Extending doubly stochastic scaling to bipartite graphs

P. A. Knight[1], L. Le Gorrec[1], S. Mouysset[2], D. Ruiz[2]

[1] University of Strathclyde, United Kingdom
p.a.knight@strath.ac.uk, legorrecluce@gmail.com
[2] IRIT, Université de Toulouse, France
{sandrine.mouysset,daniel.ruiz}@irit.fr

Mots-clés : *Bipartite graphs, Scaling, Combinatorial matrix theory.*

Abstract

Doubly stochastic scaling [1] has a long history of being used to scale a graph adjacency matrix $\mathbf{A}$. It has proven useful for graph analysis, e.g. to emphasize critical arcs in digraphs [1], or to partition graphs [2]. The conditions for the scaling to exist are well known and rely on the graph structure. How this extends when the (potentially rectangular) matrix $\mathbf{A}$ represents a bipartite graph is non straightforward though. In this work we investigate a new type of scaling that naturally enlarges doubly stochastic scaling to rectangular matrices.

If $\mathbf{A} \geq 0$ is a square non-negative matrix with total support, then we can find a diagonal scaling so that $\mathbf{DAE}$ is doubly stochastic (where $\mathbf{D}$ and $\mathbf{E}$ are diagonal matrices with positive diagonal). If $\mathbf{A} \geq 0$ is rectangular and has sufficient nonzeros, then it too can be scaled so that it has constant row and column sums (but no longer equal). Alternatively, one can prescribe arbitrary row and column sums, $\mathbf{r} \in \mathbb{R}^m$ and $\mathbf{c} \in \mathbb{R}^n$ (so long as $\sum_{i=1}^m |r_i| = \sum_{j=1}^n |c_j|$), and scale $\mathbf{A}$ so that $\mathbf{DAE}\mathbb{1}_n = \mathbf{r}$ and $\mathbf{EA}^T\mathbf{D}\mathbb{1}_m = \mathbf{c}$. Again, the precise conditions for existence of such a scaling depend on $\mathbf{r}$, $\mathbf{c}$ and $\mathbf{A}$, and were set out by Brualdi [3], but they cannot be as neatly described as in the square case.

A generic condition [4] for a given scaling to exist for $\mathbf{A}$ is that there exists a nonnegative matrix $\mathbf{B}$ with the same pattern as that of $\mathbf{A}$ for which $\mathbf{B}\mathbb{1}_n = \mathbf{r}$ and $\mathbf{B}^T\mathbb{1}_m = \mathbf{c}$. If a scaling exists, in both the square and rectangular cases, then it can be found using the Sinkhorn–Knopp algorithm [5, 6]. In fact, the existence of a scaling (particularly in the rectangular case) is confirmed by the convergence of this algorithm, although it may be more insightful to verify that Brualdi's conditions hold.

In this work we extend the class of doubly stochastic matrices to include a set of rectangular matrices. While it is impossible for a non-square nonnegative matrix to have row and column sums both equal to one, since the sum of row sums must be equal to the sum of column sums, we may however extend a weaker condition satisfied by doubly stochastic matrices. We consider nonnegative matrices for which

$$\mathbf{AA}^T\mathbb{1}_m = \mathbb{1}_m \text{ and } \mathbf{A}^T\mathbf{A}\mathbb{1}_n = \mathbb{1}_n \,. \tag{1}$$

This trivially holds for doubly stochastic matrices since

$$\mathbf{AA}^T\mathbb{1} = \mathbf{A}\mathbb{1} = \mathbb{1} \text{ and } \mathbf{A}^T\mathbf{A}\mathbb{1} = \mathbf{A}^T\mathbb{1} = \mathbb{1}.$$

But it is a bigger class, even in the square case, as can be seen with

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \end{pmatrix},$$

1. Given $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbb{1} = (1, ..., 1)^T \in \mathbb{R}^n$, it means finding two diagonal matrices $\mathbf{D}, \mathbf{E}$ such that $\mathbf{DAE}\mathbb{1} = \mathbf{EA}^T\mathbf{D}\mathbb{1} = \mathbb{1}$

for which we have

$$\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

which is doubly-stochastic even though $\mathbf{A}$ is not. Notice that $\mathbf{A}$ does not have support, so it cannot even be scaled to doubly stochastic form.

We label as "*semi-doubly stochastic*" any nonnegative matrix, square or rectangular ($\mathbf{A} \in \mathbb{R}^{m\times n}$), for which (1) holds. We first show that such a matrix is essentially the direct sum of p connected rectangular sub-components $\mathbf{A}_i$, $i = 1, \ldots, p$, where $\mathbf{A}_i \in \mathbb{R}^{m_i\times n_i}$, each having constant row sums equal to $\sqrt{n_i/m_i}$ together and constant column sums equal to $\sqrt{m_i/n_i}$. In other words, any semi-doubly stochastic matrix $\mathbf{A}$ can be permuted to block diagonal form

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ & & \mathbf{A}_p \end{bmatrix},$$

in which each rectangular sub-matrix $\mathbf{A}_i \in \mathbb{R}^{m_i\times n_i}$ is connected (e.g. the underlying bipartite graphs are connected) and has constant row and column sums equal to the above mentioned ratios.

A question that naturally arises is whether a given nonnegative matrix can be scaled to semi-doubly stochastic form. This is not always feasible, as we can see for instance with matrix $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. The question is then whether for some scalars $\alpha, \beta, \gamma, \delta > 0$ the matrix $\mathbf{B} = \begin{bmatrix} \alpha\beta & \alpha\gamma \\ 0 & \delta\gamma \end{bmatrix}$ can be semi-doubly stochastic. This would require that $(\alpha\gamma)^2 = 0$, and so $\mathbf{B}$ does not exist. We denote as "*semi-scalable*" any nonnegative matrix $\mathbf{A}$ for which we can find diagonal matrices $\mathbf{D}$ and $\mathbf{E}$ such that, if $\mathbf{B} = \mathbf{DAE}$, then $\mathbf{B}^T\mathbf{B}$ and $\mathbf{B}\mathbf{B}^T$ are doubly stochastic matrices.

While the block structure of semi-scalable matrices looks attractive, as this clearly has applications in co-clustering, there is no easy way to tell a priori whether a matrix has this property or not. In practice, if we attempt to use current scaling algorithms on such matrices without pre-existing knowledge of the underlying block structure, then they will fail to converge to anything meaningful. To remedy this, we present a new iterative scaling algorithm, which essentially targets the row sums of both $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$. We also prove that a matrix is semi-scalable if and only if the algorithm converges, providing in the limit a diagonal scaling so that $\mathbf{DAE}$ is semi-doubly stochastic. Additionally, we illustrate the behaviour of the algorithm on matrices which are not semi-scalable. The algorithm still converges to a semi-doubly stochastic matrix but in this case it is one whose nonzero pattern is included in that of the original, as certain entries are forced to go to zero.

Références

[1] D. J. Hartfiel and J. W. Spellman. A Role for Doubly Stochastic Matrices in Graph Theory. *Proceedings of the American Mathematical Society*, vol. 36, no. 2 : pp. 389–394, 1972.

[2] P. B. Slater. Hierarchical Internal Migration Regions of France. *IEEE Transaction on System, Man Cybernetics*. Vol. 4 : p. 321-324, 1976.

[3] R. A. Brualdi. Convex Sets of Non-Negative Matrices. *Canadian Journal of Mathematics*. Vol. 20 : p. 144–157, 1968.

[4] M. V. Menon. Matrix links, an extremization problem, and the reduction of a non-negative matrix to one with prescribed row and column sums. *Canadian Journal of Mathematics*, vol. 20 : p. 225-232, 1968.

[5] O. Pretzel. Convergence of the iterative scaling procedure for non-negative matrices. *Journal of the London Mathematical Society*, vol. 2, no 2 : p. 379-384, 1980.

[6] R. Sinkhorn and P. Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, vol. 21, no 2 : p. 343-348, 1967