IMPLEMENTATION OF TARJAN'S ALGORITHM IN MATLAB

Ivana Pultarová

Department of Mathematics, Faculty of Civil Engineering,
Czech Technical University in Prague

Abstract

A lot of technical problems lead to systems of linear equations. The matrices of the systems are often sparse, then a proper reordering of their rows and columns may reduce the time needed for solution and the validity of the result. We focus in Tarjan's algorithm which permutes the rows and columns of a matrix in order to obtain a block triangular matrix with irreducible diagonal blocks.

1 Irreducible matrices and graph partitioning

A solution of a set of linear equations may be affected by a pattern of nonzero elements of a matrix of the system. Using the Gauss elimination or some numerical methods, an appropriate order of rows and columns may reduce the time of computing and the accuracy of the result.

We focus in a method which determines such a permutation of rows and columns of a matrix that it becomes block upper triangular with irreducible diagonal blocks. In addition, the rows and columns are permuted symmetrically, i.e. once the columns j and k are interchanged, then also the rows j and k are switched. This method is called Tarjan's algorithm and is carefully described in [1].

The theory of irreducible matrices is closely related to the graph theory. Matrix B is *reducible* if there exists a permutation matrix P such that

$$P^T BP = \begin{pmatrix} B_1 & B_2 \\ 0 & B_3 \end{pmatrix}, \tag{1}$$

where B_1 and B_3 are square matrices. A matrix is *irreducible* if it is not reducible.

Let us introduce some basic notions from the graph theory. Any $N \times N$ matrix B can be associated to a *directed graph* consisting of a set of N *nodes* (*vertices*) $v_1, \dots, v_N$ and of a set of *edges*, each of which is an ordered pair of nodes (v_i, v_j), corresponding to some nonzero element B_{ij}, $i, j = 1, 2, \dots, n$.

A *path* from node v_i to node v_j is a sequence of edges (v_i, v_{i_1}), (v_{i_1}, v_{i_2}), $\dots$, (v_{i_m}, v_j). A path is a *cycle* when $i = j$. A *subgraph* consists of a subset of vertices and of the edges adjacent to these vertices. A subgraph is said to be *strongly connected* if there exists a path from any of its nodes to any other. A subgraph is a *strongly connected component* if it is strongly connected and cannot be enlarged to any other strongly connected subgraph by adding extra nodes of the graph and the associated edges.

Clearly, each vertex may belong to exactly one strongly connected component. A strongly connected component may consist of a single vertex or it may involve all vertices of the graph. Thus the strongly connected components define a unique partition of a graph.

Matrix B is irreducible if and only if the associate graph is strongly connected, i.e. it contains exactly one strongly connected component which includes all of the nodes. Then finding the irreducible diagonal blocks of the matrix is equal to determining all of the strongly connected components in the graph.

Examples of directed graphs are shown in Figure 1. In the first figure there are three vertices $1, 2, 3$ and four edges a, b, c, d. A matrix representation can be

$$B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & b & 0 \\ a & c & d \end{pmatrix}. \tag{2}$$

A permutation given by

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}. \tag{3}$$

yields

$$P^T BP = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & b & 0 \\ a & c & d \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} d & a & c \\ 0 & 0 & 0 \\ 0 & 0 & b \end{pmatrix}, \tag{4}$$

while the permutation given by

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \tag{5}$$

yields

$$P^T BP = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & b & 0 \\ a & c & d \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} d & c & a \\ 0 & b & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{6}$$

Thus from any of these two examples of permutations we can see that B is reducible. The first strongly connected component only includes the vertex 3 and the edge d. After omitting them from the graph and also all the edges directed to the vertex 3, we obtain a reduced graph which includes only two vertices 1 and 2 and one edge b. Then the second and third strongly connected components are vertex 1 and vertex 2 with the edge b, respectively.

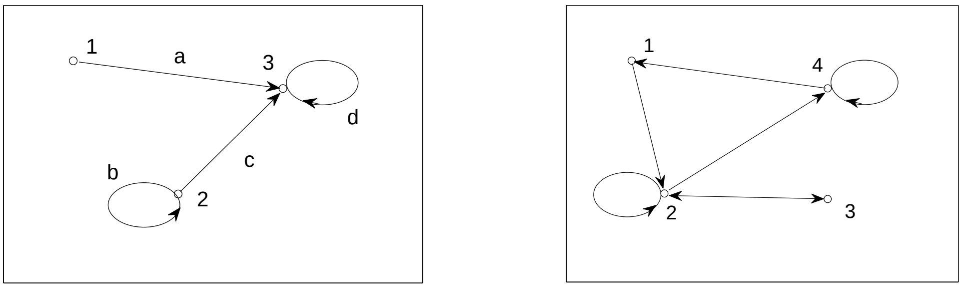

Figure 1: Graph of coincidence of matrix B given by (2) (left). An example of a strongly connected graph (right).

An algorithm for determining all of the strongly connected components is called Tarjan's algorithm and is described in [1]. The result of this algorithm is the symmetric permutation of the rows and columns of a matrix such that it has a block upper triangular form with irreducible square diagonal blocks. A decomposition of the set of vertices into the subsets belonging to the individual strongly connected components is unique up to their order.

2 Tarjan's algorithm

We now briefly introduce the Tarjan's algorithm, one can see [1] for the detailed description. Starting from some vertex it traces the paths in the graph until a cycle is found or a node is reached with no edges leaving it. Of course, all nodes in a cycle belong to the same strongly connected component. When a cycle without any outgoing edges is detected, all of its vertices with the edges of the cycle and with all of the incomming edges are removed from the original graph and they are denoted as the first strongly connected component. In a matrix notation, after a proper symmetric permutation, these nodes correspond to the indices of the first diagonal irreducible block. Then the algorithm continues in the same manner starting from some other vertex of the graph. The process finishes when all of the vertices are deleted from the graph. Then the sequence of the deleted vertices defines a permutation such that the correspondingly symmetrically permuted matrix has an upper triangular form with irreducible diagonal blocks.

Let us notice that each nonzero off-diagonal element is inspected only once. In order to minimize the complexity of the algorithm, two stacks are used, let us call them path-stack and node-stack. A node is put on them both when it is newly reached and it is removed from the path-stack but not the node-stack when backtracking takes place. With each node on the node-stack, a pointer, called lowlink, is held to the node lowest on the stack to which a path so far has been found. Lowlink is initialized to the node-stack position of the node itself.

The algorithm consists of a number of major and minor steps. Each major step begins with placing some node of the graph on the path-stack and on the node-stack and it terminates when both stacks become empty. The complexity of the algorithm is $O(N) + O(\tau)$, where N is the dimension of the matrix and τ is the number of nonzero elements.

Two examples are shown in Figures 2 and 3. In Figure 2 a block diagonal matrix B with the blocks of dimensions 50, 70 and 30, respectively, is symmetrically permuted. Then the Tarjan's algorithm is applied. It can be seen that the blocks are recovered up to their order and up to the order of the element inside the blocks. Another example with one off-diagonal nonzero block is introduced in Figure 3.

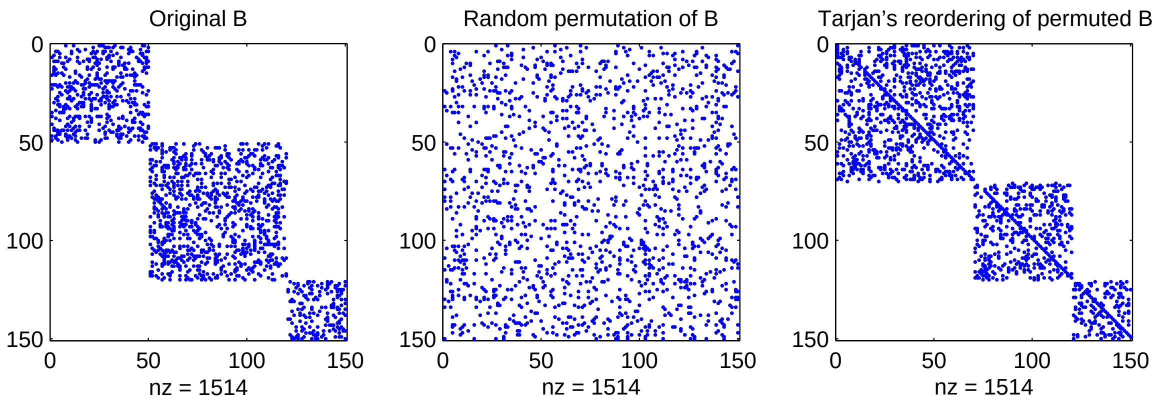

Figure 2: Spy-plot in Matlab of a block diagonal matrix (first), of a random symmetric permutation of its rows and columns (second) and of a result of Tarjan's permutation (third).

3 Concluding remarks

This sophisticated algorithm is not included in any known package of programs yet, so that the author has implemented it in an original form. The algorithm has been written in the language of Matlab.

The speed of the algorithm increases significantly when the nonzero elements of the matrix are stored in a unique vector and their coordinates are in an additional array. Complexity is approximately equal to the number of nonzero elements in B. In practise, maintaining two stacks and searching in the graph may by rather costly. Moreover, in practical examples one of

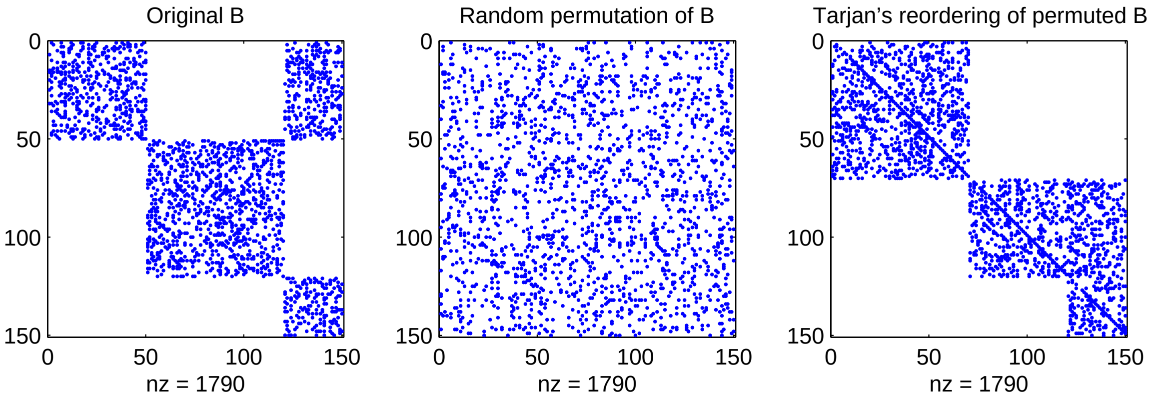

Figure 3: Spy-plot in Matlab of a block diagonal matrix with one non-zero off-diagonal block (first), of a random symmetric permutation of its rows and columns (second) and of a result of Tarjan's permutation (third).

the strongly connected components may be huge relatively to the others [2]. This is why some modifications are incorporated in the program, e.g. considering a threshold for the entries of the matrix.

Acknowledgements

This research has been supported by the Research project No. CEZ MSM 6840770001 of the Ministry of Education of Czech Republic.

References

[1] I. S. Duff, J. K. Reid. *An implementation of Tarjan's algorithm for the block triangularization of a matrix.* ACM transactions on Mathematical Software 4, 1978, pp. 137-147.

[2] K. Avrachenkov, N. Litvak, and K. S. Pham. *Distribution of PageRank mass among principle components of the Web.* Electronic journal arXiv.org, (arXiv:0709.2016v1), 2007.

Ivana Pultarová
Address: Czech Technical University in Prague, Faculty of Civil Engineering, Department of Mathematics, Thákurova 7, 166 29 Praha 6, Czech Republic; phone: +420 224 354 408, e-mail: ivana@mat.fsv.cvut.cz.