NORTH-HOLLAND

Diagonal Matrix Scaling is NP-Hard*

Leonid Khachiyan
Department of Computer Science
Rutgers University
New Brunswick, New Jersey 08903

Submitted by Richard A. Brualdi

ABSTRACT

A symmetric matrix A is said to be scalable if there exists a positive diagonal matrix X such that the row and column sums of XAX are all ones. We show that testing the scalability of arbitrary matrices is NP-hard. Equivalently, it is NP-hard to check for a given symmetric matrix A whether the logarithmic barrier function $\frac{1}{2}x^TAx - \sum \ln x_i$ has a stationary point in the positive orthant $x > 0$.

1. INTRODUCTION

An $n \times n$ symmetric matrix A is said to be *scalable* if there exists a positive diagonal matrix X such that the row and column sums of XAX are all ones:

$$XAXe = e, \qquad X = \operatorname{diag}\{x_1, \ldots, x_n\} > 0, \tag{S}$$

where e is the vector of all ones. Equivalently, the problem can be stated as the system of n nonlinear equations in n positive variables

$$Ax = x^{-1}, \qquad x = (x_1, \ldots, x_n) > 0.$$

*Research supported in part by the National Science Foundation under grant CS-9208371.

LINEAR ALGEBRA AND ITS APPLICATIONS 234:173–179 (1996)

655 Avenue of the Americas, New York, NY 10010
0024-3795/96/$15.00
SSDI 0024-3795(94)00099-Y

Here x^{-1} denotes the n-dimensional vector whose components are the reciprocals of the corresponding components of x. In particular, the scalability of A is equivalent to the existence of a stationary point of the logarithmic barrier function

$$f(x) = \tfrac{1}{2} x^T A x - \sum_{i=1}^{n} \ln x_i$$

in the positive orthant $x > 0$.

It is well known [1, 2, 6–8] that if all the entries of $A = (a_{ij})$ are nonnegative, then A is scalable if and only if there exists a doubly stochastic matrix $P = (p_{ij})$ with the same pattern ($p_{ij} > 0 \Leftrightarrow a_{ij} > 0$). Hence the scalability of *nonnegative* matrices can be formulated as a max-flow problem, and it can be tested in polynomial time.

Another important case of the scaling problem which is solvable in polynomial time is for *positive semidefinite* symmetric matrices: A is scalable if and only if $Ax = 0$, $e^T x = 1$ has no nonnegative solutions $x \geqslant 0$; see [4]. The latter condition can be checked by linear programming. In fact, the positive semidefinite case of the scaling problem provides a convenient formulation of linear programming in the context of interior point methods [5].

In this paper, we show that the general scaling problem (S) is NP-hard. Thus it is unlikely that similar good characterizations of scalability exist for arbitrary symmetric matrices.

THEOREM 1. *It is* NP*-hard to test the scalability of $n \times n$ symmetric matrices with integer entries not exceeding n in absolute value.*

We also show the NP-hardness of the problem of independent row and column matrix scaling; see Theorem 2 in Section 2.

Proof of Theorem 1. Let b and A be any given n-dimensional vector and $n \times n$ symmetric matrix, and consider the scaling problem

$$YCYe = e, \qquad Y \text{ a positive diagonal matrix} \tag{2.1}$$

for the $(n + 2) \times (n + 2)$ symmetric matrix

$$C = \begin{pmatrix} 0 & 1 & 0^T \\ 1 & 1 & b^T \\ 0 & b & A + bb^T \end{pmatrix}.$$

By the first equation of (2.1) the product of the first two components of Y is 1. So we can write $Y = \operatorname{diag}\{\tau^{-1}, \tau, x\}$, where τ is a positive scalar and x is a positive n-dimensional vector. The second of the scaling equations for C reads $\tau(\tau^{-1} + \tau + b^T x) = 1$, or equivalently $\tau + b^T x = 0$. The remaining n scaling equations are $X[\tau b + Ax + (b^T x)b] = e$, which, by $\tau + b^T x = 0$, can be written as $XAXe = e$. Since the value of $\tau > 0$ is arbitrary, we conclude that

$$C \text{ is scalable} \quad \Leftrightarrow \quad A \text{ has a scaling } X = \operatorname{diag}\{x\} > 0 \text{ such that } b^T x < 0.$$

Repeated application of the same construction leads us to the following result.

PROPOSITION 1. *Let A be an $n \times n$ symmetric matrix, and let $b_1, \ldots, b_m$ be a set of m vectors of dimension n. Define the $(n + 2m) \times (n + 2m)$ symmetric matrix C of the form*

$$C = \begin{pmatrix} 0 & 1 & & & & & & & 0^T \\ 1 & 1 & & & & & & & b_1^T \\ & & 0 & 1 & & & & & 0^T \\ & & 1 & 1 & & & & & b_2^T \\ & & & & \ddots & & & & \vdots \\ & & & & & & 0 & 1 & 0^T \\ & & & & & & 1 & 1 & b_m^T \\ 0 & b_1 & 0 & b_2 & \cdots & & 0 & b_m & A + \sum_{i=1}^{n} b_i b_i^T \end{pmatrix} \tag{2.2}$$

Then C is scalable if and only if there exists a scaling of A

$$XAXe = e, \qquad X = \operatorname{diag}\{x\} > 0, \tag{2.3}$$

satisfying the system of m strict linear inequalities

$$b_j^T x < 0, \qquad j = 1, \ldots, m. \tag{2.4}$$

We have thus shown that the scaling problem with linear constraints (2.3)–(2.4) can be polynomially transformed to the original scaling problem (S). In other words, strict homogeneous linear inequalities can be appended to the scaling problem without affecting its generality. To proceed, we shall construct a specific $n \times n$ symmetric matrix A having exponentially many discrete scalings X, and use them to represent the "upper half" of the $n - 2$-dimensional Boolean cube. Then, by adding linear constraints to the scaling problem for A, we shall obtain an NP-complete system of linear inequalities in Boolean variables.

PROPOSITION 2. *Let $n \geqslant 4$, and consider the $n \times n$ symmetric matrix*

$$A = \begin{pmatrix} 0 & 4 & 0^T \\ 4 & 8 - 8n & 4e^T \\ 0 & 4e & -I \end{pmatrix}, \tag{2.5}$$

where I is the identity matrix of order $n - 2$. A diagonal matrix X scales A if and only if

$$X = \frac{1}{\sqrt{s^2 - 1}} \operatorname{diag}\left\{ \frac{s^2 - 1}{2s}, \frac{s}{2}, s + \delta_1, \ldots, s + \delta_{n-2} \right\}, \tag{2.6}$$

where each of the $\delta_1, \ldots, \delta_{n-2}$ is either $+1$ or -1, and $s = \delta_1 + \cdots + \delta_{n-2} \geqslant 2$.

Proof. The first of the scaling equations for A implies that

$$X = \operatorname{diag}\left\{ \frac{\tau^{-1}}{2}, \frac{\tau}{2}, x \right\}, \tag{2.7}$$

where τ is a positive parameter, and where $x = (x_1, \ldots, x_{n-2}) > 0$. The second scaling equation yields

$$x_1 + \cdots + x_{n-2} = (n - 1)\tau. \tag{2.8}$$

The last $n-2$ scaling equations are $2\tau x_i - s_i^2 = 1$, that is,

$$x_i = \tau + \delta_i\sqrt{\tau^2 - 1}, \qquad \text{where} \quad \delta_i = \pm 1, \quad i = 1, \ldots, n-2. \tag{2.9}$$

Observe that for $\tau > 1$ both of the roots in (2.9) are positive. Substituting (2.9) into (2.8), we obtain

$$\tau(n-2) + (\delta_1 + \cdots + \delta_{n-2})\sqrt{\tau^2 - 1} = (n-1)\tau,$$

which can be written as $s\sqrt{\tau^2 - 1} = \tau$, using the notation $s = \delta_1 + \cdots + \delta_{n-2}$. Since τ is positive, s is positive as well. Hence,

$$\tau = \frac{s}{\sqrt{s^2 - 1}}. \tag{2.10}$$

The latter equality implies $s > 1$, and by integrality, $s \geq 2$. Finally, from (2.7), (2.10), and (2.9) we conclude that each of the scalings of A has the form (2.6), and vice versa. ■

Now consider the following problem: *Given m vectors $d_1, \ldots, d_m \in \{-1, 0, +1\}^{n-2}$, each of which has at most five nonzero components, check whether the system of $m+1$ linear equalities*

$$e^T\delta \geq 2, \qquad d_1^T\delta > 0, \ldots, \qquad d_m^T\delta > 0 \tag{P}$$

has a solution $\delta \in \{-1, +1\}^{n-2}$.

We can find in polynomial time a symmetric integral-valued matrix C of order $n + 2m$ such that

(i) C is scalable $\Leftrightarrow$ (P) is feasible;
(ii) the absolute value of any entry of C is $O(n+m)$.

We do this by letting

$$b_j^T = \left(0, 2e^Td_j, -d_j^T\right) \in Z^n, \qquad j = 1, \ldots, m, \tag{2.11}$$

and applying Propositions 1 and 2. Claim (i) follows from (2.4), (2.6), and (2.11). Claim (ii) is a direct consequence of (2.2), (2.5), (2.11), and the assumption that each d_j has at most five ± 1's.

Furthermore, by replacing C by $\begin{pmatrix} I & 0 \\ 0 & C \end{pmatrix}$, where I is the identity matrix of appropriate order, we can assume in (ii) that the absolute value of any entry of C does not exceed its order.

To complete the proof of Theorem 1, it remains to show the NP-completeness of (P), which can easily be shown as follows.

First, by replacing each d_j^T by $(0^T, d_j^T)$, where 0 is of dimension n, we can effectively remove the inequality $e^T\delta \geqslant 2$ from (P) at the expense of approximately doubling the number of the ± 1 unknowns. It is thus sufficient to show that it is NP-hard to test the feasibility of systems of linear inequalities $d_j^T\delta > 0$ in ± 1 unknowns (assuming, as before, that each inequality contains at most five nonzero coefficients ± 1). This latter fact easily follows from the following observation: the satisfiability of any 3-conjunctive normal form [3] $F(x_1, \ldots, x_n) = C_1 \wedge \cdots \wedge C_m$ is equivalent to the feasibility of the system $c_j^T\delta \geqslant -1$, $j = 1, \ldots, m$, where the ith unknown $\delta_i \in \{-1, +1\}$ in the system corresponds to the ith variable $x_i \in \{\text{false, true}\}$, and where each inequality of the system represents the corresponding clause of F and has three nonzero coefficients ± 1 (for example, $C = x_1 \vee \bar{x}_2 \vee x_3$ goes to $\delta_1 - \delta_2 + \delta_3 \geqslant -1$). Since each of the inequalities $c_j^T\delta \geqslant -1$ can be equivalently written as $c_j^T\delta + \alpha_j + \beta_j > 0$, where α_j and β_j are two additional "new" ± 1 variables for each $j = 1, \ldots, m$, Theorem 1 follows.

Theorem 1 can be extended to the problem of independent row and column scaling:

$$e^T X_1 A X_2 = e^T, \quad X_1 A X_2 e = e, \quad X_1 = \operatorname{diag}\{x_1\} > 0, \quad X_2 = \operatorname{diag}\{x_2\} > 0. \tag{2.12}$$

THEOREM 2. *It is* NP*-hard to test the solvability of* (2.12) *for* $n \times n$ *matrices* A *with integer entries not exceeding* n *in absolute value.*

Proof. Let A be an $n \times n$ matrix, and let C be defined by (2.2). The scalability of C by two positive diagonal matrices Y_1, Y_2 is equivalent to the scalability of A by two positive diagonal matrices X_1, X_2 such that

$$b_j^T x_1 < 0, \quad b_j^T x_2 < 0, \qquad j = 1, \ldots, m. \tag{2.13}$$

Hence, as before, we can append strict homogenious linear inequalities to the problem (2.12) without affecting its generality. It is also easily seen that any scaling factors X_1, X_2 of (2.5) have the form $X_1 = tX$, $X_2 = t^{-1}X$, where X is given by (2.6), and t is an arbitrary positive parameter. Since (2.13) is

invariant with respect to t, the rest of the proof is identical to that of Theorem 1.

As a final remark, note that (2.12) can be written as

$$\begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^{-1}, \qquad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} > 0,$$

which is a special case of (S). This means that Theorem 1 can be regarded as a corollary of Theorem 2.

The author thanks an anonymous referee for suggesting that Theorem 1 be extended to Theorem 2.

REFERENCES

1 R. A. Brualdi, S. V. Parter, and H. Schneider, The diagonal equivalence of a nonnegative matrix to a stochastic matrix, *J. Math. Ann. Appl.* 16:31–50 (1965).
2 D. Z. Djokovic, Note on nonnegative matrices, *Proc. Amer. Math. Soc.* 25:80–82 (1970).
3 M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, 1979.
4 B. Kalantari, Canonical problems for quadratic programming and projective methods for their solution, *Contemp. Math.* 114:243–263 (1990).
5 L. Khachiyan and B. Kalantari, Diagonal matrix scaling and linear programming, *SIAM. Optim.* 2:668–672 (1992).
6 M. London, On matrices with a doubly stochastic pattern, *J. Math. Anal. Appl.* 34:648–652 (1971).
7 H. Perfect and L. Mirsky, The distribution of positive elements in doubly stochastic matrices, *J. London Math. Soc.* 36:211–220 (1961).
8 R. Sinkhorn and P. Knopp, Concerning nonnegative matrices and doubly stochastic matrices, *Pacific J. Math.* 21:343–348 (1967).

Received 11 December 1992; final manuscript accepted 15 April 1994