A scaling algorithm to equilibrate both rows and columns norms in matrices[1]

Daniel Ruiz[2]

Scaling or equilibration of data in linear systems of equations is a topic of great importance that has already been the subject of many scientific publications, with many different developments depending on the properties one wants to obtain after scaling. It has given rise to several well known algorithms (see Duff, Erisman, and Reid (1986), Schneider and Zenios (1990), for instance).

Scaling consists in pre- and post-multiplying a matrix by two diagonal matrices, $\mathbf{D}_1$ and $\mathbf{D}_2$, respectively. Classical scalings are the well known row and column scaling. A more general purpose scaling method is the one used in the HSL 2000 routine `MC29`, which aims to make the nonzeros of the scaled matrix close to one by minimizing the sum of the squares of the logarithms of the moduli of the nonzeros (see Curtis and Reid (1972)). `MC29` reduces this sum in a global sense and therefore should be useful on a wide range of sparse matrices. Any combination of these scalings is also a possibility.

Scaling can also be combined with permutations (see Duff and Koster (1999) and the HSL 2000 routine `MC64`). The matrix is first permuted so that the product of absolute values of entries on the diagonal of the permuted matrix is maximized (other measures such as maximizing the minimum element are also options). Then the matrix is scaled so that the diagonal entries are one and the off-diagonals are less than or equal to one. This then provides a useful tool for a good pivoting strategy for sparse direct solvers, as well as for building good preconditioners for an iterative method.

In the 1960's, Bauer (1963), Bauer (1969) and van der Sluis (1969), in particular, showed some optimal properties in terms of conditions numbers for scaled matrices with all rows or all columns of equal norm of 1.

We present an iterative procedure which asymptotically scales the infinity norm of both rows and columns in a matrix to 1. To describe the algorithm, let us first denote by $\mathbf{r}_i = \mathbf{a}_{i,:}^T \in I\!\!R^{n \times 1}$, $i = 1, \ldots, m$, the row-vectors from a general $m \times n$ real matrix $\mathbf{A}$, and by $\mathbf{c}_j = \mathbf{a}_{:,j} \in I\!\!R^{n \times 1}$, $j = 1, \ldots, n$, the column-vectors from $\mathbf{A}$. With these notations, the idea of the algorithm holds in the following iteration (assuming that the matrix $\mathbf{A}$ does not have any empty row or column):

Algorithm 1 (Simultaneous row and column iterative scaling)

$$\widehat{\mathbf{A}}^{(0)} = \mathbf{A}, \quad \mathbf{D}_1^{(0)} = \mathbf{I}_m, \quad \text{and } \mathbf{D}_2^{(0)} = \mathbf{I}_n$$

for $k = 0, 1, 2, \ldots$, until convergence do :

$$\mathbf{D}_R = \text{diag}(\sqrt{\|\mathbf{r}_i^{(k)}\|_\infty})_{i=1,\ldots,m}, \quad \text{and } \mathbf{D}_C = \text{diag}(\sqrt{\|\mathbf{c}_j^{(k)}\|_\infty})_{j=1,\ldots,n}$$

$$\widehat{\mathbf{A}}^{(k+1)} = \mathbf{D}_R^{-1} \widehat{\mathbf{A}}^{(k)} \mathbf{D}_C^{-1}$$

$$\mathbf{D}_1^{(k+1)} = \mathbf{D}_1^{(k)} \mathbf{D}_R^{-1}, \quad \text{and } \mathbf{D}_2^{(k+1)} = \mathbf{D}_2^{(k)} \mathbf{D}_C^{-1}$$

[2]Ecole Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique, et d'Hydraulique de Toulouse, Institut de Recherche en Informatique de Toulouse, 2 rue Camichel, 31071 Toulouse Cedex, France.

Convergence is obtained when

$$\max_{1\leq i\leq m}\left\{|(1-\|\mathbf{r}_i^{(k)}\|_\infty)|\right\} \leq \varepsilon \;\; \text{and} \;\; \max_{1\leq j\leq n}\left\{|(1-\|\mathbf{c}_j^{(k)}\|_\infty)|\right\} \leq \varepsilon \tag{1}$$

for a given value of $\varepsilon > 0$.

The purpose of this presentation is to detail the properties of this algorithm. In particular, the convergence towards the stationary state mentioned above is at least linear, with an asymptotic rate of convergence of $\frac{1}{2}$. We highlight the case of symmetric matrices since the algorithm "*naturally*" preserves such numerical structures. In that respect, we also mention the routine MC30 in the HSL (2000) library, which is a variant of the above MC29 routine for symmetric matrices.

We also discuss possible extensions of such an algorithm when considering the one-norm or the two norm of the rows and columns of the given matrix. Following the discussion in Parlett and Landis (1982), we establish under which hypothesis the algorithm is also convergent in the case of the one-norm, and we comment on the generalisation of these results with respect to what was stated in Parlett and Landis (1982).

We plan to illustrate with some small examples the major differences between Algorithm 1 and some of the well known scaling algorithms from the litterature. We shall also investigate possible combinations of this scaling algorithm in the one-norm and the infinity-norm to reach situations where the scaled matrix can easily be permuted into a matrix with dominant elements on the diagonal. This is a favourable situation for a good pivoting strategy in Gaussian elimination, for instance, if we refer to the motivations for scaling in Duff, Erisman, and Reid (1986).

References

F. L. BAUER, (1963), *Optimally scaled matrices*, Numer. Math., **5**, 73–87.

F. L. BAUER, (1969), *Remarks on optimally scaled matrices*, Numer. Math., **13**, 1–3.

A. R. CURTIS AND J. K. REID, (1972), *On the automatic scaling of matrices for Gaussian elimination*, J. Inst. Maths. Applics., **10**, 118–124.

I. S. DUFF AND J. KOSTER, (1999), *On algorithms for permuting large entries to the diagonal of a sparse matrix*, Tech. Rep. RAL-TR-1999-030, Rutherford Appleton Laboratory. To appear in *SIAM Journal on Matrix Analysis and Applications*.

I. S. DUFF, A. M. ERISMAN, AND J. K. REID, (1986), *Direct Methods for Sparse Matrices*, Oxford University Press, London.

HSL, (2000), *A collection of Fortran codes for large scale scientific computation.* http://www.cse.clrc.ac.uk/Activity/HSL.

B. N. PARLETT AND T. L. LANDIS, (1982), *Methods for Scaling to Double Stochastic Form*, Linear Algebra and its Applications, **48**, 53–79.

M. H. SCHNEIDER AND S. ZENIOS, (1990), *A comparative study of algorithms for matrix balancing*, Operations Research, **38**, 439–455.

A. VAN DER SLUIS, (1969), *Condition Numbers and Equilibration of Matrices*, Numer. Math., **14**, 14–23.