A GOOD SUBMATRIX IS HARD TO FIND *

John J. BARTHOLDI, III

School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, U.S.A.

Received June 1982
Revised November 1982

Given a matrix, it is NP-hard to find a 'large' column, row, or arbitrary submatrix that satisfies property π, where π is nontrivial, holds for permutation matrices, and is hereditary on submatrices. Such properties include *totally unimodular, transformable to a network matrix, permutable to consecutive ones*, and many others. Similar results hold for properties such as *positive definite, of bandwidth* w, and *symmetric*.

Submatrix, NP-complete, hereditary property, monotone property

We consider the submatrix problem for fixed matrix property π:

Instance: An $m \times n$ (0, 1)-matrix A; positive integers k_1, k_2.
Question: Does A contain a $k_1 \times k_2$ submatrix which satisfies π?

(We restrict ourselves to (0, 1)-matrices since some of the specific properties in which we are interested are defined only for (0, 1)-matrices.) We show that if π is a 'nontrivial' property that holds for permutation matrices and is hereditary on submatrices, then the submatrix problem is NP-hard. (If testing for π may be accomplished in polynomial time, then the submatrix problem is NP-complete. See Garey and Johnson [7] for explanation of these terms.) These results are observed as corollaries to similar theorems for graph properties—see the remarkable paper by Yannakakis [15], the elegant work of Lewis [8], and their joint paper [9].

One area in which the submatrix problem arises is integer linear programming. A common approach is to find large submatrices of constraints with special structure exploitable by effective solution techniques. Then one solves the corresponding easy subproblems while dealing with everything else by some form of enumeration. For example, if a large row-submatrix of tractable constraints is identified, one may apply some row decomposition method such as Lagrangean relaxation. Similarly, if a large column-submatrix of special structure exists, a column decomposition method such as Bender's decomposition may prove effective. However, this paper implies that even the *finding* of special structure can itself be difficult.

Following Yannakakis, a matrix property is *non-trivial* is it holds for an infinite number of matrices and fails for an infinite number of matrices. (Note that the submatrix problem for trivial properties is easily solved by simply enumerating the fixed and finitely many possibilities.) A matrix property is *hereditary on submatrices* if, for any matrix satisfying the property, all of its submatrices also satisfy the property. Examples of nontrivial and hereditary properties for matrices include *totally unimodular* [11], *balanced* [11], *perfect* [11], *transformable to a network matrix* [1,3] and *permutable to consecutive or circular ones* [4,6,14]. In fact, any matrix property that may be characterized in terms of forbidden submatrices is a hereditary property; if, in addition, it is nontrivial and satisfied by all permutation matrices, the results of this paper apply. Thus, as a corollary to the work of Yannakakis and Lewis, we achieve

* This research was supported in part by the Office of Naval Research under Contract No. N00014-80-k-0709.

0167-6377/82/0000-0000/$02.75

mass-produced complexity results for a host of specific submatrix problems. It is not surprising that any single one of these problems is NP-hard, but it seems remarkable that they may be treated *en masse*.

We also observe analogous results for properties which are hereditary on special submatrices, such as column submatrices, row submatrices, or principal submatrices. Examples of the latter include *positive definite P-matrix,* and other properties associated with complementarity theory; see e.g. Murthy [10].

Let π be a property for (0, 1)-matrices that is non-trivial, satisfied by all permutation matrices, and is hereditary on submatrices.

Theorem 1. *The submatrix problem for π is NP-hard.*

Proof. We show that the node deletion problem for bipartite graphs is reducible to the submatrix problem. In fact, we shall work with a slight variation of the node deletion problem for bipartite graphs, the NP-hardness of which is implicit in [15]: given a bipartite graph G with m vertices on one side and n vertices on the other side, and a positive integer $k \leqslant m+n$, is there an induced subgraph on n nodes that satisfies π?

We associate bipartite graphs and (0, 1)-matrices in a familiar way. Let u_i $(i=1,\ldots,m)$ be the vertices of one side of G and let v_j $(j=1,\ldots,n)$ be the vertices of the other side. With G associate the $m \times n$ (0, 1)-matrix which has a row i for every vertex u_i of G and a column j for every vertex v_j of G, and entry $a_{ij}=1$ *else* 0 if there is an edge (u_i, v_j) in G. (Such a mapping is not unique since to G there may correspond either A or A^{T}; the inverse mapping is unique.)

Let π be a property for bipartite graphs. We say that a given (0, 1)-matrix A has property π *iff* its corresponding bipartite graph has property π. Similarly, a matrix property may be said to hold for a graph *iff* it holds for one of the matrices corresponding to the graph. Then π satisfies the hypotheses of this paper *iff* π satisfies the hypotheses of Theorem 2 of [15]. Moreover, at least one of A or A^{T} has a $k_1 \times k_2$ submatrix satisfying π *iff* G has an induced subgraph on k_1+k_2 nodes which satisfies π.

Therefore, given an instance of the node deletion problem (bipartite graph G and positive integer k), we can construct A and A^{T} in $O(mn)$ time. Then we can test both matrices for $k_1 \times k_2$ submatrices satisfying π for each k_1, k_2 such that $k_1+k_2=k$. $O(m+n)$ such tests are sufficient. If a $k_1 \times k_2$ $(k_1+k_2=k)$ submatrix is found which satisfies π, the G contains an induced subgraph on n nodes which satisfies π.

The same correspondence between (0, 1)-matrices and bipartite graphs shows that the submatrix problem is embedded in the node deletion problem.

A similar result holds when we look for column submatrices with a certain type of property. Let π be a property for (0, 1)-matrices that is non-trivial, holds for permutation matrices, and is hereditary.

The column submatrix problem for π is

Instance: An $m \times n$ (0, 1)-matrix A; positive integer $k \leqslant n$.
Question: Does A contain an $m \times k$ column submatrix that satisfies π?

Theorem 2. *The column submatrix problem for π is NP-hard.*

Proof. We modify the proof of Theorem 2 of Yannakakis [15] to establish this variation of his theorem:

Modified Theorem 2 of Yannakakis. *The node-deletion problem restricted to bipartite graphs for graph properties which are*
(i) *hereditary on induced subgraphs,*
(ii) *non-trivial on bipartite graphs,*
(iii) *satisfied by any independent set of edges,*
is NP-complete even when node deletions are confined to one specified side of the bipartite graph.

The modifications to Yannakakis' argument necessary to establish this variation are given in the Appendix.

Having established this, we argue as in the previous theorem. Note that confining node deletions to a specified side of the bipartite graph corresponds to considering only column submatrices.

An analogous theorem holds for the row submatrix problem.

Finally, a similar result holds for principal submatrices. Let π be a property for (0, 1)-matrices that is non-trivial, *holds for identity matrices*, and is hereditary on principal submatrices. Examples of such properties include *positive definite* and *positive semidefinite*, *copositive* and *strict copositive*, *P-matrix*, and *nondegenerate*. See [10] for discussion of these properties.

The principal submatrix problem for property π is

Instance: An $n \times n$ (0, 1)-matrix A; positive integer $k \leq n$.
Question: Does A contain a $k \times k$ principal submatrix with property π?

Theorem 3. *The principal submatrix problem is NP-hard.*

Proof. We show that the theorem holds even if A is restricted to be a (0, 1)-matrix with 1's along the main diagonal. If A is any such matrix it uniquely corresponds to the digraph which has directed edge (i, j) *iff* $a_{ij} = 1$, $i > j$, and has directed edge (j, i) *iff* $a_{ij} = 1$, $i < j$. Similarly any digraph (without self-loops) uniquely corresponds to such a matrix. As in Theorem 1 we say that property π holds for matrix A *iff* π holds for the unique digraph corresponding to A. Thus A has a $k \times k$ principal submatrix satisfying π *iff* its corresponding digraph has a node-induced subgraph satisfying π. But Theorem 7 of [9] establishes the NP-hardness of the node deletion problem for nontrivial digraph properties that are hereditary on induced subgraphs. Thus the principal submatrix problem is equivalent under polynomial-time transformation to the node deletion problem and so is NP-hard.

Table 2
Some properties to which Theorem 3 applies

	Property	Definition	Recognition algorithm
1.	*positive definite*	Murty [10]	[10]
2.	*co-positive*	[10]	?
3.	*is a P-matrix*	[10]	?
4.	*non-degenerate*	[10]	?
5.	*is an identity matrix*	[13]	!
6.	*of band width w* (for fixed w)	[13]	!
7.	*triangular*	[13]	!
8.	*diagonally dominant*	[13]	!
9.	*stable*	[13]	!
10.	*is a Z-matrix*	Fiedler and Ptak [5]	!
11.	*is a K-matrix*	[5]	?
12.	*symmetric*	13]	!

In Table 1 and 2 some interesting matrix properties are listed to which these results apply. References after the property indicate a polynomial time recognition algorithm, so that the submatrix problem for these properties is NP-complete. A '1' indicates that a polynomial time recognition algorithm may not appear in the literature but is straightforward. A question mark indicates that no polynomial time recognition algorithm is known to the author.

Since the NP-hardness of the submatrix problem for these properties is established for restricted inputs ((0, 1)-matrices, and, in case of Theorem 3, (0, 1)-matrices with 1's along the diag-

Table 1
Some properties to which Theorems 1, 2, 3 apply

	Property	Definition	Recognition algorithm
1.	*totally unimodular*	Padberg [11]	Seymour [12]
2.	*balanced*	[11]	?
3.	*perfect*	[11]	?
4.	*transformable to a network matrix*	Bixby and Cunningham [3]	[3]
5.	*transformable to a matching matrix*	Bartholdi and Ratliff [1]	Bartholdi and Orlin [2]
6.	*permutable to consecutive 1's*	Fulkerson and Gross [6]	[4,6]
7.	*permutable to circular 1's*	Tucker [14]	[14]
8.	*non-negative*	Strang [13]	!
9.	*of rank* $\leq r$	[13]	!

onal), the corresponding problems for arbitrary matrices over $\mathbb{R}$ must be NP-hard.

Appendix

These are the modifications to Theorem 2 of Yannakakis [15] to establish this variation of his theorem: the node deletion problem restricted to bipartite graphs for graph properties which are

(i) hereditary on induced subgraphs,

(ii) non-trivial on bipartite graphs, and

(iii) satisfied by any independent set of edges,

is NP-complete *even when node deletions are confined to one specified side of the bipartite graph.*

To modify the argument of Yannakakis, we consider bipartite graphs G with vertex partition (V_r, V_c) where all node deletions are confied to V_c. We begin as in [15] except that the cutpoint c_1 must be chosen in V_c.

Case 1. The argument proceeds exactly as in [15] except that we employ this special version of the dissociation number problem:

Instance: A bipartite graph with vertex partition (V_r, V_c) with V_c distinguished, and positive integer $k \leqslant |V_c|$.

Question: Can k or fewer vertices be deleted from V_c so that all remaining vertices of V_r are of degree 1 or 0?.

This is NP-complete since it is a special case of the set packing problem [7], which asks, for a collection C of finite sets and positive integer k, whether C contains at at least k mutually disjoint sets. To see this, construct a bipartite graph G by associating a vertex in V_c with each set in C, and a vertex in V_r with each element in $\cup C$. A vertex u in V_r is connected by an edge to a vertex v in V_c *iff* the set v contains the element u. Then C contains at least k mutually disjoint sets *iff* the dissociation number of G is no more than $|C| - k$.

Now the argument of [15], Case 1 proceeds as given, except that both the node-deletion problem and the dissociation number problem are understood to be restricted to the same specified side V_c of the bipartite graph.

Case 2. If J_0 is not a single edge, then J_0 has in its bipartition at least one more node d in the same set with $c(J_1)$, i.e. in V_c. Then G' is bipartite and the vertices in G' corresponding to vertices of G are all vertices in V_c. The argument of [15], Case 2 now proceeds unchanged except that all node deletions are confined to V_c.

These changes are sufficient to establish the modified theorem.

Acknowledgment

The authors thanks Loren K. Platzman and Mihalis Yannakakis for helpful comments.

References

[1] J.J. Bartholdi and H.D. Ratliff, "A field guide to recognizing network flow and matching problems", Technical Report, Department of Industrial and Systems Engineering, University of Florida, Gainesville (1977).

[2] J.J. Bartholdi and J.B. Orlin, "Recognition of matching problems", work in progress.

[3] R.R. Bixby and W.H. Cunningham, "Converting linear programs to network problems", *Math. Operations Res.* **5** (3), 321–357 (1980).

[4] K.S. Booth and G.S. Lueker, "Testing for the consecutive ones property, interval graphs, and graph planarity using *PQ*-tree algorithms", *J. Comput. System Sci.* **13**, 335–379 (1975).

[5] M. Fielder and V. Ptak, "On matrices with non-positive off-diagonal elements and positive principal minor", *Czech. Math. J.* **11**, 382–400 (1962).

[6] D.R. Fulkerson and D.A. Gross, "Incidence matrices and interval graphs", *Pacific J. Math.* **15**, 835–855 (1965).

[7] M.R. Garey and D.S. Johnson, *Computers and Intractability*, Freeman, San Francisco (1979).

[8] J.M. Lewis, "On the complexity of the maximum subgraph problem", *Proc. 10th Annual ACM Symposium on Theory of Computing* (1978) 265–274.

[9] J.M. Lewis and M. Yannakakis, "The node-deletion problem for hereditary properties is NP-complete" *J. Comput. System Sci.* **20**, 219–230 (1980).

[10] K.G. Murty, *Linear and Combinatorial Programming*, Wiley, New York (1976).

[11] M.W. Padberg, "Characterization of totally unimodular, balanced, and perfect matrices", in: B. Roy, Ed., *Combinatorial Programming: Methods and Applications*, D. Reidel, Dordrecht, Holland (1975) 278–284.

[12] P.D. Seymour, 'Decomposition of regular matroids", *J. Combinatorial Theory Ser. B* **28** (3), 305–359 (1980).

[13] G. Strang, *Linear Algebra and Its Applications*, 2nd ed., Academic Press, New York (1980).

[14] A. Tucker, "Matrix characterizations of circular arc graphs", *Pacific J. Math.* **39**, 535–545 (1971).

[15] M. Yannakakis, "Node-deletion problems on bipartite graphs", *Siam J. Comput.* **10** (2) (1981).