An Approximate Equation for the Condition Numbers of Well-scaled Matrices

Chin-Chieh Chiang
Department of Math and Computer Science
University of Maryland Eastern Shore
cchiang@umes.edu

John P. Chandler
Computer Science Department
Oklahoma State University
jpc@cs.okstate.edu

Abstract

It is important for anyone solving a system of linear equations, including engineers, to know whether or not the linear system is ill-conditioned. If the system is extremely ill-conditioned, the solution, even if it is exact, will not be of much practical use because the solution will be highly sensitive to tiny changes in the problem. Little work has been done on the scaling of matrices in recent years. In this paper, we introduce the properties of our SCALGM algorithm and present an approximate equation for the Euclidean condition number κ_2 of any matrix $A \in \mathbf{R}^{m \times n}$ after the matrix has been rescaled using our SCALGM algorithm, in terms of the minimum angle between a column vector of A and the linear subspace spanned by the remaining columns of A. Numerical evidence is presented supporting this relationship.

Introduction

Matrix scaling, or equilibration, has been an important subject in the scientific computing on linear algebraic systems. The relative change in the solution can be as large as the product of the relative change in the problem and the "condition number" of the matrix of coefficients, as the latter has been defined by Wilkinson [21]. Related to this is the fact that the number of digits, base b, that are unavoidably lost in solving a linear system using finite precision arithmetic can be as large as the logarithm, base b, of the condition number. Thus, a condition number of 10^7 can mean that a solution using 32-bit floating point arithmetic (type REAL or float) has no accurate digits in it.

Unfortunately, poorly scaled linear systems will often appear to be ill-conditioned, using the usual definition of condition number, when they really are not (that is, when the sensitivity is actually low or moderate). Our work addresses the problem of measuring the "true condition" number of a matrix, independent of its scaling. The true condition number turns

out to be correlated with the minimum angle between a column vector of the matrix and the subspace defined by the n-1 other column vectors.

We apply our iterative algorithm, called SCALGM, which works on any given nonzero matrix $A \in \mathbf{R}^{m \times n}$ to produce the row and column scale factors in the form of diagonal matrices D and E, respectively, such that the scaled matrix DAE is well-scaled. More properties of the scaled matrix performed by SCALGM will be introduced later.

Let $A \in \mathbf{R}^{m \times n}$ and let S = SCALGM(A). For some column j of S, let $\theta_j(S)$ be the angle between the jth column of S and the linear subspace spanned by the remaining columns of S, and let $\theta_{\min}(S)$ be the minimum angle among them. We believe the minimum angle $\theta_{\min}(A)$ contains the essential information pertaining to the condition number $\kappa_2(A)$ of any well-scaled matrix A. If $\theta_{\min} = 90^{\circ}$, the matrix is optimally well-conditioned; as $\theta_{\min} \to 0^{\circ}$, the "true condition number" increases without limit. Below is an approximate equation we have found from our experiments for the relation between the condition numbers $\kappa_2(S)$ and the minimum angles $\theta_{\min}(S)$ for well-scaled matrices S obtained from our scaling method, namely the SCALGM algorithm:

$$\kappa_2(S) \approx 1.4\ (90/\ \theta_{\min}(S)) - 0.4\ ,$$

where $\theta_{\min}(S)$ is measured in degrees.

Matrix Scaling

The objective of scaling matrices A is to find suitable row and column scale factors, written as diagonal matrices, D and E, respectively, such that the scaled matrices DAE satisfy the desired properties. For instance:

- Forsythe and Moler [9] present the motivation of scaling to make pivoting work well.
- Berman, Parlett, and Plemmons [2] give a necessary and sufficient condition for A to be diagonally equivalent to an orthogonal matrix Q, and offer an algorithm that either produces positive diagonal matrices D and E such that DAE is orthogonal or it fails if no such pair D, E exists. (Note: Every orthogonal matrix Q is really orthonormal; the Euclidean norm of every column is unity, and hence is, in a sense, optimally well-scaled.)
- Bunch [4] presents an algorithm for any symmetric matrix A (with no null rows) such that the scaled matrix DAE is equilibrated in the ∞-norm.
- Curtis and Reid [7] propose an algorithm for scaling based on the assumption that the given matrix can be scaled into the required form such that all scaled matrix elements are of comparable size. Thus, the usual pivoting strategies for Gaussian elimination can be applied on the scaled matrix.
- Fulkerson and Wolfe [11] present a method for finding scale factors that minimize the ratio of the matrix entry of largest absolute value to that of smallest non-zero absolute

value. These authors state that it is believed that such a number is a useful condition number.

- Rothblum, Schneider, and Schneider [16] present an algorithm so that for a given nonnegative symmetric matrix A and a positive vector r, it either finds a positive diagonal matrix D such that $B = DAD$ has row maxima prescribed by r or shows that no such D exists.
- Parlett and Reinsch [15] present an algorithm based on the work of Osborne [14] on balancing a matrix for calculation of eigenvalues and eigenvectors.
- Skeel [18] shows the effect of scaling on the stability of Gaussian elimination.
- The problem of optimal scaling of matrices with respect to the condition number κ_2 has been extensively studied, as seen in papers presented by Bauer [1], Braatz and Morari [3], Businger [5], Forsythe and Strauss [10], Golub and Varah [12], McCarthy and Strang [13], Rump [17], and Watson [20].

We apply our iterative convergence algorithm, called SCALGM, below.

Basic "Scale up"" algorithm:
Divide each row (or column) in a matrix by the smallest non-zero magnitude of any element in that row (or column).

Basic "Scale down" algorithm:
Divide each row (or column) in a matrix by the largest magnitude of any element in that row (or column).

SCALGM algorithm:
Iterate until converged.
Call the matrix at the beginning of an iteration "the original matrix."
Step 1) Scale the original matrix up by rows, then by columns,
saving the scale factors.
Step 2) Scale the original matrix up by columns, then by rows,
saving the scale factors.
Step 3) Scale up the rows of the original matrix by the geometric mean
of the row factors from Step 1 and Step 2 above,
and scale up the columns of the original matrix analogously.
Steps 4–6) Scale down the resulting matrix similarly.
After the above iteration converges, iterate Steps 4–6 only,
until converged.

In practice, the matrix to be rescaled is not altered; all rescaling is done by modifying row and column scale factors. SCALGM works on any given non-zero matrix $A \in \mathbf{R}^{m \times n}$ to produce the row and column scale factors in the form of diagonal matrices D and E, respectively, such that the scaled matrix DAE satisfies the following properties:

1. The maximum magnitude of elements in DAE is 1.
2. The non-zero rows and columns of DAE are equilibrated in the ∞-norm.

3. The ratio of the minimum magnitude of nonzero elements to the maximum magnitude of elements in *DAE* is maximized.
4. The minimum magnitude *m* of non-zero elements in *DAE* must occur in a pair or more.
5. Algorithm SCALGM preserves symmetry.

Measuring Linear Dependence by Angles

Let $A = [a_1|\ldots|a_n] \in \mathbf{R}^{m\times n}$, where a_j is the jth column vector of A with

$$a_j = \begin{bmatrix} a_{1j} \\ \cdot \\ \cdot \\ \cdot \\ a_{mj} \end{bmatrix}.$$

Let $x = [x_1,\ldots,x_n]^T \in \mathbf{R}^{n\times 1}$. We may view $Ax \in range(A)$ as

$$\sum_{j=1}^{n} x_j a_j \in L(a_1, \ldots, a_n),$$

where $L(a_1, \ldots, a_n)$ is the linear span of the column vectors a_j of A, or the linear space spanned by the column vectors a_j of A. In fact, $range(A) = L(a_1, \ldots, a_n)$. In this section, we will investigate this problem: let $A = [a_1|\ldots|a_n] \in \mathbf{R}^{m\times n}$ be given with $m \geq n$, and A does not have to be of full rank.(If m < n, simply consider A^T instead of A.) From the column vectors of A, we compute the angle θ_j between a_j and $L(a_1,\ldots,a_{j-1},a_{j+1},\ldots,a_n)$, as illustrated in Figure 1 .

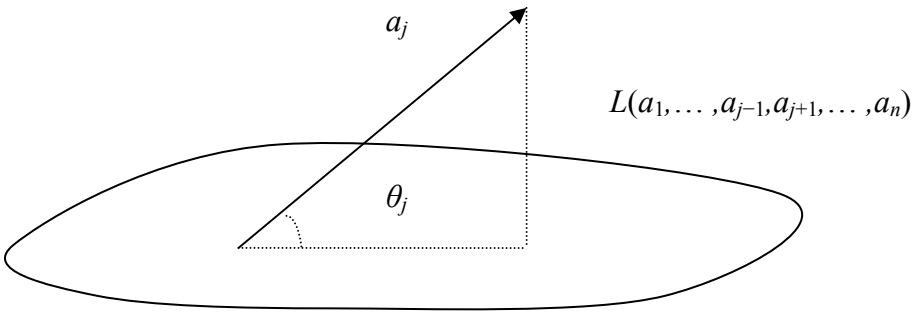

Figure 1: Computing the Angle θ_j

Note that $L(a_1,\ldots,a_{j-1},a_{j+1},\ldots,a_n) \in \mathbf{R}^{m\times(n-1)}$, denoted by $\hat{A}_j$, is a submatrix of A by removing the jth column vector from A, and thus

$$range(\hat{A}_j) = L(a_1,\ldots,a_{j-1},a_{j+1},\ldots,a_n).$$

Now, we are ready to find the orthogonal projector $P_j \in \mathbf{R}^{m\times m}$ that projects any vector onto $range(\hat{A}_j)$. In particular, for $a_j \in \mathbf{R}^m, 1 \le j \le n$, we have

$$(a_j - P_j a_j) \perp P_j a_j .$$

Let $\theta_j = \angle(a_j, P_j a_j)$. Then $0 \le \theta_j \le \pi/2$. Solve for each θ_j, and let $\theta_{\min} = \min\{\theta_j : 1 \le j \le n\}$, then $0 \le \theta_{\min} \le \pi/2$. Our approach is based on the following properties for determining the orthogonal projector P that projects a_j onto the linear subspace S_j, so as to find the angle θ_j by using the property of the inner product:

$$a_j^T (P_j a_j) = \|a_j\| \, \|P_j a_j\| \cos(\theta_j).$$

Property 1. If $A = [a_1|\ldots|a_n] \in \mathbf{R}^{m\times n}$ has full rank $n \le m$, then the orthogonal projector P that projects onto $range(A) = L(a_1, \ldots, a_n) \in \mathbf{R}^m$ is

$$P = AA^{\dagger} \in \mathbf{R}^{m\times m},$$

where $A^{\dagger} = (A^T A)^{-1} A^T \in \mathbf{R}^{m\times m}$ is the pseudoinverse of A.

Property 2. If $Q = [q_1|\ldots|q_n] \in \mathbf{R}^{m\times n}$ is an orthogonal matrix, then the orthogonal projector P that projects onto $range(Q) = L(q_1,\ldots,q_n) \in \mathbf{R}^m$ is

$$P = QQ^T .$$

It is well known that every matrix $A \in \mathbf{R}^{m\times n} (m \ge n)$, regardless of the rank of A, has a full QR decomposition $A = QR$, and, hence, a reduced QR decomposition $A = \hat{Q}\hat{R}$, where $Q \in \mathbf{R}^{m\times m}$, $\hat{Q} \in \mathbf{R}^{m\times n}$ are orthogonal matrices and $R \in \mathbf{R}^{m\times n}$, $\hat{R} \in \mathbf{R}^{n\times n}$ are upper-triangular matrices. Moreover, if A is of full rank = n, then A has a unique reduced QR decomposition $A = \hat{Q}\hat{R}$ with $r_{jj} > 0$. Therefore, we have the following property:

Property 3. If A is of full rank and $A = \hat{Q}\hat{R}$ is the reduced QR decomposition, then

(i) $range(A) = range(\hat{Q})$, and

(ii) the orthogonal projector P that projects onto $range(A)$ is $P = \hat{Q}\,\hat{Q}^T$.

What if the given matrix $A \in \mathbf{R}^{m\times n}$ is not of full rank? For example, $rank(A) = k < n \le m$. The reduced QR decomposition is still our choice for the rank-deficient matrix as to how to find an orthogonal matrix $\hat{Q} \in \mathbf{R}^{m\times k}$ and an upper-triangular matrix $\hat{R} \in \mathbf{R}^{k\times k}$ such that $A = \hat{Q}\hat{R}$ and thus

$$range(A) = range(\hat{Q}).$$

By Property 3, the orthogonal projector P that projects vectors onto $range(A)$ is

$$\hat{Q}\,\hat{Q}^T \in \mathbf{R}^{m\times m}.$$

The QR decomposition is important in numerical linear algebra and useful in its applications. Two well-known stable algorithms are Modified Gram Schmidt Orthogonalization (MGS) and Householder Triangularization.

The important property of the QR decomposition is that every matrix $A \in \mathbf{R}^{m\times n}$ with $m \ge n$, not necessarily of full rank, has the QR decomposition, written as

$$A = QR,$$

satisfying that $Q \in \mathbf{R}^{m\times m}$ is an orthogonal matrix and $R \in \mathbf{R}^{m\times n}$ is an uppertriangular matrix.

Example 1. Consider $A = [a_1|\ldots|a_n] \in \mathbf{R}^{n\times n}$:
If $\theta_{\min}(A) = 0$, then clearly, for some column vector a_j of A, $a_j \in range(\hat{A}_j)$, that is, a_j is a linear combination of other columns of A, and, hence, A is singular with $\kappa_2(A) = \infty$.
If $\theta_{\min}(A) = \pi/2$, then $a_i \perp a_j$ for all $i \ne j$. Thus, we have

$$A^TA = diag(\|a_1\|^2,\ldots,\|a_n\|^2),$$

and hence,

$$\kappa_2(A) = \frac{\max\{\|a_1\|,\ldots,\|a_n\|\}}{\min\{\|a_1\|,\ldots,\|a_n\|\}}.$$

Next, consider $A \in \mathbf{R}^{n\times n}$ and let $S = [s_1|\ldots|s_n] = \text{SCALGM}(A)$:
If $\theta_{\min}(S) = 0$, then $\kappa_2(S) = \infty$.
If $\theta_{\min}(S) = \pi/2$, then, by the property of SCALGM [Row and Column Equilibrated in ∞-norm], we have

$$1 \le \min\{\|s_1\|,\ldots,\|s_n\|\} \le \max\{\|s_1\|,\ldots,\|s_n\|\} \le \sqrt{n},$$

since $0 \le |s_{ij}| \le 1$, the upper bound $\sqrt{n}$ for 2-norm of columns s_j is attained if $s_j = [1,\dots,1]^T$, and the lower bound 1 for 2-norm of columns s_j is attained if s_j = some unit vector. Thus,

$$\kappa_2(S) = \frac{\max\{\|s_1\|,\dots,\|s_n\|\}}{\min\{\|s_1\|,\dots,\|s_n\|\}} \le \sqrt{n}\,.$$

The condition numbers $\kappa_2(A)$ of badly scaled matrices A can be very large, even when $\theta_{\min}(A) = \pi/2$. For instance,

$$A = \begin{bmatrix} 10^5 & 0 \\ 0 & 10^{-5} \end{bmatrix},$$

we have $\theta_{\min}(A) = \pi/2$ and $\kappa_2(A) = 10^{10}$, but if S = SCALGM(A), then S is the identity I_2 with $\theta_{\min}(S) = \pi/2$ and $\kappa_2(S) = 1$.

Now, let S^* be the optimal two-sided scaled matrix of A with $\kappa_2(S^*) = \mu$ and denote $\rho(A)$ the spectral radius of A (i.e., the largest absolute value $|\lambda|$ of an eigenvalue λ of A):

$$\rho(A) = \max\{|\lambda| : \lambda \in \Lambda(A)\}.$$

Then, an estimated interval, due to Rump [17], for the condition number of the optimal two-sided scaled matrix $D^*_1 A D^*_2$, is

$$[\,\rho(|A^{-1}|\,|A|)/n\,,\ \rho(|A^{-1}|\,|A|)].$$

Finally, to approach the condition number $\mu = \mu(A)$ of the optimal two-sided scaled matrix $D^*_1 A D^*_2$, we construct an objective function and apply the pattern search method [8, 19] and its C code [6] on it to find the optimal scale factors that minimize the condition number. The experimental results are sampled in the next section.

Experimental Results

Let $A = [a_1|\dots|a_n] \in \mathbf{R}^{m\times n}$ be given, and let $\theta_j = \theta_j(A)$ be the angle, measured in degrees, between the jth column vector a_j of A and the linear subspace $L(a_1,\dots,a_{j-1},a_{j+1},\dots,a_n)$ spanned by other columns of A. Denote $\theta_{\min} = \theta_{\min}(A) = \min\{\theta_1, \dots,\theta_n\}$. In this section, we will compare the condition number of some matrices that have been scaled using SCALGM to an empirical formula

$$f(A) = 1.4(90/\theta_{\min}(A)) - 0.4$$

that we have found to be a reasonable, if rough, approximation for $\kappa_2(A)$, in terms of the minimum angle $\theta_{\min}(A)$ discussed above. All angles will be measured in degrees.

Example 2. Consider $A \in \mathbf{R}^{4\times 4}$ as follows:

$$A = \begin{bmatrix} 0.1480770 & 7.2100848 & 0.1305800 & 15.4921243 \\ 9.9878992 & 57.2956969 & 1.4177203 & 0.0343257 \\ 0.0175972 & 7.9973998 & 22.2709231 & 1.2756261 \\ 0.3718530 & 0.1517864 & 80.524769 & 0.5036384 \end{bmatrix}$$

Then, we have $\theta_1 = 7.6290198$, $\theta_2 = 7.5595667$, $\theta_3 = 85.4822349$, and $\theta_4 = 47.8432547$. Thus, $\theta_{\min}(A) = 7.5595667$. The 2-norm condition number, $\kappa_2(A) = 64.1742100$, and $1.4(90/\theta_{\min}(A)) - 0.4 = 16.2676219$.

Now, let $S = \text{SCALGM}(A)$, then we have

$$S = \begin{bmatrix} 0.0155002 & 0.1315657 & 0.0021193 & 1.0000000 \\ 1.0000000 & 1.0000000 & 0.0220076 & 0.0021193 \\ 0.0050962 & 0.4037436 & 1.0000000 & 0.2278069 \\ 0.0297831 & 0.0021193 & 1.0000000 & 0.0248746 \end{bmatrix},$$

and we have $\theta_1 = 15.8723160$, $\theta_2 = 15.0347404$, $\theta_3 = 49.5444269$, and $\theta_4 = 59.2475015$. Thus, $\theta_{\min}(S) = 15.0347404$. The 2-norm condition number, $\kappa_2(S) = 8.0578672$, and observe $1.4(90/\theta_{\min}(S)) - 0.4 = 7.9805903$, which is very close to $\kappa_2(S)$. The ratio $\rho(S)$ of min/max in magnitude is 0.0021193.

The estimated interval for the condition number μ of the optimal two-sided scaled matrix is [0.6306089,2.5224355] .

Next, let $T = \textit{PatternSearch}(S)$, then we have

$$T = \begin{bmatrix} 0.1577278 & 0.5316732 & 0.0011684 & 2.1062209 \\ 1.8848862 & 0.7485433 & 0.0022474 & 0.0008268 \\ 0.0618953 & 1.9473666 & 0.6580056 & 0.5726785 \\ 0.9153895 & 0.0258673 & 1.6651569 & 0.1582436 \end{bmatrix},$$

and we have $\theta_1 = 61.1521800$, $\theta_2 = 53.8934825$, $\theta_3 = 62.2106164$, and $\theta_4 = 61.2947051$. Thus, $\theta_{\min}(S) = 53.8934825$. The 2-norm condition number, $\kappa_2(T) = 2.0020228$, and observe $1.4(90/\theta_{\min}(T)) - 0.4 = 1.9379450$, which is very close to $\kappa_2(T)$ and within the above estimated interval. The ratio $\rho(T)$ of min/max in magnitude is 0.00082680/2.10622093 = 0.00039255. Table 1 shows the summary from the previous results.

Table 1: Interval for μ: [0.6306089, 2.5224355]

	A	S	T
θ_1	7.6290198	15.8723160	61.1521800
θ_2	7.5595667	15.0347404	53.8934825
θ_3	85.4822349	49.5444269	62.2106164
θ_4	47.8432547	59.2475015	61.2947051
$\theta_{\min}$	7.5595667	15.0347404	53.8934825
ρ	-	0.0021193	0.00039255
κ_2	64.1742100	8.0578672	2.0020228
$f(\theta_{\min})$	-	7.9805903	1.9379450

Example 3. Consider $A \in \mathbf{R}^{4\times4}$ as follows:

$$A = \begin{bmatrix} 0.0926612 & 17.0784926 & 0.3127063 & 12.7526810 \\ 1.7811361 & 54.0213314 & 1.4953060 & 14.7655003 \\ 0.3460217 & 0.0680433 & 0.2626770 & 0.0227214 \\ 1.3745248 & 45.1500312 & 0.0505958 & 1.4314422 \end{bmatrix}$$

Now, let S = SCALGM(A), then we have

$$S = \begin{bmatrix} 0.0243606 & 0.3890154 & 0.1082947 & 1.0000000 \\ 0.3805432 & 1.0000000 & 0.4208410 & 0.9409447 \\ 1.0000000 & 0.0170377 & 1.0000000 & 0.0195857 \\ 0.3513716 & 1.0000000 & 0.0170377 & 0.1091433 \end{bmatrix}.$$

Next, let T = *PatternSearch*(S), then we have

$$T = \begin{bmatrix} 0.1171181 & 0.6907233 & 0.4922486 & 1.4434180 \\ 1.0922594 & 1.0600407 & 1.1420381 & 0.8108517 \\ 1.1754749 & 0.0073965 & 1.1113581 & 0.0069121 \\ 0.9256365 & 0.9729141 & 0.0424350 & 0.0863229 \end{bmatrix}.$$

The results for matrices A, S, and T regarding the angles θ_j, $\theta_{\min}$, the ratio ρ of min/max in magnitude, the 2-norm condition numbers κ_2, the values of the fitting function $f(\theta_{\min})$, as well as the estimated interval for the optimal condition number μ are summarized in Table 2.

Table 2: Interval for μ: [3.6213799, 14.4855194]

	A	S	T

θ_1	4.8352583	6.3541657	10.8802034
θ_2	4.9049531	14.7800779	14.5441070
θ_3	10.4000159	6.6880465	12.9141023
θ_4	10.0054682	17.3776228	16.0505446
$\theta_{\min}$	4.8352583	6.3541657	10.8802034
ρ	-	0.01703767	0.00039255
κ_2	460.2705191	23.9129780	14.4856257
$f(\theta_{\min})$	-	19.4295112	11.1806658

Example 4. Consider $A \in \mathbf{R}^{6\times 3}$ as follows:

$$A = \begin{bmatrix} 4.2436341 & 0.0269694 & 0.0622383 \\ 0.0149600 & 10.8894167 & 49.9185153 \\ 2.3807229 & 0.0453816 & 0.4947489 \\ 0.0910166 & 0.0291091 & 0.0240973 \\ 0.3740146 & 0.2130833 & 0.2439024 \\ 0.0650553 & 0.1287199 & 0.0238598 \end{bmatrix}$$

Now, let S = SCALGM(A), then we have

$$S = \begin{bmatrix} 1.0000000 & 0.0036574 & 0.0020965 \\ 0.0020965 & 0.8782185 & 1.0000000 \\ 1.0000000 & 0.0109700 & 0.0297065 \\ 1.0000000 & 0.1840528 & 0.0378462 \\ 1.0000000 & 0.3278651 & 0.0932186 \\ 0.8782185 & 1.0000000 & 0.0460428 \end{bmatrix}.$$

Next, let T = *PatternSearch*(S), then we have

$$T = \begin{bmatrix} 1.0638298 & 0.0166244 & 0.0209650 \\ 0.0005098 & 0.9124348 & 2.2857143 \\ 1.7021277 & 0.0797817 & 0.4753033 \\ 0.0000696 & 0.0000547 & 0.0000248 \\ 0.0000696 & 0.0000975 & 0.0000610 \\ 0.5338714 & 2.5974026 & 0.2631015 \end{bmatrix}.$$

The results for matrices A, S, and T regarding the angles θ_j, $\theta_{\min}$, the ratio ρ of min/max in magnitude, the 2-norm condition numbers κ_2, the values of the fitting function $f(\theta_{\min})$, as well as the estimated interval for the optimal condition number μ are summarized in Table 3.

Table 3: Interval for μ: [0.3607154, 2.1642922]

	A	S	T
θ_1	88.6713931	55.7560570	73.4727064
θ_2	1.1215613	36.8741891	61.8206507
θ_3	1.1215571	42.4600963	63.7554493
$\theta_{\min}$	1.1215571	36.8741891	61.8206507
ρ	-	0.00209650	0.00000953
κ_2	245.2917162	4.1852132	1.6924102
$f(\theta_{\min})$	-	3.0170243	1.6381539

Conclusion

In practice, we suggest measuring the true condition number of a matrix by rescaling the matrix using our algorithm and then computing the value of the conventional condition number, while solving the re-scaled linear system. This method does not greatly increase the time to solve a problem and can give a much better measure of the sensitivity of the solution to small perturbations in the problem than does the usual method. We believe that solving linear systems in engineering problems by this method will provide valuable information to the practitioner concerning the reliability and sensitivity of the solution.

References

[1] Bauer, F. L. "Optimally Scaled Matrices." *Numerische Mathematik*, 5:73–87, 1963.

[2] Berman, A., B.N.Parlett, and R.J.Plemmons. "Diagonal Scaling to an Orthogonal Matrix." *SIAM Journal on Algebraic and Discrete Methods*, 2(1):57–65, March 1981.

[3] Braatz, Richard D. and Manfred Morari. "Minimizing the Euclidean Condition Number." *SIAM Journal on Control and Optimization*, 32(6):1763–1768, November 1994.

[4] Bunch, J. R. "Equilibration of Symmetric Matrices in the Max-norm." *J. Assoc. Comput. Mach.*, 18:566–572, 1971.

[5] Businger, P. A. "Matrices Which Can Be Optimally Scaled." *Numerische Mathematik*, 12:346–348, 1968.

[6] Chandler, John P. Private communication.

[7] Curtis, A. R. and J. K. Reid. "On the Automatic Scaling of Matrices for Gaussian Elimination." *J. Inst. Maths. Applics.*, 10:118–124, 1972.

[8] Vogelaere, René de. "Remark on Algorithm 178: Direct search." *Communications of the ACM*, 11(7):498, 1968.

[9] Forsythe, G. and C. Moler. *Computer Solution of Linear Algebraic Systems*. Prentice–Hall, Englewood Cliffs, 1967.

[10] Forsythe, G. E. and E. G. Strauss. “On Best Conditioned Matrices.” *Proc. Amer. Math. Soc*., 6:340–345, 1955.
[11] Fulkerson, D. R. and P. Wolfe. “An Algorithm for Scaling Matrices.” *SIAM Review*, 4:142–146, 1962.
[12] Golub, G. H. and J. M. Varah. “On a Characterization of the Best *L*2-scaling of a Matrix.” *SIAM J. Numer. Anal.*, 11:472–479, 1974.
[13] McCarthy C. and G. Strang. “Optimal Conditioning of Matrices.” *SIAM J. Numer. Anal*., 10:370–388, 1973.
[14] Osborne, E. E.. “On Preconditioning of Matrices.” *J. Assoc. Comput. Mach.*, 7:338–345, 1960.
[15] Parlett, B. N. and C. Reinsch. “Balancing a Matrix for Calculation of Eigenvalues and Eigenvectors.” *Numerische Mathematik*, 13:292–304, 1969.
[16] Rothblum, Uriel G., Hans Schneider, and Michael H. Schneider. “Scaling Matrices to Prescribed Row and Column Maxima.” *SIAM J. Matrix Anal. Appl.*, 15(1):1–14, 1994.
[17] Rump, Siegfried M. “Optimal Scaling for *P*-norms and Componentwise Distance to Singularity.” *IMA J. Numer. Anal.*, 2002. To appear.
[18] Skeel, R. D.. “Scaling for Numerical Stability in Gaussian Elimination.” *J. Assoc. Comput. Mach.,* 26:494–526, 1979.
[19] Smith, Lyle B.. “Remark on Algorithm 178 [E4]: Direct search.” *Communications of the ACM*, 12(11):638–638, November 1969.
[20] Watson, G. A.. “An Algorithm for Optimal 2 Scaling of Matrices.” *IMA J. Numer. Anal.*, 11:481–492, 1991.
[21] Wilkinson, J. H.. “The Algebraic Eigenvalue Problem.” *Monographs on Numerical Analysis*, p. 662, Clarendon Press, 1965