Scaling and structural condition numbers

Arnold Neumaier

Institut für Mathematik, Universität Wien
Strudlhofgasse 4, A-1090 Wien, Austria
email: neum@cma.univie.ac.at

revised, August 1996

Abstract. We introduce structural condition numbers and discuss their significance for the proper scaling of nonsymmetric and symmetric matrices.

1 Introduction

Direct methods for solving a system of linear equations $Ax = b$ with a square nonsingular coefficient matrix A are generally based on matrix factorizations. Unless the matrix is positive definite, the factorization always involves some form of pivoting in order to have satisfactory numerical stability properties. However, to get good pivot choices, the matrices must be properly scaled.

A common suggestion is to scale such that the scaled matrix be equilibrated, i.e., all its rows have l_1-norm 1. However, this can be disastrous for the subsequent factorization when the entries in some row of A have widely differing magnitudes, and it is particularly unsatisfactory for sparse matrices (Curtis & Reid [3]).

One can improve the behavior by enforcing also that the transposed matrix is equilibrated, i.e., that all columns have norm 1. Ignoring the signs, the scaled matrix is then doubly stochastic, i.e., the entries of each row and column sum to 1. Now the scaling matrices must satisfy nonlinear equations that are not easy to solve. Parlett & Landis describe in [8] (expensive) iterative procedures for scaling nonnegative matrices to double stochastic form, and they present results of tests comparing the new algorithms with other methods.

Recently, OLSCHOWKA & NEUMAIER [7] introduced another idea for choosing scaling matrices D_1 and D_2 in such a way that D_1AD_2 has the structure of a permuted I-matrix.

In this paper we relate scaling problems to questions of the closeness of a given matrix to structurally singular matrices, measured by a structural condition number. Among other results, we show that both doubly stochastic equilibration and scaling to permuted I-matrix form leads to bounded structural condition numbers, thus justifying their observed good performance in the context of Gauss elimination.

We also show that the permuted I-matrix scaling is naturally related to rank one bounds for the absolute value of A. Such bounds are required in NEUMAIER [6] for the construction of good scales for modified Cholesky factorizations of indefinite symmetric matrices.

2 Structural singularity and structural condition numbers

A square matrix $A \in \mathbb{R}^{n\times n}$ is called *structurally singular* if every matrix A' is singular that has zeros $A'_{ik} = 0$ whenever $A_{ik} = 0$, and *structurally nonsingular* otherwise. A *transversal* of A is a permutation π of $\{1,\dots,n\}$ such that $A_{i,\pi i} \neq 0$ for $i = 1,\dots,n$. Every nonsingular matrix has a transversal. More generally, since the determinant is the sum of signed products of entries on transversals, a square matrix has a transversal iff it is structurally nonsingular.

The *structural condition number* of $A \in \mathbb{R}^{n\times n}$ is defined as

$$q(A) = \begin{cases} \infty & \text{when } A \text{ is structurally singular,} \\ \dfrac{\max\limits_{i,k} |A_{ik}|}{\sup\{\delta \mid A^\delta \text{ structurally nonsingular}\}} & \text{otherwise.} \end{cases}$$

Here, A^δ denotes the matrix with entries

$$A^\delta_{ik} = \begin{cases} A_{ik} & \text{if } |A_{ik}| \ge \delta, \\ 0 & \text{otherwise.} \end{cases}$$

The structural condition number is a measure of how close A is to a structurally singular matrix. The relevance of the structural condition number is the fact that a nonsingular matrix A with large $q(A)$ contains in some of its small entries significant information about its inverse.

Since $A^\delta = 0$ when $\delta > \max\limits_{i,k} |A_{ik}|$, we always have $q(A) \ge 1$. Therefore, the structurally best behaved matrices are those with $q(A) = 1$. If this holds, the matrix A^δ with $0 < \delta = \max\limits_{i,k} |A_{ik}|$ has all its entries $\in \{0, \delta, -\delta\}$ and is nonsingular, hence has a transversal. After permuting this transversal to the diagonal and dividing the ith rows of the permuted matrix PA by $(PA)_{ii} \in \{\delta, -\delta\}$, we obtain an *$I$-matrix*, i.e., a matrix B with

$$B_{ii} = 1 \ge |B_{ik}| \quad \text{for all } i, k.$$

Conversely, every permuted I-matrix B has $q(B) = 1$ and is structurally optimally conditioned. The main result of Olschowka & Neumaier [7] is that one can scale every structurally nonsingular matrix to a permuted I-matrix, cf. Section 3 below. As we shall show in Section 4, structurally nonsingular symmetric matrices can be scaled to a *symmetric* permuted I-matrix. Therefore, for both nonsymmetric and symmetric matrices, *large structural condition numbers are a reflection of bad scaling.*

For relative perturbations we have the following 'structural' analogue of the standard result

$$\|A^{-1}\|\ \|B - A\| \le \varepsilon \quad \Rightarrow \quad \frac{1-\varepsilon}{1+\varepsilon} \operatorname{cond}(A) \le \operatorname{cond}(B) \le \frac{1+\varepsilon}{1-\varepsilon} \operatorname{cond}(A);$$

in the following, absolute values are taken componentwise.

2.1. Proposition.

$$|B - A| \le \varepsilon |A| \quad \Rightarrow \quad \frac{1-\varepsilon}{1+\varepsilon} q(A) \le q(B) \le \frac{1+\varepsilon}{1-\varepsilon} q(A).$$

Proof. Straightforward. □

We have the following characterization of structural condition numbers in terms of submatrices with small entries.

2.2. Proposition. *$A \in \mathbb{R}^{n\times n}$ has structural condition number $q(A) \geq \max\limits_{i,k} |A_{ik}|/\mu$ iff there is a submatrix A_{JK} with $|J|+|K|>n$ and $\max\{|A_{jk}| \mid i \in J, k \in K\} \leq \mu$.*

Proof. Proposition 2.3 of [7] states that A is structurally nonsingular iff $m_1 + m_2 \leq n$ for all $m_1 \times m_2$ - submatrices containing zeros only. Thus A^δ is structurally singular iff some A_{JK} with $|J| + |K| > n$ satisfies $\max\{|A_{jk}| \mid i \in J, k \in K\} < \delta$. This implies the assertion. □

Parlett & Landis [8] suggest to scale matrices to doubly stochastic form. That this scaling procedure is reasonable from the point of view of structural condition follows from our next theorem.

2.3. Theorem. *Every doubly stochastic matrix $A \in \mathbb{R}^{n\times n}$ satisfies*

$$q(A) \leq n(n+1).$$

Proof. This is the case $\alpha = \beta = \gamma = 1$ of the following result. □

2.4. Proposition. *Suppose $A \in \mathbb{R}^{n\times n}$ satisfies*

$$|A|e \geq \alpha e, \quad |A^T|e \leq \beta e, \quad |A| \leq \gamma e e^T,$$

where $e = (1,\ldots,1)^T$; i.e., $|A|$ has row sums $\geq \alpha$, column sums $\leq \beta$, and entries $\leq \gamma$. Then

$$\frac{\alpha}{\beta} \geq \frac{n}{n+1} \quad \Rightarrow \quad q(A) \leq n(n+1)\frac{\gamma}{\beta}.$$

Proof. We consider a submatrix of the form in Proposition 2.2. Let $r = |J|, s = |K|$ and denote by c and d the sum of all A_{ik} with $i \in J$ and $k \notin K$ or $k \in K$, respectively. Looking at column sums, we find $c \leq \beta(n-s)$; looking at row sums, we find $c + d \geq \alpha r$. Hence $d \geq \alpha r - \beta(n-s)$. Since

$\alpha/\beta \geq n/(n+1)$ by assumption, and $r+s \geq n+1$, there is some A_{ik} with $i \in J, k \in K$ satisfying

$$\begin{aligned} |A_{ik}| &\geq \frac{\alpha r - \beta(n-s)}{rs} = \frac{\alpha}{s} - \frac{\beta(n-s)}{rs} \\ &\geq \frac{\beta n}{(n+1)s} - \frac{\beta(n-s)}{(n+1-s)s} \\ &= \frac{\beta}{(n+1)(n+1-s)} \geq \frac{\beta}{(n+1)n}. \end{aligned}$$

By, Proposition 2.2, $|A_{ik}| \leq q(A)^{-1}$, and the result follows. □

3 Optimal rank one bounds

Let $A \in \mathbb{R}^{n\times n}$. Suppose we have (diagonal and nonsingular) scaling matrices D_1 and D_2 such that all entries of D_1AD_2 have absolute value ≤ 1. Then the rank one matrix

$$B = (\beta_i\gamma_k)_{i,k=1}^n \tag{1}$$

with $\beta_i = (D_1)_{ii}^{-1}, \quad \gamma_i = (D_2)_{ii}^{-1}$ is a componentwise upper bound for $|A|$, and we may pose the scaling problem as one of finding small $\beta_i, \gamma_k > 0$ such that

$$|A_{ik}| \leq \beta_i\gamma_k \quad \text{for all } i,k. \tag{2}$$

3.1. Theorem. *Let A be structurally nonsingular. Among all rank one matrices* (1) *satisfying $|A| \leq B$, the minimal value of $\prod(\beta_i\gamma_i)$ is attained for values such that* (2) *holds with equality for $k = \pi i$ $(i = 1,\ldots,n)$, for some transversal π.*

Proof. (2) implies

$$\prod_i |A_{i,\pi i}| \leq \prod_i (\beta_i\gamma_{\pi i}) = \prod_i (\beta_i\gamma_i)$$

for any transversal π, so that

$$\prod_i (\beta_i\gamma_i) \geq \max_\pi \prod_i |A_{i,\pi i}|.$$

The right hand side is positive since A is structurally nonsingular. Equality holds iff $\beta_i\gamma_{\pi i} = |A_{i,\pi i}|$ for some transversal π, i.e., iff D_1AD_2 is a permuted I-matrix, where $D_1 = \mathrm{Diag}(\beta_i^{-1}|i = 1,\ldots n)$ and $D_2 = \mathrm{Diag}(\gamma_i^{-1}|i = 1,\ldots n)$. But, according to [7], such a transversal exists. $\square$

If we write $i \to k$ if $A_{ik} \neq 0$ and introduce

$$\alpha_{ik} = \log|A_{ik}| \qquad \text{if } i \to k,$$

$$y_i = \log\beta_i, \quad z_k = \log\gamma_k,$$

the problem to minimize $\prod(\beta_i\gamma_i)$ subject to (2) can be rewritten as the linear program

$$\begin{array}{ll} \min & \sum y_k + \sum z_k \\ \text{s.t.} & y_i + z_k \geq \alpha_{ik} \quad \text{for} \quad i \sim k, \end{array} \tag{3}$$

The dual of the linear program (3) is a linear program associated with the weighted matching problem for the directed graph with adjacency relation $\to$ and weights α_{ik} (for details see, e.g., [7]); thus (3) can be solved by algorithms for the weighted bipartite matching problem [1, 2, 4, 5, 7].

4 Optimal symmetric scaling

For symmetric matrices $A \in \mathbb{R}^{n\times n}$ one generally wants to find a scaling procedure that preserves symmetry and hence uses the same scaling matrix D on the left and on the right. Equivalently, we want to find a good bound for $|A|$ by a rank one matrix

$$B = (\beta_i\beta_k)_{i,k=1}^n. \tag{4}$$

Thus we want to find small $\beta_k > 0$ such that

$$|A_{ik}| \leq \beta_i\beta_k \quad \text{for all } i, k. \tag{5}$$

Numbers β_i satisfying (5) are required in NEUMAIER [6] for the construction of good scales for modified Cholesky factorizations of indefinite symmetric matrices.

4.1. Theorem. *Let the symmetric matrix A be structurally nonsingular. Among all rank one matrices (4) with $|A| \le B$, the minimal value of $\prod \beta_i$ is attained for values such that (5) holds with equality for $k = \pi i$ ($i = 1, \dots, n$), for some transversal π.*

Proof. (5) implies

$$\prod_i |A_{i,\pi i}| \le \prod_i (\beta_i \beta_{\pi i}) = \prod_i \beta_i^2$$

for any transversal π, so that

$$\prod_i \beta_i^2 \ge \max_\pi \prod_i |A_{i,\pi i}|. \tag{6}$$

Now suppose that we have a solution of the problem to minimize $\prod(\beta_i \gamma_i)$ subject to (2). Then the $\beta_i' = \sqrt{\beta_i \gamma_i}$ satisfy

$$\beta_i' \beta_k' = \sqrt{\beta_i \gamma_i \beta_k \gamma_k} = \sqrt{\beta_i \gamma_k \beta_k \gamma_i} \ge \sqrt{|A_{ik}|\ |A_{ki}|} = |A_{ik}|$$

and

$$\prod_i (\beta_i')^2 = \prod_i (\beta_i \gamma_i) = \max_\pi \prod_i |A_{i,\pi i}|$$

by Theorem 3.1. A comparison with (6) shows the optimality of β_i'. Hence equality holds in the argument leading to (6), and for any maximizing π, (5) holds with equality for $k = \pi i$ ($i = 1, \dots, n$). $\square$

If we write $i \sim k$ if $A_{ik} \ne 0$ and introduce

$$\alpha_{ik} = \alpha_{ki} = \log |A_{ik}| \qquad \text{if } i \sim k,$$

$$x_k = \log \beta_k,$$

the problem to minimize $\prod \beta_i^2$ subject to (5) can be rewritten as the linear program

$$\begin{array}{ll} \min & \sum x_k \\ \text{s.t.} & x_i + x_k \ge \alpha_{ik} \ \text{ for } i \sim k. \end{array} \tag{7}$$

From a solution of the nonsymmetric problem (3) we can find a solution of (7) by setting $x_k = \frac{1}{2}(y_k + z_k)$. So we can solve (7), too, using matching algorithms. But perhaps some work can be saved there by exploiting symmetry and/or looking only for involutory paths, cf. the following remark.

4.2. Remark. When the dominant transversal is unique (and this holds with probability one under natural stochastic assumptions) then we must have $\pi^{-1} = \pi$ since π^{-1} is again a transversal. Thus π^2 is the identity. Therefore the active constraints

$$x_i + x_{\pi i} = \alpha_{i,\pi i}$$

fix x_i when $\pi i = i$, but if $\pi i \neq i$ we only get

$$x_i = \frac{1}{2}\alpha_{i,\pi i} + w_i \text{ with } w_i + w_{\pi i} = 0.$$

In particular, if the dominant transversal is diagonal (π = identity, which can be checked directly by testing (5) with $\beta_i = |A_{ii}|$), the solution of (7) is unique. Otherwise, there may be multiple solutions and we want to select a small one. Since $x_i^2 + x_{\pi i}^2 = \frac{1}{2}\alpha_{i,\pi i}^2 + 2w_i^2$, we get the x_i with smallest Euclidean norm, by solving the convex quadratic program

$$\begin{array}{ll} \min & \sum w_i^2 \\ \text{s.t.} & w_i + w_k \geq \bar{\alpha}_{ik} := \alpha_{ik} - \frac{1}{2}\alpha_{i,\pi i} - \frac{1}{2}\alpha_{k,\pi k}, \\ & w_i + w_{\pi i} = 0. \end{array}$$

The involution π and a starting point are available from solving the matching problem (3). This gives small $|w_i|$, hence small x_i and β_i.

References

1. R. E. Burkard; U. Derigs. Assignment and Matching Problems: Solution Methods with FORTRAN-Programs. Lecture Notes in Econ. and Math. Systems, Springer, Berlin 1980.

2. G. Carpaneto; P. Toth. Solution of the assignment problem (Algorithm 548). ACM Trans. Math. Softw. (1980), 104-111.

3. A. R. Curtis; J. K. Reid. On the automatic scaling of matrices for Gaussian elimination. J. Inst. Math. Appl. 10 (1972), 118-124.

4. U. Derigs; A. Metz. An efficient labeling technique for solving sparse assignment problems. Computing 36 (1986), 301-311.

5. H. W. Kuhn. The Hungarian method for the assignment problem. Naval Research Logistics Quarterly 2 (1955) 83.

6. A. Neumaier, On satisfying second-order optimality conditions by modified Cholesky factorizations, submitted.

7. M. Olschowka and A. Neumaier, A new pivoting strategy for Gaussian elimination, Linear Algebra Appl. 240 (1996), 131-151.

8. B. N. Parlett; T. L. Landis. Methods for scaling to double stochastic form. Linear Algebra Appl. 48 (1982), 53-79.