Mathematical Programming 32 (1985) 1-10
North-Holland

# COMPUTING OPTIMAL SCALINGS BY PARAMETRIC NETWORK ALGORITHMS

J.B. ORLIN

*Massachusetts Institute of Technology, Sloan School of Management, Cambridge, MA 02139, USA*

U.G. ROTHBLUM

*Technion-Israel Institute of Technology, Faculty of Industrial Engineering and Management, Haifa 32000, Israel*

Received 14 March 1984
Revised manuscript received 5 November 1984

A *symmetric scaling* of a square matrix $A \neq 0$ is a matrix of the form $XAX^{-1}$ where $X$ is a nonnegative, nonsingular, diagonal matrix having the same dimension of $A$. An *asymmetric scaling* of a rectangular matrix $B \neq 0$ is a matrix of the form $XBY^{-1}$ where $X$ and $Y$ are nonnegative, nonsingular, diagonal matrices having appropriate dimensions. We consider two objectives in selecting a symmetric scaling of a given matrix. The first is to select a scaling $A'$ of a given matrix $A$ such that the maximal absolute value of the elements of $A'$ is lesser or equal that of any other corresponding scaling of $A$. The second is to select a scaling $B'$ of a given matrix $B$ such that the maximal absolute value of ratios of nonzero elements of $B'$ is lesser or equal that of any other corresponding scaling of $B$. We also consider the problem of finding an optimal asymmetric scaling under the maximal ratio criterion (the maximal element criterion is, of course, trivial in this case). We show that these problems can be converted to parametric network problems which can be solved by corresponding algorithms.

*Key words*: Scaling, Network Algorithms.

## 1. Introduction

A *symmetric scaling* of a square matrix $A \neq 0$ is a matrix of the form $XAX^{-1}$ where $X$ is a nonnegative, nonsingular, diagonal matrix having the same dimension of $A$. An *asymmetric scaling* of a rectangular matrix $B \neq 0$ is a matrix of the form $XBY^{-1}$ where $X$ and $Y$ are nonnegative, nonsingular, diagonal matrices having appropriate dimensions. There are many models in which a matrix can be replaced by any one of its scalings without changing the character of the problem. For example, an asymmetric scaling of the underlying matrix in linear programming corresponds to the change of units in which the variables are measured or the constraints are expressed (e.g., Fulkerson and Wolfe (1962)). Of course, in this case, the matrix being scaled consists of the coefficient matrix with an added column (the right hand side) and an added row (the cost vector). Also, a symmetric scaling of a (square) input-output matrix of a Leontief production model corresponds to the

This research was supported by NSF Grant ECS-83-10213.

change of units in which the activities are measured (e.g., Gale (1960) or Bachrach (1970)). Additional applications of scalings of matrices include telecommunication (e.g., Orchard and Hays (1968)), approximation theory (e.g., Golitschek (1982)). We note that in some applications a combination of symmetric and asymmetric scalings might be used, e.g., linear programming where groups of constraints and groups of variables might express the same physical commodity for which common measurement units are used.

A number of recent investigations studied the problem of finding and characterizing scalings of a matrix which are efficient in some way. One criterion used to measure the efficiency of a given scaling is the ratio of the largest absolute value of a nonzero element to the smallest. We refer to this criterion as the *min-max-ratio criterion.* It was first used by Fulkerson and Wolfe (1962) who developed an algorithm that approximates a corresponding best scaling. Alternative algorithms were obtained by Diliberto and Strauss (1951), Aumann (1958, 1959) and Golitschek (1980). One motivation for using this criterion is the fact that high ratios cause difficulties in executing the simplex method. Another criterion used to measure the efficiency of a scaling is the largest absolute value of an element. We refer to this criterion as the *min-max criterion.* In particular, Saunders and Schneider (1979) characterized matrices having scaling for which the corresponding measure is below a predetermined level (see also Golitschek, Rothblum and Schneider (1983)). Alternative approaches to measures of good scalings were used by Bauer (1963, 1969), Curtis and Reid (1972), Dantzig (1983) and Golitschek, Rothblum and Schneider (1983). Tomlin (1975) summarizes the issues of selecting criteria for good scalings by saying that:

> "The scaling of linear programming problems remains a rather poorly understood subject (as indeed it does for linear equations). Although many scaling techniques have been proposed, the rationale behind them is not always evident and very few numerical results are available."

The purpose of this paper is to show that the symmetric scaling problem, under the min-max criterion, and the asymmetric scaling problem, under the min-max-ratio criterion can be converted to parametric network optimization problems. Specifically, we show (in Sections 3 and 4, respectively) that the problems of determining best symmetric scalings under the min-max criterion and the best asymmetric scalings under the min-max-ratio criterion can be converted into a single parameter network optimization problem. Also, we show (in Section 5) that determining best symmetric scalings under the min-max-ratio criterion can be converted into a two parameter network optimization problem. Of course, the min-max criterion is meaningless for asymmetric scalings since for every rectangular matrix $B$, one can clearly select corresponding diagonal matrices $X$ and $Y$ for which the absolute values of the elements of $XBY^{-1}$ are (uniformly) arbitrarily small. We emphasize that our approach is elementary. Our proofs depend only on the Farkas Lemma and we obtain simple explicit programs for computing optimal scalings.

Our results imply that algorithms that solve parametric network problems can be used to compute good scalings. In particular, efficient algorithms for one-parameter network optimization problems have been obtained by Dantzig, Blattner and Rao (1967), Golitschek (1980, 1982), Karp (1978), Karp and Orlin (1981), Lawler (1967, 1976) and Megiddo (1979) (see also references in these papers). Golitschek and Schneider (1983) obtained an algorithm for computing symmetric scalings under the min-max-ratio criterion. Their algorithm essentially solves a two-parameter network optimization problem; but, they do not give a bound on the computational effort of their algorithm. Also, it is shown in Rothblum and Schneider (1980, Theorem 2), that the computation of an optimal scaling of an $n \times n$ matrix having $S$ nonzero elements, under the min-max-ratio criterion, corresponds to the solution of a linear program whose coefficient matrix has dimension $(2^n+n+1)(2^n+2S)$. The extreme points of the corresponding polytopes were characterized in Rothblum and Schneider (1980). Orlin and Rothblum (1984) obtain an algorithm for solving multi-parameter network optimization problems and provide explicit bounds on the computational complexity of these algorithms.

It was shown in Rothblum and Schneider (1980) that the logarithm of the optimal objective under the min-max-ratio criterion for asymmetric scaling equals the minimum average cycle in a corresponding graph. This property was used in Golitschek (1980) to compute that optimal value. Related algorithms were developed by Golitschek and Schneider (1983). Evidently, parametric network problems have been used earlier to solve minimum average cycle problems (e.g., Lawler (1976)).

## 2. Graph theoretic conventions

A *directed graph* $G$ is an ordered pair $(G_V, G_E)$, where $G_V = \{1, \dots, n\} \equiv \langle n \rangle$ for some positive integer $n$ and where $G_E$ is a (completely) ordered set. Elements of $G_V$ are called *vertices* and elements of $G_E$ are called *edges*. For every edge $e \in E$ we associate two vertices, called the *initial vertex* of $e$ and the *final vertex* of $e$, and we denote these vertices $s(e)$ and $t(e)$, respectively. If there exist no $i, j \in G_V$ for which there is more than a single edge $e$ having $s(e) = i$ and $t(e) = j$, we say that *$G$ has no multiple edges*. In this case, we identify an edge $e$ with the ordered pair $(s(e), t(e))$, and the order on $G_E$ is typically the lexicographic order on these pairs.

Let $G = (G_V, G_E)$ be a directed graph where $G_V = \langle n \rangle$ and $G_E = (e_1, \dots, e_S)$. The *node arc incidence matrix* of $G$, denoted $\Gamma(G)$, is the $n \times S$ matrix defined by

$$\Gamma(G)_{iq} = \begin{cases} 1 & \text{if } s(e_q) = i \text{ and } t(e_q) \neq i, \\ -1 & \text{if } s(e_q) \neq i \text{ and } t(e_q) = i, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

(In particular, $\Gamma(G)_{iq} = 0$ if $s(e_q) = t(e_q) = i$.) A *directed circulation* of $G$ is a vector $u \in \mathbb{R}^S$ for which $u \geqslant 0$ and $\Gamma(G)u = 0$. A *directed cycle* of $G$ is a nonzero directed circulation $u$ of $G$ for which $u_q \in \{0, 1\}$ for $q = 1, \ldots, S$ and for which the only circulation $z$ of $G$ with $\{q = 1, \ldots, S\colon v_q \neq 0\} \subset \{q = 1, \ldots, S\colon u_q \neq 0\}$ is the circulation $v = 0$. It is well known (see, Tutte (1956)) that every directed circulation is the sum of directed cycles.

A directed graph $H = (H_V, H_E)$ is called *$(m, n)$-bipartite* where $m$ and $n$ are positive integers, if

$$H_V = \langle m+n \rangle \quad \text{and} \quad H_E = \{(i, m+j)\colon i \in \langle m \rangle, j \in \langle n \rangle\} \cup \{(m+j, i)\colon i \in \langle m \rangle, j \in \langle n \rangle\},$$

where $H_E$ is ordered lexicographically. (There are no multiple arcs in such graphs.)

## 3. Symmetric scalings under the min–max criterion

Let $0 \neq A \in \mathbb{R}^{n \times n}$ be a given (square) matrix. The problem of finding an optimal symmetric scaling of $A$ under the min–max criterion can be formulated as the following optimization problem.

**Program 1**

$$\min \Lambda,$$

$$X_i |A_{ij}| X_j^{-1} \leq \Lambda \quad \text{whenever } A_{ij} \neq 0,$$

$$X_1, \ldots, X_n, \Lambda \text{ positive.}$$

For $i, j \in \langle n \rangle$ with $A_{ij} \neq 0$, let $a_{ij} = \lg |A_{ij}|$. We observe that program 1 can be converted to the following additive variant of that program by the change of variables $x_i = \lg X_i$, $i = 1, \ldots, n$, and $\lambda = \lg \Lambda$.

**Program 2**

$$\min \lambda,$$

$$x_i + a_{ij} - x_j \leq \lambda \quad \text{whenever } A_{ij} \neq 0,$$

$$x_1, \ldots, x_n, \lambda \text{ unrestricted.}$$

Denote the optimal value of Program 2 by $\lambda_*$. We observe that once $\lambda_*$ is determined, an optimal scaling of $A$ can be obtained from any solution to the system of inequalities $x_i + a_{ij} - x_j \leq \lambda_*$ for all $i, j \in \{1, \ldots, n\}$ with $A_{ij} \neq 0$. (The latter can be done, for example, by solving corresponding shortest path problems.)

In the following we shall find it convenient to consider the *graph associated with* the matrix $A$, denoted $G(A)$, which has $G(A)_V \equiv \langle n \rangle$ and $G(A)_E \equiv \{(i, j) \in \langle n \rangle \times \langle n \rangle\colon A_{ij} \neq 0\}$ where $G(A)_E$ is ordered lexicographically. Also, the *incidence matrix associated with* $A$, denoted $\Gamma(A)$, is defined by $\Gamma(A) \equiv \Gamma[G(A)]$.

Let $S$ be the number of nonzero elements of $A$. Then $|G(A)_E| = S$. Also, let $e \equiv (1, \ldots, 1)^T \in \mathbb{R}^S$ and let $a \in \mathbb{R}^S$ be the vector whose components are given by $a_q = a_{ij}$ for $q = 1, \ldots, S$, if the $q$-th edge of $G(A)_E$ (when ordered lexicographically) is $(i, j)$. The constraints determining feasibility for Program 2 can be written in matrix notation as

$$x^{\mathrm{T}} \Gamma(A) \leqslant -a^{\mathrm{T}} + \lambda e^{\mathrm{T}}. \tag{1}$$

By the Farkas Lemma (see Gale (1960, Theorem 4.6, p. 46)), for each $\lambda \in \mathbb{R}$, a vector $x \in \mathbb{R}^n$ satisfying (1) exists if and only if the following implication holds:

$$\Gamma(A) u = 0,\ u \geqslant 0 \quad \Rightarrow \quad (-a^{\mathrm{T}} + \lambda e^{\mathrm{T}}) u \geqslant 0, \tag{2}$$

i.e., $(-a^{\mathrm{T}} + \lambda e^{\mathrm{T}}) u \geqslant 0$ for every directed circulation $G(A)$. As each directed circulation of $G(A)$ is the sum of directed cycles of $G(A)$, (2) is clearly equivalent to the implication:

$$u \text{ is a directed cycle of } G(A) \Rightarrow (-a^{\mathrm{T}} + \lambda e^{\mathrm{T}}) u \geqslant 0. \tag{3}$$

(The equivalence of feasibility of Program 2, for a fixed $\lambda$, to (3) has been derived in Saunders and Schneider (1979) by different arguments.) We conclude that $\lambda_*$, the optimal value of Program 2, can be determined as the optimal value of the following program.

**Program 3**

$$\min \lambda,$$

$$(-a^{\mathrm{T}} + \lambda e^{\mathrm{T}}) u \geqslant 0 \quad \text{for every directed cycle } u \text{ of } G(A).$$

We have seen that $\lambda_*$ is the lowest value of the parameter $\lambda$ for which the graph $G(A)$ has no negative cycles under the parametric cost function which assigns cost $-a_{ij} + \lambda$ to arc $(i, j)$. This latter problem arises in the minimum ratio cycle problem, e.g., Dantzig, Blattner and Rao (1967), Lawler (1967), Lawler (1976) and Megiddo (1979) and corresponding efficient algorithms have been developed. In particular, Karp (1978) obtained an $O(Sn)$ algorithm, Megiddo (1979) obtained $O(S^2 n^2)$ and $O(Sn^2 \lg n)$ algorithms and Karp and Orlin (1981) obtained $O(n^3)$ and $O(Sn \lg n)$ algorithms. Also, Golitschek (1980, 1982) obtained an improved variant of Karp's algorithm.

We next comment about the general applicability of the above algorithms. We will use the standard graph theoretic terminology concerning connectedness of a graph and the partition of a graph into its strong components (e.g., Lawler (1976)). We remark that some of the above algorithms require special structure of the corresponding graphs. In particular, they all apply when the graph is connected. We note that when the graph is not connected, one can first identify the strong components of the graph, which can be done in $O(S + n)$ steps (e.g., Lawler (1976)), and then apply the algorithm to each strong component, independently. The optimal value of the corresponding program will then be determined as the least of the optimal values computed for the strong components.

## 4. Asymmetric scalings under the min–max-ratio criterion

Let $0 \neq B \in \mathbb{R}^{m\times n}$ be a given (rectangular) matrix. The problem of finding an optimal asymmetric scaling of $A$ under the min–max-ratio criterion can be formulated as the following optimization problem.

**Program 4**

$$\min \Pi/\Psi,$$

$$\Psi \leqslant X_i|B_{ij}|Y_j^{-1} \leqslant \Pi \quad \text{whenever } B_{ij} \neq 0,$$

$$X_1, \ldots, X_m,\ Y_1, \ldots,\ Y_n,\ \Psi,\ \Pi \text{ positive.}$$

We next observe that if $(X_1, \ldots, X_m, Y_1, \ldots, Y_n, \Psi, \Pi)$ is feasible for Program 4, then so is $(X_1/\Psi, \ldots, X_m/\Psi, Y_1, \ldots, Y_n, 1, \Pi/\Psi)$, and that the objective values of these two feasible solutions coincide, equaling $\Pi/\Psi$. We conclude that one can impose the requirement $\Psi = 1$ in Program 4, thereby obtaining the following simplification of that program.

**Program 4′**

$$\min \Pi,$$

$$1 \leqslant X_i|B_{ij}|Y_j^{-1} \leqslant \Pi \quad \text{whenever } B_{ij} \neq 0,$$

$$X_1, \ldots, X_m,\ Y_1, \ldots,\ Y_n,\ \Pi \text{ positive.}$$

For $i \in \langle n\rangle$ and $j \in \langle m\rangle$ with $B_{ij} \neq 0$, let $b_{ij} = \lg|B_{ij}|$. We observe that Program 4′ can be converted to the following additive variant of that program by the change of variables $x_i = \lg X_i$, $i = 1, \ldots, m$, $y_j = \lg Y_j$, $j = 1, \ldots, n$, and $\pi = \lg \Pi$.

**Program 5**

$$\min \pi$$

$$0 \leqslant x_i + b_{ij} - y_j \leqslant \pi \quad \text{whenever } B_{ij} \neq 0,$$

$$x_1, \ldots, x_m,\ y_1, \ldots, y_n,\ \pi \text{ unrestricted.}$$

Denote the optimal value of Program 5 by $\pi_*$. We observe that once $\pi_*$ is determined, an optimal scaling of $A$ can be obtained from any solution to the corresponding system of inequalities.

In the following we shall find it convenient to consider the $(m, n)$*-bipartite graph associated with* the matrix $B$, denoted $H(B)$, which has $H(B)_V \equiv \langle m+n \rangle$ and

$$H(B)_E = \{(i, m+j): i \in \langle m \rangle, j \in \langle n \rangle, B_{ij} \neq 0\} \cup \{m+j, i): i \in \langle m \rangle, j \in \langle n \rangle, B_{ij} \neq 0\}.$$

The $(m, n)$*-bipartite incidence matrix associated with* $B$, denoted $\Delta(B)$, is defined by $\Delta(B) \equiv \Gamma[H(B)]$.

Let $S$ be the number of nonzero elements of $B$. Then, $|H(B)_E| = 2S$. Also, let $c \in \mathbb{R}^{2S}$ be the vector whose components are given by $c_q = b_{ij}$ for $q = 1, \ldots, S$, if the $q$-th edge of $H(B)_E$ (when ordered lexicographically) is $(i, m+j)$, and $c_q = -b_{ij}$ for $q = S+1, \ldots, 2S$, if the $q$-th edge of $H(B)_E$ (when ordered lexicographically) is $(m+j, i)$. Also, let $f \equiv (1, \ldots, 1, 0, \ldots, 0)^{\mathrm{T}} \in \mathbb{R}^{2S}$ where $f$ has $S$ ones and $S$ zeros. The constraints determining feasibility for Program 5 can be represented in matrix notation as

$$(x^{\mathrm{T}}, y^{\mathrm{T}})\Delta(B) \leqslant -c^{\mathrm{T}} + \pi f^{\mathrm{T}}. \tag{4}$$

The arguments used in Section 3 show that for a given $\pi$, the existence of $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$ which satisfy (4) is equivalent to the implication

$$u \text{ a directed cycle of } H(B) \Rightarrow (-c^{\mathrm{T}} + \pi f^{\mathrm{T}})u \geqslant 0. \tag{5}$$

We conclude that $\pi_*$, the optimal value of Program 5, can be determined as the optimal value of the following program.

**Program 6**

$$\min \pi,$$
$$(-c^{\mathrm{T}} + \pi f^{\mathrm{T}})u \geqslant 0 \quad \text{for every directed cycle } u \text{ of } H(B).$$

We have seen that $\pi_*$ is the lowest value of the parameter $\pi$ for which the graph $H(B)$ has no negative cycles under the parametric cost function that assigns cost $-b_{ij} + \pi$ to edge $(i, m+j)$ and cost $b_{ij}$ to edge $(m+j, i)$. This problem seems to be more general than the parametric problem resulting from Program 3, as the parameter $\pi$ does not appear on all arcs. However, observing that half the edges of each cycle of $H$ are of the type to which $\pi$ is assigned, we conclude that, with $e = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^{m+n}$, the optimal value of program 6 coincides with that of the following problem.

**Program 7**

$$\min 2^{-1}\pi,$$
$$(-c^{\mathrm{T}} + \pi e^{\mathrm{T}})u \geqslant 0 \quad \text{for every directed cycle } u \text{ of } G(A).$$

Methods for solving Program 7 are discussed at the end of Section 3.

## 5. Symmetric scalings under the min–max-ratio criterion

Let $0 \neq \mathbb{R}^{n\times n}$ be a given (rectangular) matrix. The problem of finding an optimal symmetric scaling of $A$ under the min–max-ratio criterion can be formulated as the following optimization problem.

**Program 8**

$$\min \Pi / \Psi,$$

$$\Psi \leqslant X_i|A_{ij}|X_j^{-1} \leqslant \Pi \quad \text{whenever } A_{ij} \neq 0,$$

$$X_1, \ldots, X_n, \Psi, \Pi \text{ positive.}$$

For $i, j \in \langle n\rangle$ with $A_{ij} \neq 0$, let $a_{ij} = \lg|A_{ij}|$. As done in Sections 3 and 4, taking logarithms and using corresponding change of variables, one can convert Program 8 into the following additive variant of that program.

**Program 9**

$$\min \pi - \psi,$$

$$\psi \leqslant x_i + a_{ij} - x_j \leqslant \pi,$$

$$x_1, \ldots, x_n, \psi, \pi \text{ unrestricted.}$$

In the following we shall find it convenient to consider the *double graph associated with* $A$, denoted $F(A)$, which has $F(A)_V = \langle n\rangle$ and $F(A)_E$ consisting of the following two sets, ordered consecutively:

$$F(A)_E^1 \equiv \{(i,j) \in \langle n\rangle \times \langle n\rangle\colon A_{ij} \neq 0\} \quad \text{and} \quad F(A)_E^2 \equiv \{(j,i) \in \langle n\rangle \times \langle n\rangle\colon A_{ij} \neq 0\}$$

where $F(A)_E^1$ and $F(A)_E^2$ are each ordered lexicographically. Also, the *double incidence matrix associated with* $A$, denoted $\Omega(A)$, is defined by $\Omega(A) \equiv \Gamma[F(A)]$.

Let $S$ be the number of nonzero elements of $B$. Then $|F(A)_E| = 2S$. Also, let $c \in \mathbb{R}^{2S}$ be the vector whose components are given by $c_q = a_{ij}$ for $q = 1, \ldots, S$, if the $q$-th edge of $F(A)_E^1$ (when ordered lexicographically) is $(i,j)$, and $c_q = -a_{ij}$ for $q = S+1, \ldots, 2S$, if the $q$-th edge of $F(A)_E^2$ (when ordered lexicographically) is $(j,i)$. Also, let $f \equiv (1, \ldots, 1, 0, \ldots, 0)^{\mathrm{T}} \in \mathbb{R}^{2S}$ and $g = (0, \ldots, 0, 1, \ldots, 1)$ where both $f$ and $g$ have $S$ ones and $S$ zeros. The constraints determining feasibility of Program 9 can be represented in matrix notation as

$$x^{\mathrm{T}}\Omega(A) \leqslant -c^{\mathrm{T}} + \pi f^{\mathrm{T}} - \psi g^{\mathrm{T}}. \tag{6}$$

The arguments used in Section 3 show that for a given $\pi$ and $\psi$, the existence of $x \in \mathbb{R}^n$ which satisfies (6) is equivalent to the implication 7

$$u \text{ a directed cycle of } F(A) \Rightarrow (-c^{\mathrm{T}} + \pi f^{\mathrm{T}} - \psi g^{\mathrm{T}})u \geqslant 0. \tag{7}$$

We conclude that values $\pi_*$ and $\psi_*$ for which there exist $x_1, \ldots, x_n$ such that $x_1, \ldots, x_n, \pi_*, \psi_*$ are optimal for program 9 can be determined by the following program. (Of course, once $\pi_*$ and $\psi_*$ are determined, the corresponding values $x_1, \ldots, x_n$ are easily computable (see corresponding comments in Section 3).

**Program 10**

$$\min \pi - \psi,$$

$$(-c^{\mathrm{T}} + \pi f^{\mathrm{T}} - \psi g^{\mathrm{T}})u \geq 0 \text{ for every directed cycle } u \text{ of } F(A).$$

We have seen that the problem of finding an optimal symmetric scaling under the min-max-ratio criterion was converted to a problem of determining the lowest value of the difference of two parameters $\pi$ and $\psi$ for which the graph $F(A)$ has no negative cycles, under a corresponding parametric cost function (involving *both* parameters). Orlin and Rothblum (1984) developed algorithms for solving such multi-parameter network problems. Also, Golitschek and Schneider (1984) obtained a method for identifying optimal symmetric scalings under the min-max-ratio criterion which in fact can be used for solving Program 10 explicitly.

## References


G. Aumann, "Über approximative Nomo-graphie, I", *Bayerische Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche Klasse, Sitzungsberichte* (1958) 137–155.

G. Aumann, "Über approximative Nomo-graphie, II", *Bayerische Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche Klasse, Sitzungsberichte* (1959) 103–109.

M. Bacharach, *Biproportional matrices and input-output change* (Cambridge University Press, Cambridge, 1970).

F.L. Bauer, "Optimally scaled matrices", *Numerische Mathematik* 5 (1963) 73–87.

F.L. Bauer, "Remarks on optimally scaled matrices", *Numerische Mathematik* 13 (1969) 1–3.

A.R. Curtis and J.K. Reid, "On the automatic scaling of matrices for Gaussian elimination", *Journal of the Institute of Mathematics and Its Applications* 10 (1972) 118–124.

G.B. Dantzig, W. Blattner and M.R. Rao, "Finding a cycle in a graph with minimum cost to time ratio with application to a ship routing problem", in: P. Rosenstiehl, ed., *Theory of Graphs* (Dunod, Paris, and Gordon and Breach, NY, 1967) 77–84.

G.B. Dantzig, "The assignment problem for matrix scaling", Abstracts of talks presented at the ORSA/TIMS joint national meeting (1983) 78.

S.P. Diliberto and E.G. Straus, "On the approximation of a function of several variables by the sum of functions of fewer variables", *Pacific Journal of Mathematics* 1 (1951) 195–210.

D.R. Fulkerson and P. Wolfe, "An algorithm for scaling matrices", *SIAM Review* 4 (1962) 142–146.

D. Gale, *The theory of linear economic models* (McGraw-Hill, New York, 1960).

M. v. Golitschek, "An algorithm for scaling matrices and computing the minimum cycle mean in a diagraph", *Numerische Mathematik* 35 (1980) 45–55.

M. v. Golitschek, "Optimal cycles in doubly weighted graphs and approximation of bivariate functions by univariate ones", *Numerische Mathematik* 39 (1982) 65–84.

M. v. Golitschek, U.G. Rothblum and H. Schneider, "A conforming decomposition theorem, a piecewise linear theorem of the alternative, and scalings of matrices satisfying lower and upper bounds", *Mathematical Programming* 27 (1983) 291–306.

M. v. Golitschek and H. Schneider, "Applications of shortest path algorithms to matrix scalings", unpublished manuscript (1983).

R. Karp, "A characterization of the minimum cycle mean in a digraph", *Discrete Mathematics* 23 (1978) 309-311.

R. Karp and J.B. Orlin, "Parametric shortest path algorithms with an application to cycle staffing", *Discrete Applied Mathematics* 3 (1981) 37-45.

E.L. Lawler, "Optimal cycles in doubly weighted linear graphs", in: P. Rosenstiehl, ed., *Theory of graphs* (Dunod, Paris, and Gordon and Breach, NY, 1967) 209-214.

E.L. Lawler, *Combinatorial optimization: Networks and matroids* (Holt, Rinehart and Winston, NY, 1976).

N. Megiddo, "Combinatorial optimization with rational objective functions", *Mathematics of Operations Research* 4 (1979) 414-424.

W. Orchard-Hays, *Advanced linear programming techniques* (McGraw-Hill, New York, 1968).

J.B. Orlin and U.G. Rothblum, "Algorithms for multiparameter network flow problems", in preparation (1984).

U.G. Rothblum and H. Schneider, "Characterizations of optimal scalings of matrices", *Mathematical Programming* 19 (1980) 121-136.

B.D. Saunders and H. Schneider, "Cones, graphs and optimal scalings of matrices", *Linear and Multilinear Algebra* 8 (1979) 121-135.

J.A. Tomlin, "On scaling linear programming problems", *Mathematical Programming Study* 4 (1975) 146-166.

W.T. Tutte, "A class of Abelian groups", *Canadian Journal of Mathematics* 8 (1956) 13-28.