitbucket.io/hpl-ai/.
[9] D. L. DONOHO AND X. HUO, *Uncertainty principles and ideal atomic decomposition*, IEEE Trans. Inform. Theory, 47 (2001), pp. 2845–2862, https://doi.org/10.1109/18.959265.
[10] L. V. FOSTER, *Gaussian elimination with partial pivoting can fail in practice*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 1354–1362, https://doi.org/10.1137/S0895479892239755.
[11] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, 4th ed., Johns Hopkins University Press, Baltimore, MD, 2013.
[12] C. E. GONZÁLEZ-GUILLÉN, C. PALAZUELOS, AND I. VILLANUEVA, *Euclidean distance between Haar orthogonal and Gaussian matrices*, J. Theor. Probab., 31 (2018), pp. 93–118, https://doi.org/10.1007/s10959-016-0712-6.
[13] T. GUDMUNDSSON, C. S. KENNEY, AND A. J. LAUB, *Small-sample statistical estimates for matrix norms*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 776–792, https://doi.org/10.1137/S0895479893243876.

Downloaded 06/18/21 to 192.41.114.224. Redistribution subject to CCBY license

© 2021 SIAM. Published by SIAM under the terms of the Creative Commons 4.0 license

[14] A. Haidar, A. Abdelfattah, M. Zounon, P. Wu, S. Pranesh, S. Tomov, and J. Dongarra, *The design of fast and energy-efficient linear solvers: On the potential of half-precision arithmetic and iterative refinement techniques*, in Computational Science—ICCS 2018, Y. Shi, H. Fu, Y. Tian, V. V. Krzhizhanovskaya, M. H. Lees, J. Dongarra, and P. M. A. Sloot, eds., Springer International Publishing, Cham, 2018, pp. 586–600, https://doi.org/10.1007/978-3-319-93698-7_45.

[15] A. Haidar, H. Bayraktar, S. Tomov, J. Dongarra, and N. J. Higham, *Mixed-precision iterative refinement using tensor cores on GPUs to accelerate solution of linear systems*, Proc. Roy. Soc. London A, 476 (2020), 20200110, https://doi.org/10.1098/rspa.2020.0110.

[16] A. Haidar, S. Tomov, J. Dongarra, and N. J. Higham, *Harnessing GPU tensor cores for fast FP16 arithmetic to speed up mixed-precision iterative refinement solvers*, in Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis, SC18 (Dallas, TX), IEEE Press, Piscataway, NJ, 2018, pp. 603–613, https://doi.org/10.1109/SC.2018.00050.

[17] H. V. Henderson and S. R. Searle, *On deriving the inverse of a sum of matrices*, SIAM Rev., 23 (1981), pp. 53–60, https://doi.org/10.1137/1023004.

[18] N. J. Higham, *The Matrix Computation Toolbox*, http://www.maths.manchester.ac.uk/~higham/mctoolbox.

[19] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, 2002, https://doi.org/10.1137/1.9780898718027.

[20] N. J. Higham and D. J. Higham, *Large growth factors in Gaussian elimination with pivoting*, SIAM J. Matrix Anal. Appl., 10 (1989), pp. 155–164, https://doi.org/10.1137/0610012.

[21] N. J. Higham and T. Mary, *A new preconditioner that exploits low-rank approximations to factorization error*, SIAM J. Sci. Comput., 41 (2019), pp. A59–A82, https://doi.org/10.1137/18M1182802.

[22] N. J. Higham and S. Pranesh, *Simulating low precision floating-point arithmetic*, SIAM J. Sci. Comput., 41 (2019), pp. C585–C602, https://doi.org/10.1137/19M1251308.

[23] N. J. Higham, S. Pranesh, and M. Zounon, *Squeezing a matrix into half precision, with an application to solving linear systems*, SIAM J. Sci. Comput., 41 (2019), pp. A2536–A2551, https://doi.org/10.1137/18M1229511.

[24] *IEEE Standard for Floating-Point Arithmetic,* IEEE Std 754-2019 (Revision of IEEE 754-2008), IEEE, New York, 2019, https://doi.org/10.1109/IEEESTD.2019.8766229.

[25] Intel Corporation, *BFLOAT16—Hardware Numerics Definition*, white paper, document number 338302-001US, https://software.intel.com/en-us/download/bfloat16-hardware-numerics-definition, 2018.

[26] M. Jankowski and H. Woźniakowski, *Iterative refinement implies numerical stability*, BIT, 17 (1977), pp. 303–311, https://doi.org/10.1007/BF01932150.

[27] T. Jiang, *Maxima of entries of Haar distributed matrices*, Probab. Theory Relat. Fields, 131 (2005), pp. 121–144, https://doi.org/10.1007/s00440-004-0376-5.

[28] T. Jiang, *How many entries of a typical orthogonal matrix can be approximated by independent normals?*, Ann. Probab., 34 (2006), pp. 1497–1529, https://doi.org/10.1214/009117906000000205.

[29] C. S. Kenney and A. J. Laub, *Small-sample statistical condition estimates for general matrix functions*, SIAM J. Sci. Comput., 15 (1994), pp. 36–61, https://doi.org/10.1137/0915003.

[30] F. C. Leone, L. S. Nelson, and R. B. Nottingham, *The folded normal distribution*, Technometrics, 3 (1961), pp. 543–550, https://doi.org/10.1080/00401706.1961.10489974.

[31] F. Mezzadri, *How to generate random matrices from the classical compact groups*, Notices Amer. Math. Soc., 54 (2007), pp. 592–604, http://www.ams.org/notices/200705/fea-mezzadri-web.pdf.

[32] C. C. Paige and M. Wei, *History and generality of the CS decomposition*, Linear Algebra Appl., 208/209 (1994), pp. 303–326, https://doi.org/10.1016/0024-3795(94)90446-4.

[33] *Parallel Computing Toolbox*, The MathWorks, Inc., Natick, MA, http://www.mathworks.co.uk/products/parallel-computing/.

[34] G. Poole and L. Neal, *The rook's pivoting strategy*, J. Comput. Appl. Math., 123 (2000), pp. 353–369, https://doi.org/10.1016/S0377-0427(00)00406-4.

[35] R. D. Skeel, *Iterative refinement implies numerical stability for Gaussian elimination*, Math. Comp., 35 (1980), pp. 817–832, https://doi.org/10.1090/S0025-5718-1980-0572859-4.

[36] G. W. Stewart, *The efficient generation of random orthogonal matrices with an application to condition estimators*, SIAM J. Numer. Anal., 17 (1980), pp. 403–409, https://doi.org/10.1137/0717034.

[37] G. W. Stewart and J. Sun, *Matrix Perturbation Theory*, Academic Press, London, 1990.

Downloaded 06/18/21 to 192.41.114.224. Redistribution subject to CCBY license

© 2021 SIAM. Published by SIAM under the terms of the Creative Commons 4.0 license

[38] L. N. Trefethen and R. S. Schreiber, *Average-case stability of Gaussian elimination*, SIAM J. Matrix Anal. Appl., 11 (1990), pp. 335–360, https://doi.org/10.1137/0611023.
[39] J. H. Wilkinson, *Error analysis of direct methods of matrix inversion*, J. ACM, 8 (1961), pp. 281–330, https://doi.org/10.1145/321075.321076.
[40] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, Oxford, UK, 1965.
[41] S. J. Wright, *A collection of problems for which Gaussian elimination with partial pivoting is unstable*, SIAM J. Sci. Comput., 14 (1993), pp. 231–238, https://doi.org/10.1137/0914013.

Downloaded 06/18/21 to 192.41.114.224. Redistribution subject to CCBY license

© 2021 SIAM. Published by SIAM under the terms of the Creative Commons 4.0 license