PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Optimally Scalable Matrices

T. I. Fenner and G. Loizou

Phil. Trans. R. Soc. Lond. A 1977 **287**, 307-349
doi: 10.1098/rsta.1977.0147

OPTIMALLY SCALABLE MATRICES

By T. I. FENNER and G. LOIZOU

Department of Computer Science, Birkbeck College, University of London, Malet Street, London WC1E 7HX

(*Communicated by J. H. Wilkinson, F.R.S. – Received* 18 *October* 1976)

CONTENTS

The characterization of matrices which can be optimally scaled with respect to various modes of scaling is studied. Particular attention is given to the following two problems:

(*a*) The characterization of those square matrices for which

$$\inf_{D} \operatorname{lub}(D^{-1}AD)$$

is attainable for some non-singular diagonal matrix D.

(*b*) The characterization of those square non-singular matrices A for which

$$\inf_{D_1,\, D_2} \operatorname{cond}_{12}(D_1AD_2)$$

is attainable for some non-singular diagonal matrices D_1 and D_2.

For norms having certain properties, various necessary and sufficient conditions for optimal scalability are obtained when, in problem (*a*), the matrix A and, in problem (*b*), both A and A^{-1} have chequerboard sign distribution. The characterizations so established impose various conditions on the combinatorial and spectral structure of the matrices. These are investigated by using results from the Perron–Frobenius theory of non-negative matrices and combinatorial matrix theory. It is shown that the Hölder or l_p-norms have the required properties, and that, in general, the only norms having all of the properties needed, for both the necessary and the sufficient conditions to be satisfied, are variants of the l_p-norms. For the special cases $p = 1$ and $p = \infty$, the characterizations obtained hold for all matrices, irrespective of sign distribution.

1. Introduction

Several types of matrix scaling are of interest in the solution of various numerical problems. These may be utilized in trying to minimize an appropriate condition number for the particular problem. The most common problems of this type are (Bauer 1968):

(I) $\psi_1(A) \overset{\Delta}{=} \inf_{D} \operatorname{lub}_{11}(D^{-1}AD)$, similarity scaling,

(II) $\chi_{12}(A) \overset{\Delta}{=} \inf_{D_2} \operatorname{cond}_{12}(AD_2)$

(III) $\chi'_{12}(A) \overset{\Delta}{=} \inf_{D_1} \operatorname{cond}_{12}(D_1 A)$ } one-sided scaling,

(IV) $\kappa_{12}(A) \overset{\Delta}{=} \inf_{D_1, D_2} \operatorname{cond}_{12}(D_1 A D_2)$, two-sided scaling,

where $\overset{\Delta}{=}$ stands for 'is defined as' and D, D_1, D_2 range over the set of non-singular diagonal matrices. As usual, $\operatorname{cond}_{12}(A)$ is the condition number of a non-singular real or complex matrix A with respect to two vector norms $\|\cdot\|_1$, $\|\cdot\|_2$, defined by (Bauer 1963):

$$\operatorname{cond}_{12}(A) = \operatorname{lub}_{12}(A)\operatorname{lub}_{21}(A^{-1}).$$

In fact $\operatorname{cond}_{12}(A)$ can be defined for rectangular A, having full column rank, as $\operatorname{lub}_{12}(A)/\operatorname{glb}_{12}(A)$, but if the columns of A are not independent $\operatorname{glb}_{12}(A) = 0$, so $\operatorname{cond}_{12}(A)$ is not defined. In problem (I), obviously, A must be square, but need not be non-singular.

If A is non-singular and Hermitian then under two-sided scaling this property is preserved if $D_1 = D_2^{\mathrm{H}}$. This gives rise to the problem

(V) $\kappa^*_{12}(A) \overset{\Delta}{=} \inf_{D} \operatorname{cond}_{12}(D^{\mathrm{H}}AD)$, symmetric scaling,

which, for practical purposes, is mainly of interest when A is positive definite.

There is some simplification in dealing with problems (II) to (V), if they are generalized by replacing A and A^{-1} by two arbitrary *rectangular* matrices B and C, respectively, where B and C^{T} have the same dimensions (Bauer 1963; McCarthy & Strang 1973), yielding

(IIA) $\chi_{12}(B; C) \overset{\Delta}{=} \inf_{D_2}\{\operatorname{lub}_{12}(BD_2)\operatorname{lub}_{21}(D_2^{-1}C)\}$,

(IIIA) $\chi'_{12}(B; C) \overset{\Delta}{=} \inf_{D_1}\{\operatorname{lub}_{12}(D_1B)\operatorname{lub}_{21}(CD_1^{-1})\}$,

(IVA) $\kappa_{12}(B; C) \overset{\Delta}{=} \inf_{D_1, D_2}\{\operatorname{lub}_{12}(D_1BD_2)\operatorname{lub}_{21}(D_2^{-1}CD_1^{-1})\}$,

(VA) $\kappa^*_{12}(B; C) \overset{\Delta}{=} \inf_{D}\{\operatorname{lub}_{12}(D^{\mathrm{H}}BD)\operatorname{lub}_{21}(D^{-1}C(D^{\mathrm{H}})^{-1})\}$, where in (VA) both B and C are Hermitian. [Note that $\chi'_{12}(B; C) = \chi_{21}(C; B)$, cf. Fenner & Loizou (1974).]

The various quantities on the left hand sides of (I) to (V) and (II A) to (V A) are called *minimal condition numbers* for the corresponding scaling problems.

In various papers the problem of finding, or finding bounds on, some of the minimal condition numbers has been studied extensively (Bauer 1963; Bauer 1969; Fenner & Loizou 1974; McCarthy & Strang 1973; van der Sluis 1969). In addition the problem of characterizing those matrices (or pairs of matrices) which are *optimally scaled* (best conditioned) in one of the above senses, i.e. those for which the corresponding condition number is minimal, has been investigated (Bauer 1963; Businger 1969; Fadeeva 1967; Forsythe & Straus 1955; Golub & Varah 1974; McCarthy & Strang 1973).

The purpose of this paper is to examine the related problem of *optimal scalability*, i.e. to characterize those matrices (or pairs of matrices) for which the infimum on the right hand side of the appropriate equation of (I) to (V), (II A) to (V A), is *attainable*. This problem, which has received relatively little attention in the literature, has been completely solved in cases (II) and (III) for an important class of norms (Businger 1968), but for cases (I), (IV) and (V) only partial results, under very special conditions, have been obtained (Businger 1968; Ström 1972). In this paper results for cases (I) and (IV) (and case (IV A)) are derived, and it is shown that the previously mentioned results are special cases of the results obtained here. In particular, when A, B and C have *chequerboard sign distribution* (Bauer 1963), the problem is completely solved in cases (I) and (IV A), for a certain class of norms. This also yields a solution for case (IV) if both A and A^{-1} have chequerboard sign distribution.

For cases (II) and (III) the results of Businger (1968) can be slightly generalized. It is there shown that any non-singular matrix A is optimally scalable in senses (II) and (III) with respect to a single *absolute* norm (Bauer, Stoer & Witzgall 1961), $\|.\|_1 = \|.\|_2$. It is readily seen that Businger's proof can be generalized to show that any non-singular matrix A is optimally scalable in sense (II) with respect to two norms $\|.\|_1$ and $\|.\|_2$, provided $\|.\|_2$ is absolute, but not necessarily $\|.\|_1$. Similarly, if $\|.\|_1$ is absolute, A is optimally scalable in sense (III).

Problems (I) to (V) are of importance in several numerical problems (Bauer 1963; Bauer 1966); in particular, (I) for eigenvalue problems (Bauer 1968; Fadeeva 1967; Osborne 1960); (II) and (III) for inclusion theorems (Bauer & Householder 1961; McCarthy & Strang 1973) and as a measure of the linear independence of the columns of A; (IV) and (V) in error analyses for the solution of systems of linear equations (Bauer 1966; Todd 1968).

In § 2 the various concepts, notations and properties used in the paper are defined, and several preliminary results involving these are then obtained in § 3. § 4 is devoted to various combinatorial results which, when utilized together with the conditions for optimal scalability subsequently established in § 5, indicate the combinatorial and spectral structure of those matrices which are optimally scalable. The applicability of the results of § 5 depends on the norms concerned satisfying certain properties, and a principal one of these properties is characterized in § 6. Finally, the paper ends with a discussion, in § 7, on the various conditions, both necessary and sufficient, for optimal scalability, and on the classes of norms for which these are valid.

2. Definitions

It will become apparent in the sequel that characterizations of optimally scalable matrices depend heavily on certain aspects of the combinatorial structure of the matrices involved. The concepts defined below will be useful in subsequently obtaining certain combinatorial results.

The matrices and vectors considered in this paper can, in general, be either real or complex. X_k will denote the Euclidean space of real or complex k-dimensional column vectors, i.e. $\mathbb{R}^k$ or $\mathbb{C}^k$. The X_k, for all k, will be either all real or all complex, and all matrices will be real or complex, correspondingly.

Let A be a square matrix, then $G(A)$ denotes the *digraph* (directed graph) associated with A (Fenner & Loizou 1971), i.e. the matrix $A^{\#}$, obtained from A by replacing all non-zero elements by unity, is the *adjacency matrix* of $G(A)$ (Wilson 1972). It is noted that $G(A)$ is *strongly connected* if and only if A is *irreducible* (Varga 1962), and $G(A)$ is (*weakly*) *connected* (Harary, Norman & Cartwright 1965) if and only if A is not *completely* (totally) *decomposable* (Harary 1962), i.e. there exists no permutation matrix P such that

$$P^{\mathrm{T}}AP = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix},$$

where A_{11} and A_{22} are *square* non-vacuous matrices. Note that a *strongly connected* digraph contains a directed path from any given vertex to any other given vertex; thus, if A is irreducible, for any given i, j, there exists a sequence $k_1, k_2, \ldots, k_s$ such that $a_{ik_1}, a_{k_1k_2}, \ldots, a_{k_{s-1}k_s}, a_{k_sj}$ are all non-zero.

Another related property is that of full indecomposability (Fenner & Loizou 1971; Fenner & Loizou 1977; Marcus & Minc 1963), namely

Definition 2.1. A square matrix A is *fully indecomposable* if there exist no permutation matrices P and Q such that

$$P^{\mathrm{T}}AQ = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \tag{1}$$

where A_{11} and A_{22} are *square* non-vacuous matrices.

A square matrix having the form of the right hand side of (1), where A_{11} and A_{22} are square and non-vacuous, is said to be in *reduced form*.

If A is rectangular, and non-square, then $G(A)$ is not defined, but, for *any* rectangular matrix A, $K(A)$ denotes the directed *bigraph* (bipartite graph) associated with A (Dulmage & Mendelsohn 1958; Fenner & Loizou 1971), i.e. the bipartite graph with vertex sets $X = \{x_i\}$ and $Y = \{y_j\}$ in which the arc $\overrightarrow{x_i y_j}$ is present if and only if $a_{ij} \neq 0$.

It is similarly noted that $K(A)$ is connected if and only if A is *chainable* (Hartfiel & Maxson 1975), i.e. there exist no permutation matrices P and Q such that

$$P^{\mathrm{T}}AQ = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}, \tag{2}$$

where at least one of the zero submatrices on the right hand side of (2) is non-vacuous. (Note that, in general, neither A_{11} nor A_{22} need be square.)

If A_{11} and A_{22} are non-vacuous, the matrix on the right hand side of (2) is called the *direct sum* of A_{11} and A_{22}, denoted by $A_{11} \oplus A_{22}$. This notation will also be used if A_{22} is vacuous, in which case if A_{11} is $k \times l$ and A_{22} is $(m-k) \times (n-l)$ (so either $k = m$ or $l = n$) $A_{11} \oplus A_{22}$ will denote the $m \times n$ matrix with leading $k \times l$ submatrix A_{11}, and all other elements zero. If A_{11} is vacuous, $A_{11} \oplus A_{22}$ is defined similarly.

A *connected component* (Harary, Norman & Cartwright 1965) of a digraph is a *maximal* connected subgraph of the digraph, and a *strong component* is a maximal *strongly* connected subgraph.

Two matrices B and C are *co-conformable* if the products BC and CB both exist, i.e. if B is $m \times n$ and C is $n \times m$, for some m, n.

Two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, having the same dimension, have *coinciding zeros* if, for each i, $x_i = 0$ if and only if $y_i = 0$.

It is well known (see theorem 2.7 of Varga (1962)) that any square non-negative matrix A has a non-negative eigenvalue equal to its *spectral radius*, $\rho(A)$, and, furthermore, that A possesses a non-negative right eigenvector $\boldsymbol{x}$ and a non-negative left eigenvector $\boldsymbol{y}^{\mathrm{T}}$, corresponding to $\rho(A)$. If A is *irreducible* then $\boldsymbol{x}$ and $\boldsymbol{y}^{\mathrm{T}}$, which are unique in this case, are the right and left *Perron eigenvectors* of A, respectively, and are positive. For arbitrary non-negative matrices this is generalized as follows:

Definition 2.2. A *right* (*left*) *P-vector* of a non-negative matrix A is a non-negative right (left) eigenvector of A corresponding to $\rho(A)$.

Note that if A is reducible its P-vectors are not necessarily unique.

Definition 2.3. A non-negative matrix A has *property* C if given any right P-vector $\boldsymbol{x}$ of A there exists a left P-vector $\boldsymbol{y}^{\mathrm{T}}$ of A such that $\boldsymbol{x}$ and $\boldsymbol{y}$ have coinciding zeros.

Definition 2.4. A non-negative matrix A has *property* C^{l} if given any right P-vector $\boldsymbol{x}$ of A there exists a left P-vector $\boldsymbol{y}^{\mathrm{T}}$ of A such that, for each i, $y_i = 0$ implies $x_i = 0$.

Definition 2.5. A non-negative matrix A has *property* C^{r} if given any left P-vector $\boldsymbol{y}^{\mathrm{T}}$ of A there exists a right P-vector $\boldsymbol{x}$ of A such that, for each i, $x_i = 0$ implies $y_i = 0$.

Obviously A has property C^{l} if and only if A^{T} has property C^{r}.

Throughout this paper P, Q, R, P_0, Q_0, R_0, etc., denote permutation matrices; I_n denotes the identity matrix of order n, this being shortened to just I if the order is apparent from context; $\boldsymbol{e}_j$ denotes the jth axis vector (i.e. the jth column of I), and $\boldsymbol{e}$ denotes a vector with all components equal to unity; $\mathscr{D}_n$ denotes the set of all $n \times n$ non-singular *non-negative* diagonal matrices, this similarly being abbreviated to just $\mathscr{D}$, if the order is evident.

For any vectors $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)^{\mathrm{T}}$, $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^{\mathrm{T}}$, and any $m \times n$ matrices $A = (a_{ij})$, $B = (b_{ij})$,

$|\boldsymbol{x}|$ denotes the vector $(|x_1|, |x_2|, \ldots, |x_m|)^{\mathrm{T}}$;

$\boldsymbol{x}^q$ denotes the vector $(x_1^q, x_2^q, \ldots, x_m^q)^{\mathrm{T}}$, for any $q > 0$;

$\boldsymbol{x} \oplus \boldsymbol{y}$ denotes the *direct sum* of $\boldsymbol{x}$ and $\boldsymbol{y}$, namely the vector $(\boldsymbol{x}^{\mathrm{T}} \boldsymbol{y}^{\mathrm{T}})^{\mathrm{T}} = (x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n)^{\mathrm{T}}$;

if $m = n$,

$$\left.\begin{array}{ll} \boldsymbol{x} \geqslant \boldsymbol{y} & \text{means } x_i \geqslant y_i \\ \boldsymbol{x} > \boldsymbol{y} & \text{means } x_i > y_i \end{array}\right\} \quad \text{for} \quad i = 1, 2, \ldots, n;$$

$$\left.\begin{array}{ll} A \geqslant B & \text{means } a_{ij} \geqslant b_{ij} \\ A > B & \text{means } a_{ij} > b_{ij} \end{array}\right\} \quad \text{for} \quad i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, n;$$

$|A|$ denotes the $m \times n$ matrix whose (i,j)th element is $|a_{ij}|$.

Definition 2.6. An $m \times n$ matrix A has a *chequerboard sign distribution* (Bauer 1963), hereinafter abbreviated to c.b.s.d., if there exist matrices E_1, E_2 with $|E_1| = I_m$, $|E_2| = I_n$, such that

$$A = E_1 |A| E_2. \tag{3}$$

As in Householder (1964) and Stoer (1964*a*), *all norms* are assumed to be *strictly homogeneous*. A norm $\|.\|$ is *absolute* if $\| |\boldsymbol{x}| \| = \|\boldsymbol{x}\|$ for every vector $\boldsymbol{x}$. It is *monotonic* if $|\boldsymbol{x}| \leqslant |\boldsymbol{y}|$ implies that $\|\boldsymbol{x}\| \leqslant \|\boldsymbol{y}\|$. A well-known result of Bauer *et al.* (1961) states that *a norm is absolute if and only if it is monotonic*.

Definition 2.7. A norm $\|.\|$ is *strongly monotonic* (cf. van der Sluis 1969) if $|\boldsymbol{x}| \leqslant |\boldsymbol{y}|$ and $|\boldsymbol{x}| \neq |\boldsymbol{y}|$ implies that $\|\boldsymbol{x}\| < \|\boldsymbol{y}\|$.

It is easily seen, by the continuity of a norm, that a strongly monotonic norm is monotonic.

The most commonly encountered norms are the l_p-norms (Hölder norms), $1 \leqslant p \leqslant \infty$, where, for $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$,

$$\|\boldsymbol{x}\| = l_p^n(\boldsymbol{x}) = \begin{cases} \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} & (1 \leqslant p < \infty), \\ \max_{1\leqslant i\leqslant n} |x_i| & (p = \infty). \end{cases}$$

Obviously these norms are absolute and, except for $p = \infty$, are strongly monotonic.

Any norm $\|.\|$ on a space of (column) vectors induces a norm on the space of dual (row) vectors; this norm, called the *dual norm* (Bauer *et al.* 1961), is defined by

$$\|\boldsymbol{y}^{\mathrm{H}}\|^{\mathrm{D}} = \sup_{\boldsymbol{x}\neq \boldsymbol{0}} \frac{|\boldsymbol{y}^{\mathrm{H}}\boldsymbol{x}|}{\|\boldsymbol{x}\|}. \tag{4}$$

It is well known (Bauer *et al.* 1961) that the dual of an absolute norm is itself absolute. Equation (4) gives rise to the concept of a *dual pair of vectors* (Bauer *et al.* 1961; Stoer 1964*a*).

Definition 2.8. A pair of *non-zero* vectors, $\boldsymbol{y}^{\mathrm{H}}$ and $\boldsymbol{x}$, are *dual* with respect to the norm $\|.\|$ if

$$\|\boldsymbol{y}^{\mathrm{H}}\|^{\mathrm{D}}\,\|\boldsymbol{x}\| = |\boldsymbol{y}^{\mathrm{H}}\boldsymbol{x}|. \tag{5}$$

This is written symbolically as $\boldsymbol{y}^{\mathrm{H}}\|\boldsymbol{x}$. It is thus seen that if $\boldsymbol{x}$ is dual to $\boldsymbol{y}^{\mathrm{H}}$ then $\boldsymbol{x}$ maximizes the right hand side of (4). The relation of duality is in fact quasi-symmetric, since

$$\|\boldsymbol{x}\| = \|\boldsymbol{x}\|^{\mathrm{DD}} = \sup_{\boldsymbol{y}^{\mathrm{H}}\neq \boldsymbol{0}} \frac{|\boldsymbol{y}^{\mathrm{H}}\boldsymbol{x}|}{\|\boldsymbol{y}^{\mathrm{H}}\|^{\mathrm{D}}}.$$

If $\|.\| = l_p^n(.)$, $1 \leqslant p \leqslant \infty$, its dual norm is given by

$$\|\boldsymbol{y}^{\mathrm{H}}\|^{\mathrm{D}} = l_q^n(\boldsymbol{y}), \quad \text{where} \quad 1/p + 1/q = 1. \tag{6}$$

Furthermore, $\boldsymbol{y}^{\mathrm{H}}\|\boldsymbol{x}$ if and only if

(i) there exists θ such that $\bar{y}_i x_i = \mathrm{e}^{\mathrm{i}\theta}|\bar{y}_i x_i|$, for all i, and

(ii) (*a*) for $1 < p < \infty$, there exists $\alpha > 0$ such that

$$|x_i|^p = \alpha|y_i|^q \quad \text{for all } i, \tag{7}$$

or (*b*) for $p = 1$,

$$|x_i|\,(\max_{1\leqslant k\leqslant n} |y_k| - |y_i|) = 0 \quad \text{for all } i, \tag{8}$$

or (*c*) for $p = \infty$,

$$|y_i|\,(\max_{1\leqslant k\leqslant n} |x_k| - |x_i|) = 0 \quad \text{for all } i. \tag{9}$$

If $\boldsymbol{x}, \boldsymbol{y} \geqslant \boldsymbol{0}$ then condition (i) is fulfilled automatically.

Throughout it will be assumed that the norms $\|.\|_1$ and $\|.\|_2$ are defined on X_m and X_n, respectively, $m, n \geqslant 2$. If $m = n$ then n will generally be used, and if $\|.\|_1 = \|.\|_2$ the norm may be simply denoted by $\|.\|$.

Given two vector norms $\|.\|_1$ and $\|.\|_2$, *the least upper bound norm* of an $m \times n$ matrix A, with respect to $\|.\|_1$ and $\|.\|_2$, is defined by any of the following characterizations:

$$\text{lub}_{12}(A) = \sup_{x \neq 0} \frac{\|Ax\|_1}{\|x\|_2} = \sup_{y^H \neq 0} \frac{\|y^H A\|_2^D}{\|y^H\|_1^D} = \sup_{x,\, y^H \neq 0} \frac{|y^H A x|}{\|y^H\|_1^D \|x\|_2}. \tag{10}$$

If $\|.\|_1 = \|.\|_2 = \|.\|$ then $\text{lub}_{11}(.)$ may be shortened to $\text{lub}(.)$, and similarly for $\text{cond}_{11}(.)$, $\kappa_{11}(.)$, etc.

Let U and V be $m \times m$ and $n \times n$ non-singular matrices, respectively. Then, given norms $\|.\|_1$ and $\|.\|_2$, define

$$\|x_1\|_{1U} = \|Ux_1\|_1 \quad \text{for} \quad x_1 \in X_m,$$
$$\|x_2\|_{2V} = \|Vx_2\|_2 \quad \text{for} \quad x_2 \in X_n.$$

It is easily shown that $\|.\|_{1U}$ and $\|.\|_{2V}$ are norms, and that the dual norms are given by

$$\|y_1^H\|_{1U}^D = \|y_1^H U^{-1}\|_1^D \quad \text{for} \quad y_1 \in X_m,$$

and similarly for $\|.\|_{2V}^D$. If $\|.\|_1 = \|.\|_2 = \|.\|$ then $\|.\|_{1U}$ is simply written as $\|.\|_U$, etc. Let A be an $m \times n$ matrix, and $\text{lub}_{1\,2}^{UV}(A)$ the lub norm of A with respect to the norms $\|.\|_{1U}$ and $\|.\|_{2V}$, then

$$\text{lub}_{1\,2}^{UV}(A) = \text{lub}_{12}(UAV^{-1}).$$

Definition 2.9. A norm $\|.\|$ is *symmetric* (Fenner & Loizou 1974) if, for all permutation matrices P and all vectors x,

$$\|x\|_P = \|x\|, \quad \text{i.e.} \quad \|Px\| = \|x\|.$$

It is easy to verify that the dual of a symmetric norm is also symmetric, and when $\|.\|_1$ and $\|.\|_2$ are both symmetric then, for any A and any permutation matrices P and Q,

$$\text{lub}_{1\,2}^{PQ}(A) = \text{lub}_{12}(A). \tag{11}$$

Definition 2.10. A norm $\|.\|$ is *subspace monotonic* if, for every P, and any $x = P(x_1 \oplus x_2) \in X_n$,

$$\|x_1 \oplus 0\|_P \leqslant \|x\|.$$

Obviously every monotonic (absolute) norm is subspace monotonic, and it is straightforward to show that the dual of a subspace monotonic norm is itself subspace monotonic. In addition, the following remark holds.

Remark 2.11. If the norm $\|.\|$ is subspace monotonic, and $x = P(x_1 \oplus 0)$ for some P, then for any $y^H = (y_1^H \oplus y_2^H)P^T$ that is dual to x, with x_1 and y_1 having the same dimension, the vector $(y_1^H \oplus 0)P^T$ is also dual to x.

The proof of this follows directly from the definition of duality and (4), by using the fact that $\|.\|^D$ is subspace monotonic.

Definition 2.12. A pair of norms $\|.\|_1$, $\|.\|_2$ has *property* L_{12} if, for any P and Q, and any $x = P(x_1 \oplus x_2) \in X_m$, $y = Q(y_1 \oplus y_2) \in X_n$,

$$\left.\begin{array}{l}\|x_1 \oplus 0\|_{1P} \leqslant \|y_1 \oplus 0\|_{2Q} \\ \text{and} \quad \|0 \oplus x_2\|_{1P} \leqslant \|0 \oplus y_2\|_{2Q}\end{array}\right\} \Rightarrow \|x\|_1 \leqslant \|y\|_2. \tag{12}$$

If the pair $\|.\|_1$, $\|.\|_2$ has property L_{12}, it is easy to see that the implication of (12) may be generalized to direct sums of more than two vectors, i.e. if

$$x = P(x_1 \oplus x_2 \oplus \ldots \oplus x_t), \quad y = Q(y_1 \oplus y_2 \oplus \ldots \oplus y_t),$$

and

$$\|\boldsymbol{x}_1 \oplus \mathbf{0} \oplus \mathbf{0} \ldots \oplus \mathbf{0}\|_{1P} \leqslant \|\boldsymbol{y}_1 \oplus \mathbf{0} \oplus \mathbf{0} \ldots \oplus \mathbf{0}\|_{2Q},$$
$$\|\mathbf{0} \oplus \boldsymbol{x}_2 \oplus \mathbf{0} \ldots \oplus \mathbf{0}\|_{1P} \leqslant \|\mathbf{0} \oplus \boldsymbol{y}_2 \oplus \mathbf{0} \ldots \oplus \mathbf{0}\|_{2Q},$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$
$$\|\mathbf{0} \oplus \mathbf{0} \oplus \ldots \mathbf{0} \oplus \boldsymbol{x}_t\|_{1P} \leqslant \|\mathbf{0} \oplus \mathbf{0} \oplus \ldots \mathbf{0} \oplus \boldsymbol{y}_t\|_{2Q},$$

then $\|\boldsymbol{x}\|_1 \leqslant \|\boldsymbol{y}\|_2$. This follows directly from the definition of property L_{12} since $\oplus$ is associative. It is immediately seen that the pair of norms $l_p^m(.)$, $l_p^n(.)$, $1 \leqslant p \leqslant \infty$, has property L_{12}.

Remark 2.13. If the pair of norms $\|.\|_1$, $\|.\|_2$ has property L_{12} then $\|.\|_2$ is subspace monotonic.

This follows directly from (12) on choosing $\boldsymbol{x}_1$ such that $\|\boldsymbol{x}_1 \oplus \mathbf{0}\|_1 = \|\boldsymbol{y}_1 \oplus \mathbf{0}\|_{2Q}$, $\boldsymbol{x}_2 = \mathbf{0}$ and $P = I$.

The pair of norms $\|.\|_1$, $\|.\|_2$ is said to have *property* L if both properties L_{12} and L_{21} hold; thus a single norm $\|.\|_1$ is said to have property L if property L_{11} holds. It is easy to see that a single norm having property L must be monotonic, and thus absolute.

Property L may be characterized directly as follows:

Remark 2.14. The pair of norms $\|.\|_1$, $\|.\|_2$ has property L if and only if (12) holds with equality throughout, where P, Q, $\boldsymbol{x}$ and $\boldsymbol{y}$ are as in definition 2.12.

Proof. The necessity of the condition follows immediately from property L. If (12) holds with equality throughout it is easy to show that $\|\boldsymbol{y}_1 \oplus (-\boldsymbol{y}_2)\|_{2Q} = \|\boldsymbol{y}_1 \oplus \boldsymbol{y}_2\|_{2Q}$ and thence that $\|.\|_2$ is subspace monotonic. By taking suitable multiples of $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ so that equality holds on the left hand side of (12), and by using the triangle inequality, it then follows that property L_{12} holds; similarly property L_{21} holds, which completes the proof.

(Cf. *Monotonic* and *absolute* norms on *direct sums* in Lancaster & Farahat (1972), in particular Theorem 1.)

A pair of vectors $\boldsymbol{y}^H$, $\boldsymbol{x}$ for which the supremum of $\dfrac{|\boldsymbol{y}^H A\boldsymbol{x}|}{\|\boldsymbol{y}^H\|_1^D \|\boldsymbol{x}\|_2}$ is attained is called a *maximizing pair* for A with respect to $\|.\|_1$ and $\|.\|_2$, or, more briefly, a maximizing pair for $\mathrm{lub}_{12}(A)$. For *absolute* norms any non-negative matrix A has a non-negative maximizing pair, and, furthermore, if $\boldsymbol{y}^H$, $\boldsymbol{x}$ is a maximizing pair then so is $|\boldsymbol{y}^H|$, $|\boldsymbol{x}|$.

It is easily seen from the definition of dual norms that, if $\boldsymbol{y}^H$, $\boldsymbol{x}$ is a maximizing pair for $\mathrm{lub}_{12}(A)$,

$$\mathrm{lub}_{12}(A)\|\boldsymbol{y}^H\|_1^D \|\boldsymbol{x}\|_2 = \|\boldsymbol{y}^H\|_1^D \|A\boldsymbol{x}\|_1 = \|\boldsymbol{y}^H A\|_2^D \|\boldsymbol{x}\|_2 = |\boldsymbol{y}^H A\boldsymbol{x}|. \tag{13}$$

Thus, for a maximizing pair,

$$\boldsymbol{y}^H\|_1 A\boldsymbol{x} \quad \text{and} \quad \boldsymbol{y}^H A\|_2 \boldsymbol{x}, \tag{14}$$

where $\|_1$ means 'dual with respect to $\|.\|_1$', etc. However, although these two dualities do not, in general, imply that $\boldsymbol{y}^H$, $\boldsymbol{x}$ is a maximizing pair, certain results of Bauer (1963) depend on this being true. For this reason Bauer defined a pair of norms having *property* S_{12} as follows:

Definition 2.15. A pair of norms $\|.\|_1$, $\|.\|_2$ has *property* S_{12} if, for every $m \times n$ matrix $A > 0$ and every pair of vectors $\boldsymbol{y}^T$, $\boldsymbol{x} \geqslant \mathbf{0}$,

$$\boldsymbol{y}^T \|_1 A\boldsymbol{x} \quad \text{and} \quad \boldsymbol{y}^T A \|_2 \boldsymbol{x} \quad \Rightarrow \quad \boldsymbol{y}^T, \boldsymbol{x} \text{ is a maximizing pair for } \mathrm{lub}_{12}(A). \tag{15}$$

Although Bauer only defined property S_{12} for $m = n$, it will be useful in the sequel to consider the above generalization to the rectangular case, which is clearly still a well-defined concept. Stoer & Witzgall (1962) have shown that property S_{12} holds for Hölder norms $\|.\|_1 = \|.\|_2 = l_p^n(.)$,

$1 < p < \infty$ (see footnote on page 78 of Bauer (1963)), and it is not difficult to see that their proof is still valid for $\|.\|_1 = l_p^m(.)$, $\|.\|_2 = l_p^n(.)$, $1 < p < \infty$, with $m \neq n$. For $p = 1$ or ∞ this will be proved in § 3.

A slightly weaker property than property S_{12} is defined by:

DEFINITION 2.16. A pair of norms $\|.\|_1$, $\|.\|_2$ has *property* S'_{12} if, for every $m \times n$ matrix $A > 0$ and every pair of vectors $\boldsymbol{y}^T$, $\boldsymbol{x} > \boldsymbol{0}$, the implication (15) holds.

It is noted that lemma I (i) of Bauer (1963) is still valid under the assumption that the norms have properties S'_{12} and S'_{21}, since the only use made of (15) in the proof of this result is when $\boldsymbol{y}^T$ and $\boldsymbol{x}$ are both strictly positive.

In this paper certain results of Stoer & Witzgall (1962) and Bauer (1963), which were proved for strictly positive matrices, are extended to certain classes of non-negative and c.b.s.d. matrices, and in some cases to arbitrary matrices. It is therefore necessary to define new properties as follows:

DEFINITION 2.17. A pair of norms $\|.\|_1$, $\|.\|_2$ has property $\bar{S}_{12}$ if, for every $m \times n$ matrix $A \geqslant 0$ and every pair of vectors $\boldsymbol{y}^T$, $\boldsymbol{x} > \boldsymbol{0}$, the implication (15) holds.

DEFINITION 2.18. A pair of norms $\|.\|_1$, $\|.\|_2$ has *property* S^*_{12} if, for every $m \times n$ matrix $A \geqslant 0$, where $A = P(A_1 \oplus 0)\, Q^T$ with A_1 being $k \times l$, $1 \leqslant k \leqslant m$, $1 \leqslant l \leqslant n$, and every pair of positive vectors $\boldsymbol{y}_1 \in X_k$, $\boldsymbol{x}_1 \in X_l$, the implication (15) holds with $\boldsymbol{y}^T = (\boldsymbol{y}_1 \oplus \boldsymbol{0})^T P^T$ and $\boldsymbol{x} = Q(\boldsymbol{x}_1 \oplus \boldsymbol{0})$.

Obviously property S^*_{12} implies property $\bar{S}_{12}$. Furthermore, if the norms are subspace monotonic, by generalizing theorem 2.11 of Gries & Stoer (1967), it can be shown that property S^*_{12} holds for the pair $\|.\|_1$, $\|.\|_2$ if and only if property $\bar{S}_{12}$ holds for every pair of norms *induced* on a pair of *coordinate subspaces* of X_m and X_n by $\|.\|_1$ and $\|.\|_2$, respectively. [See Gries (1967) and Gries & Stoer (1967) for definitions of *induced* norms and *coordinate* subspaces, and also Fiedler & Pták (1960).]

Although properties $\bar{S}_{12}$ and S^*_{12} are more restrictive than property S'_{12}, it will be shown in § 3 that, for $\|.\|_1 = l_p^m(.)$, $\|.\|_2 = l_p^n(.)$, $1 \leqslant p \leqslant \infty$, these properties also hold.

From the definition of property S_{12} it is easy to show:

REMARK 2.19. If the pair of norms $\|.\|_1$, $\|.\|_2$ has property S_{12}, then the pair $\|.\|_{1U}$, $\|.\|_{2V}$ also has property S_{12} if $U = P_1 D_1$, $V = P_2 D_2$ for some permutation matrices P_1, P_2 and some $D_1 \in \mathscr{D}_m$, $D_2 \in \mathscr{D}_n$. A similar result holds for properties S'_{12}, $\bar{S}_{12}$ and S^*_{12}.

The pair of norms $\|.\|_1$, $\|.\|_2$ is said to have *property* S if both properties S_{12} and S_{21} hold; consequently a single norm is said to have property S if property S_{11} holds; similarly for properties S′, $\bar{S}$ and S*.

DEFINITION 2.20. A pair of norms $\|.\|_1$, $\|.\|_2$ is *lub-absolute* if, for every pair of co-conformable matrices B and C,

$$\text{lub}_{12}(|B|) = \text{lub}_{12}(B), \quad \text{lub}_{21}(|C|) = \text{lub}_{21}(C).$$

A single norm $\|.\|$ is said to be lub-absolute if $\text{lub}(|A|) = \text{lub}(A)$, for every A. It can be shown that the pair $\|.\|_1 = l_{p_1}^m(.)$, $\|.\|_2 = l_{p_2}^n(.)$ is lub-absolute if and only if $p_1 = p_2 = 1$ or $p_1 = p_2 = \infty$. It is also true that if the pair $\|.\|_1$, $\|.\|_2$ is lub-absolute then both of these norms must be absolute. This may be shown by considering matrices of rank unity (Fenner 1977).

3. Preliminary results

In this section various preliminary results used subsequently are stated or derived.

3.1. *Minimal condition numbers*

If $\|.\|_1 = \|.\|_2$ then

$$\operatorname{lub}(A) \geqslant \rho(A), \tag{16}$$

which, from (I) (see Introduction), implies that

$$\psi(A) \geqslant \rho(A). \tag{17}$$

Also, for arbitrary $\|.\|_1$, $\|.\|_2$, from (IVA), by using the multiplicativity of lub norms,

$$\kappa_{12}(B;\, C) \geqslant \rho(BC). \tag{18}$$

Given two *absolute* norms, $\|.\|_1$ and $\|.\|_2$, for any two matrices A and G such that $A \geqslant |G|$, it is easy to show (Bauer 1963) that

$$\operatorname{lub}_{12}(A) \geqslant \operatorname{lub}_{12}(G). \tag{19}$$

Now, for *absolute* norms, by virtue of lemma III of Bauer (1963), if D_1 and D_2 are non-singular diagonal matrices such that D_1AD_2 exists, then

$$\operatorname{lub}_{12}(D_1AD_2) = \operatorname{lub}_{12}(|D_1|\,A\,|D_2|). \tag{20}$$

Therefore, for *absolute* norms, there is no loss of generality in assuming that D, D_1, D_2 are non-negative in (I) to (V) and (IIA) to (VA), so it may be assumed that all scaling matrices D, D_1, D_2, etc., are non-negative, i.e. are in $\mathscr{D}$.

Under this assumption, for *absolute* $\|.\|_1 = \|.\|_2$, it follows from (19) that, if $A \geqslant |G|$,

$$\psi(A) \geqslant \psi(G). \tag{21}$$

Similarly, for two *absolute* norms $\|.\|_1$ and $\|.\|_2$, if $B \geqslant |F|$ and $C \geqslant |H|$,

$$\kappa_{12}(B;\, C) \geqslant \kappa_{12}(F;\, H). \tag{22}$$

Suppose now that A has c.b.s.d., then from (3) and (20)

$$\operatorname{lub}_{12}(A) = \operatorname{lub}_{12}(|A|). \tag{23}$$

Thus, for *absolute* $\|.\|_1 = \|.\|_2$, and A having c.b.s.d., it follows from (23) that

$$\psi(A) = \psi(|A|). \tag{24}$$

Similarly, for two *absolute* norms $\|.\|_1$ and $\|.\|_2$, and B and C both having c.b.s.d.,

$$\kappa_{12}(B;\, C) = \kappa_{12}(|B|;\, |C|). \tag{25}$$

Obviously (24) and (25) also hold for non-c.b.s.d. matrices if the pair of norms is lub-absolute.

3.2. lub *norms of direct sums*

The following two lemmas are concerned with pairs of norms having property L_{12} or L.

Lemma 3.1. If the pair $\|.\|_1$, $\|.\|_2$ has property L_{12}, then the pair $\|.\|_{1U}$, $\|.\|_{2V}$ also has property L_{12} if $U = R_1D_1$ and $V = R_2D_2$ for some permutation matrices R_1, R_2 and some $D_1 \in \mathscr{D}_m$, $D_2 \in \mathscr{D}_n$.

Proof. Consider, for arbitrary P and Q, any two vectors,

$$\boldsymbol{x} = P(\boldsymbol{x}_1 \oplus \boldsymbol{x}_2) \in \mathrm{X}_m \quad \text{and} \quad \boldsymbol{y} = Q(\boldsymbol{y}_1 \oplus \boldsymbol{y}_2) \in \mathrm{X}_n,$$

satisfying

$$\|\boldsymbol{x}_1 \oplus \boldsymbol{0}\|_{1UP} \leqslant \|\boldsymbol{y}_1 \oplus \boldsymbol{0}\|_{2VQ}, \quad \|\boldsymbol{0} \oplus \boldsymbol{x}_2\|_{1UP} \leqslant \|\boldsymbol{0} \oplus \boldsymbol{y}_2\|_{2VQ}. \tag{26}$$

Now let $P^{\mathrm{T}}D_1P = D_1^{(1)} \oplus D_1^{(2)}$, $Q^{\mathrm{T}}D_2Q = D_2^{(1)} \oplus D_2^{(2)}$, where $D_1^{(1)}$ and $D_2^{(1)}$ are square and conformable with $\boldsymbol{x}_1$ and $\boldsymbol{y}_1$, respectively. Then (26) may be written as

$$\|D_1^{(1)}\boldsymbol{x}_1 \oplus \boldsymbol{0}\|_{1R_1P} \leqslant \|D_2^{(1)}\boldsymbol{y}_1 \oplus \boldsymbol{0}\|_{2R_2Q}, \quad \|\boldsymbol{0} \oplus D_1^{(2)}\boldsymbol{x}_2\|_{1R_1P} \leqslant \|\boldsymbol{0} \oplus D_2^{(2)}\boldsymbol{y}_2\|_{2R_2Q}.$$

On using property L_{12} for the pair $\|.\|_1$, $\|.\|_2$, this now implies that $\|R_1D_1\boldsymbol{x}\|_1 \leqslant \|R_2D_2\boldsymbol{y}\|_2$, yielding the result.

Lemma 3.2. If the pair $\|.\|_1$, $\|.\|_2$ has property L then $\|.\|_1$ and $\|.\|_2$ both have property L.

Proof. Consider, for arbitrary P and Q, any two vectors, $\boldsymbol{x} = P(\boldsymbol{x}_1 \oplus \boldsymbol{x}_2)$, $\boldsymbol{x}' = Q(\boldsymbol{x}_1' \oplus \boldsymbol{x}_2') \in \mathrm{X}_m$ (where $\boldsymbol{x}_1$ and $\boldsymbol{x}_1'$ are not necessarily of the same dimension), which satisfy

$$\|\boldsymbol{x}_1 \oplus \boldsymbol{0}\|_{1P} = \|\boldsymbol{x}_1' \oplus \boldsymbol{0}\|_{1Q}, \quad \|\boldsymbol{0} \oplus \boldsymbol{x}_2\|_{1P} = \|\boldsymbol{0} \oplus \boldsymbol{x}_2'\|_{1Q}.$$

Now, for any k, $1 \leqslant k \leqslant n-1$, let $\boldsymbol{y}_1 \in \mathrm{X}_k$, $\boldsymbol{y}_2 \in \mathrm{X}_{n-k}$ be chosen such that

$$\|\boldsymbol{y}_1 \oplus \boldsymbol{0}\|_2 = \|\boldsymbol{x}_1 \oplus \boldsymbol{0}\|_{1P} = \|\boldsymbol{x}_1' \oplus \boldsymbol{0}\|_{1Q},$$
$$\|\boldsymbol{0} \oplus \boldsymbol{y}_2\|_2 = \|\boldsymbol{0} \oplus \boldsymbol{x}_2\|_{1P} = \|\boldsymbol{0} \oplus \boldsymbol{x}_2'\|_{1Q}.$$

On using property L, by remark 2.14, these equalities imply that $\|\boldsymbol{y}_1 \oplus \boldsymbol{y}_2\|_2 = \|\boldsymbol{x}\|_1 = \|\boldsymbol{x}'\|_1$. Hence, by remark 2.14, $\|.\|_1$ has property L, and it follows similarly that $\|.\|_2$ also has property L, which completes the proof.

Various results on lub norms of direct sums and other partitioned matrices are now established.

Lemma 3.3. Let $\|.\|_1$ and $\|.\|_2$ be a pair of norms, with $\|.\|_2$ subspace monotonic, and for some P and Q let

$$\|\boldsymbol{y}_1\|_3 = \|\boldsymbol{y}_1 \oplus \boldsymbol{0}\|_{1P} \quad \text{for} \quad \boldsymbol{y}_1 \in \mathrm{X}_k,$$
$$\|\boldsymbol{x}_1\|_4 = \|\boldsymbol{x}_1 \oplus \boldsymbol{0}\|_{2Q} \quad \text{for} \quad \boldsymbol{x}_1 \in \mathrm{X}_l,$$

where $1 \leqslant k \leqslant m$ and $1 \leqslant l \leqslant n$. Then, if $A = P(A_1 \oplus 0)\,Q^{\mathrm{T}}$ is an $m \times n$ matrix with A_1 being $k \times l$, $\mathrm{lub}_{12}(A) = \mathrm{lub}_{34}(A_1)$.

Proof. From (10),

$$\mathrm{lub}_{12}(A) = \sup_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\|A\boldsymbol{x}\|_1}{\|\boldsymbol{x}\|_2} = \sup_{\substack{\boldsymbol{x}_1,\,\boldsymbol{x}_2 \\ \boldsymbol{x}_1 \oplus \boldsymbol{x}_2 \neq \boldsymbol{0}}} \frac{\|P(A_1\boldsymbol{x}_1 \oplus \boldsymbol{0})\|_1}{\|Q(\boldsymbol{x}_1 \oplus \boldsymbol{x}_2)\|_2}$$
$$= \sup_{\boldsymbol{x}_1 \neq \boldsymbol{0}} \frac{\|A_1\boldsymbol{x}_1 \oplus \boldsymbol{0}\|_{1P}}{\|\boldsymbol{x}_1 \oplus \boldsymbol{0}\|_{2Q}} = \sup_{\boldsymbol{x}_1 \neq \boldsymbol{0}} \frac{\|A_1\boldsymbol{x}_1\|_3}{\|\boldsymbol{x}_1\|_4} = \mathrm{lub}_{34}(A_1),$$

since $\|.\|_2$ is subspace monotonic. This completes the proof.

By considering the dual characterization of $\mathrm{lub}_{12}(A)$ it is seen that lemma 3.3 also holds if $\|.\|_1$ is subspace monotonic instead of $\|.\|_2$.

Lemma 3.4. For any subspace monotonic norms $\|.\|_1$ and $\|.\|_2$, if

$$A = \begin{bmatrix} A_1 & Z \\ W & A_2 \end{bmatrix}$$

then $\mathrm{lub}_{12}(A) \geqslant \max\{\mathrm{lub}_{12}(A_1 \oplus 0), \mathrm{lub}_{12}(0 \oplus A_2)\}$.

(Note that, given the dimensions of A_1, the dimensions of the – possibly vacuous – zero matrix in the expression $\mathrm{lub}_{12}(A_1 \oplus 0)$ are implicit from the fact that $\|.\|_1$ and $\|.\|_2$ are defined on X_m and X_n, respectively; similarly for $\mathrm{lub}_{12}(0 \oplus A_2)$.)

Proof. Let A_1 be $k \times l$, then for all non-zero $\boldsymbol{x}_1 \in \mathrm{X}_l$,

$$\mathrm{lub}_{12}(A) \geqslant \frac{\|A(\boldsymbol{x}_1 \oplus \mathbf{0})\|_1}{\|\boldsymbol{x}_1 \oplus \mathbf{0}\|_2} = \frac{\|A_1\boldsymbol{x}_1 \oplus W\boldsymbol{x}_1\|_1}{\|\boldsymbol{x}_1 \oplus \mathbf{0}\|_2} \geqslant \frac{\|A_1\boldsymbol{x}_1 \oplus \mathbf{0}\|_1}{\|\boldsymbol{x}_1 \oplus \mathbf{0}\|_2},$$

since $\|.\|_1$ is subspace monotonic. For all $\boldsymbol{x}_2 \in \mathrm{X}_{n-l}$, it then follows that

$$\mathrm{lub}_{12}(A) \geqslant \frac{\|A_1\boldsymbol{x}_1 \oplus \mathbf{0}\|_1}{\|\boldsymbol{x}_1 \oplus \boldsymbol{x}_2\|_2} = \frac{\|(A_1 \oplus 0)\,(\boldsymbol{x}_1 \oplus \boldsymbol{x}_2)\|_1}{\|\boldsymbol{x}_1 \oplus \boldsymbol{x}_2\|_2},$$

since $\|.\|_2$ is also subspace monotonic. Therefore $\mathrm{lub}_{12}(A) \geqslant \mathrm{lub}_{12}(A_1 \oplus 0)$, and the other inequality follows similarly. This completes the proof.

Note that if A_1 is vacuous ($k = 0$ or $l = 0$), or A_2 is vacuous ($k = m$ or $l = n$), this result is still valid by a slight modification of the proof. Note also that if W and Z are both zero (possibly vacuously) the requirement that $\|.\|_1$ be subspace monotonic may be dropped. Alternatively, by considering the dual characterization of $\mathrm{lub}_{12}(A)$, it follows that if W and Z are both zero the requirement that $\|.\|_2$ be subspace monotonic could be dropped, provided $\|.\|_1$ is subspace monotonic, since the dual of a subspace monotonic norm is subspace monotonic.

Corollary 3.5. For any norms $\|.\|_1$ and $\|.\|_2$, with $\|.\|_1$ or $\|.\|_2$ subspace monotonic,

$$\mathrm{lub}_{12}(A_1 \oplus A_2 \oplus \ldots \oplus A_t) \geqslant \max_{1 \leqslant j \leqslant t}\{\mathrm{lub}_{12}(0 \oplus \ldots 0 \oplus A_j \oplus 0 \ldots \oplus 0)\}.$$

Equality can be achieved in corollary 3.5 for pairs of norms having property L_{12}. This follows from the following result.

Lemma 3.6. For a pair of norms $\|.\|$, $\|.\|_2$ having property L_{12},

$$\mathrm{lub}_{12}(A_1 \oplus A_2) \leqslant \max\{\mathrm{lub}_{12}(A_1 \oplus 0), \mathrm{lub}_{12}(0 \oplus A_2)\}.$$

Proof. Let A_1 be $k \times l$, and let $\alpha = \max\{\mathrm{lub}_{12}(A_1 \oplus 0), \mathrm{lub}_{12}(0 \oplus A_2)\}$. Then, for any $\boldsymbol{x}_1 \in \mathrm{X}_l$,

$$\|A_1\boldsymbol{x}_1 \oplus \mathbf{0}\|_1 \leqslant \mathrm{lub}_{12}(A_1 \oplus 0)\|\boldsymbol{x}_1 \oplus \mathbf{0}\|_2 \leqslant \|\alpha\boldsymbol{x}_1 \oplus \mathbf{0}\|_2. \tag{27}$$

Similarly, for any $\boldsymbol{x}_2 \in \mathrm{X}_{n-l}$, $\|\mathbf{0} \oplus A_2\boldsymbol{x}_2\|_1 \leqslant \|\mathbf{0} \oplus \alpha\boldsymbol{x}_2\|_2$, which with (27) yields

$$\|(A_1 \oplus A_2)\,(\boldsymbol{x}_1 \oplus \boldsymbol{x}_2)\|_1 \leqslant \|\alpha\boldsymbol{x}_1 \oplus \alpha\boldsymbol{x}_2\|_2 = \alpha\|\boldsymbol{x}_1 \oplus \boldsymbol{x}_2\|_2$$

by property L_{12}. Therefore $\mathrm{lub}_{12}(A_1 \oplus A_2) \leqslant \alpha$, which completes the proof.

Note that the result holds trivially if A_1 or A_2 is vacuous.

Corollary 3.7. For a pair of norms $\|.\|_1$, $\|.\|_2$ having property L_{12},

$$\mathrm{lub}_{12}(A_1 \oplus A_2 \oplus \ldots \oplus A_t) = \max_{1 \leqslant j \leqslant t}\{\mathrm{lub}_{12}(0 \oplus \ldots 0 \oplus A_j \oplus 0 \ldots \oplus 0)\}.$$

Proof. This follows directly from corollary 3.5 and lemma 3.6, on using remark 2.13.

Lemma 3.8. For a pair of norms $\|.\|_1$, $\|.\|_2$ having property L, given a pair of co-conformable matrices $B = B_1 \oplus B_2$, $C = C_1 \oplus C_2$, where B_1 and C_1 are non-zero and also co-conformable, there exist positive constants, λ and μ, such that

$$\mathrm{lub}_{12}(B_1 \oplus \lambda\mu^{-1}B_2)\,\mathrm{lub}_{21}(C_1 \oplus \lambda^{-1}\mu C_2) = \max\{\mathrm{lub}_{12}(B_1 \oplus 0)\,\mathrm{lub}_{21}(C_1 \oplus 0), \mathrm{lub}_{12}(0 \oplus B_2)\,\mathrm{lub}_{21}(0 \oplus C_2)\}$$

Proof. By corollary 3.7. for *any* positive constants, λ and μ,

$$\text{lub}_{12}(B_1 \oplus \lambda\mu^{-1}B_2)\,\text{lub}_{21}(C_1 \oplus \lambda^{-1}\mu C_2)$$
$$= \max\{\text{lub}_{12}(B_1 \oplus 0), \lambda\mu^{-1}\text{lub}_{12}(0 \oplus B_2)\} \times \max\{\text{lub}_{21}(C_1 \oplus 0), \lambda^{-1}\mu\,\text{lub}_{21}(0 \oplus C_2)\}.$$

This yields the required result if λ and μ are defined as follows:

$$\lambda = \text{lub}_{12}(B_1 \oplus 0), \quad \mu = \text{lub}_{12}(0 \oplus B_2), \quad \text{if} \quad B_2 \neq 0,$$
$$\lambda = \text{lub}_{21}(0 \oplus C_2), \quad \mu = \text{lub}_{21}(C_1 \oplus 0), \quad \text{if} \quad C_2 \neq 0,$$
$$\lambda = \mu = 1, \quad \text{if} \quad B_2 = 0, \quad C_2 = 0.$$

Corollary 3.9. With the same conditions as in lemma 3.8, there exists $D \in \mathscr{D}_m$ such that

$$\text{lub}_{12}(DB)\,\text{lub}_{21}(CD^{-1}) = \max\{\text{lub}_{12}(B_1 \oplus 0)\,\text{lub}_{21}(C_1 \oplus 0), \text{lub}_{12}(0 \oplus B_2)\,\text{lub}_{21}(0 \oplus C_2)\}.$$

Lemma 3.10. Let $\|.\|_1 = l_1^m(.)$, $\|.\|_2 = l_1^n(.)$, and let

$$B = \begin{bmatrix} B_1 & U_1 \\ 0 & B_2 \end{bmatrix}, \quad C = \begin{bmatrix} C_1 & V_1 \\ 0 & C_2 \end{bmatrix} \tag{28}$$

be $m \times n$, $n \times m$ matrices, respectively, where B_1, C_1 are $k \times l$, $l \times k$, respectively, $1 \leqslant k \leqslant m$, $1 \leqslant l \leqslant n$, and

$$\text{lub}_{12}(B_1 \oplus 0)\,\text{lub}_{21}(C_1 \oplus 0) > \text{lub}_{12}(0 \oplus B_2)\,\text{lub}_{21}(0 \oplus C_2). \tag{29}$$

Then there exist $D_1 \in \mathscr{D}_m$, $D_2 \in \mathscr{D}_n$ such that

$$\text{lub}_{12}(D_1 B D_2)\,\text{lub}_{21}(D_2^{-1} C D_1^{-1}) = \text{lub}_{12}(B_1 \oplus 0)\,\text{lub}_{21}(C_1 \oplus 0). \tag{30}$$

Proof. Let $D_1 = I_k \oplus \lambda I_{m-k}$, $D_2 = I_l \oplus \mu^{-1} I_{n-l}$, where $\lambda, \mu > 0$. Now, since

$$\text{lub}_{12}(B) = \max_j\,(\boldsymbol{e}^{\mathrm{T}}|B|)_j,$$

$$\text{lub}_{12}(B) = \max\{\text{lub}_{12}(B_1 \oplus 0), \text{lub}_{12}(U^* + (0 \oplus B_2))\} \tag{31}$$

where $U^* = B - (B_1 \oplus B_2)$. Therefore, by (31),

$$\text{lub}_{12}(D_1 B D_2) = \text{lub}_{12}(B_1 \oplus 0) \tag{32}$$

if

$$\text{lub}_{12}(B_1 \oplus 0) \geqslant \text{lub}_{12}(\mu^{-1}U^* + \mu^{-1}\lambda(0 \oplus B_2)).$$

This inequality certainly holds if

$$\mu\,\text{lub}_{12}(B_1 \oplus 0) \geqslant \text{lub}_{12}(U^*) + \lambda\,\text{lub}_{12}(0 \oplus B_2). \tag{33}$$

Similarly,

$$\text{lub}_{21}(D_2^{-1} C D_1^{-1}) = \text{lub}_{21}(C_1 \oplus 0) \tag{34}$$

if

$$\lambda\,\text{lub}_{21}(C_1 \oplus 0) \geqslant \text{lub}_{21}(V^*) + \mu\,\text{lub}_{21}(0 \oplus C_2), \tag{35}$$

where $V^* = C - (C_1 \oplus C_2)$. Now let

$$\beta_1 = \text{lub}_{12}(B_1 \oplus 0), \quad \beta_2 = \text{lub}_{12}(0 \oplus B_2), \quad \sigma = \text{lub}_{12}(U^*),$$
$$\gamma_1 = \text{lub}_{21}(C_1 \oplus 0), \quad \gamma_2 = \text{lub}_{21}(0 \oplus C_2), \quad \tau = \text{lub}_{21}(V^*).$$

It is easily verified that the set

$$\{\langle\lambda, \mu\rangle \mid \lambda, \mu > 0 \quad \text{and} \quad \mu\beta_1 \geqslant \sigma + \lambda\beta_2 \quad \text{and} \quad \lambda\gamma_1 \geqslant \tau + \mu\gamma_2\}$$

is non-empty by virtue of (29). Thus there exist $\lambda, \mu > 0$ satisfying (33) and (35), so the result (30) follows immediately from (32) and (34). This completes the proof.

A corresponding result for the $l_\infty(.)$ norms holds for matrices having the form of the transposes of those in (28).

3.3. *The* S *properties*

In this subsection it is shown that, as mentioned at the end of § 2, properties $\bar{S}_{12}$ and S^*_{12} hold for appropriate pairs of l_p-norms.

LEMMA 3.11. Let A be a non-negative $m \times n$ matrix with $K(A)$ connected, and let $\boldsymbol{r}$ and $\boldsymbol{s}$ be non-negative m and n vectors, respectively. If the zeros of $A^{\mathrm{T}}\boldsymbol{r}$ and $\boldsymbol{s}$ coincide, and also the zeros of $A\boldsymbol{s}$ and $\boldsymbol{r}$, then $\boldsymbol{r}$ and $\boldsymbol{s}$ are either both zero or both strictly positive.

Proof. If $\boldsymbol{r} = \boldsymbol{0}$ then clearly $\boldsymbol{s} = \boldsymbol{0}$, and conversely. So assume that both are non-zero.

If $\boldsymbol{r} > \boldsymbol{0}$ then, as $K(A)$ is connected, $A^{\mathrm{T}}\boldsymbol{r} > \boldsymbol{0}$ and thus $\boldsymbol{s} > \boldsymbol{0}$. Conversely, if $\boldsymbol{s} > \boldsymbol{0}$ then $\boldsymbol{r} > \boldsymbol{0}$, so assume that $\boldsymbol{r}, \boldsymbol{s} \not> \boldsymbol{0}$. Now let P and Q be such that

$$\boldsymbol{r}^{\mathrm{T}}P = [\hat{\boldsymbol{r}}^{\mathrm{T}}\ \boldsymbol{0}], \quad Q^{\mathrm{T}}\boldsymbol{s} = \begin{bmatrix}\hat{\boldsymbol{s}}\\ \boldsymbol{0}\end{bmatrix}, \tag{36}$$

where $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{s}}$ are positive k and l vectors, respectively, with $0 < k < m$, $0 < l < n$. Also, let

$$P^{\mathrm{T}}AQ = \begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}, \tag{37}$$

where A_{11} is $k \times l$. Using (36) and (37), it then follows from the coincidence of the zeros of $A^{\mathrm{T}}\boldsymbol{r}$ and $\boldsymbol{s}$ that $\hat{\boldsymbol{r}}^{\mathrm{T}}A_{12} = \boldsymbol{0}$, which implies that $A_{12} = 0$. Similarly, the coincidence of the zeros of $A\boldsymbol{s}$ and $\boldsymbol{r}$ yields $A_{21} = 0$, which contradicts the connectedness of $K(A)$.

COROLLARY 3.12. Let A be a non-negative $m \times n$ matrix with $K(A)$ connected, and let $\|.\|_1$ and $\|.\|_2$ be norms with the property that any dual pair of vectors has coinciding zeros. Then every pair of non-negative vectors, $\boldsymbol{y}^{\mathrm{T}}$ and $\boldsymbol{x}$, satisfying

$$\boldsymbol{y}^{\mathrm{T}} \parallel_1 A\boldsymbol{x} \quad \text{and} \quad \boldsymbol{y}^{\mathrm{T}}A \parallel_2 \boldsymbol{x}, \tag{38}$$

is strictly positive.

Proof. The zeros of $A^{\mathrm{T}}\boldsymbol{y}$ and $\boldsymbol{x}$ coincide, as do those of $\boldsymbol{y}$ and $A\boldsymbol{x}$, so the result follows by the lemma, as $\boldsymbol{x}$ and $\boldsymbol{y}$ cannot be zero.

It can be shown that norms which are both strongly monotonic and have strongly monotonic duals satisfy the conditions of corollary 3.12. In particular, corollary 3.12 holds for $\|.\|_1 = l^m_{p_1}(.)$, $\|.\|_2 = l^n_{p_2}(.)$, $1 < p_1, p_2 < \infty$.

COROLLARY 3.13. Let A be a non-negative *square* matrix with $G(A)$ connected, and $\boldsymbol{r}, \boldsymbol{s}$ non-negative vectors for which the zeros of $\boldsymbol{r}$, $\boldsymbol{s}$, $A^{\mathrm{T}}\boldsymbol{r}$ and $A\boldsymbol{s}$ all coincide. Then $\boldsymbol{r}$ and $\boldsymbol{s}$ are either both zero or both strictly positive.

Proof. Since the zeros of $\boldsymbol{r}$ and $\boldsymbol{s}$ coincide, it follows that $Q = P$ in (36), which yields the result.

LEMMA 3.14. Let A be a non-negative matrix with $K(A)$ connected, then AA^{T} is irreducible.

Proof. Suppose that AA^{T} is reducible, then there exists P such that $P^{\mathrm{T}}AA^{\mathrm{T}}P = F_1 \oplus F_2$, where F_1 and F_2 are square and non-vacuous. Since $K(A)$ is connected A has no zero rows or columns, so, by lemma 4.5 (proved in the next section), there exists Q such that $P^{\mathrm{T}}AQ = A_1 \oplus A_2$ with A_1 and A_2 non-vacuous. This contradicts the connectedness of $K(A)$, so AA^{T} must be irreducible.

This result can also be proved using theorem 2.1 of Hartfiel & Maxson (1975).

Lemma 3.15. Let A be a non-negative $m \times n$ matrix with $K(A)$ connected, and let $\|.\|_1$ and $\|.\|_2$ be *absolute* norms with the property that any dual pair of vectors has coinciding zeros. Then, up to constant multiples, every maximizing pair of vectors for $\mathrm{lub}_{12}(A)$ is strictly positive.

Proof. Let w^H, z be a maximizing pair for $\mathrm{lub}_{12}(A)$, then

$$\mathrm{lub}_{12}(A)\|w^H\|_1^D\|z\|_2 = |w^H Az| \leqslant |w|^T A|z|. \tag{39}$$

Thus, as the norms are absolute, it follows that equality must hold in (39), and that $|w|^T$, $|z|$ is also a maximizing pair for $\mathrm{lub}_{12}(A)$. Now, by (13), $y^T = |w|^T$ and $x = |z|$ satisfy (38), so by corollary 3.12 $|w|$ and $|z|$ are positive. Consequently, since equality holds in (39),

$$\arg(\bar{w}_i z_j) = \nu, \quad \forall i,j \quad \text{such that} \quad a_{ij} > 0, \tag{40}$$

where ν is a constant. As $K(A)$ is connected, it follows from lemma 3.14 that AA^T is irreducible, so for any i, k there exists a sequence of positive elements of A:

$$a_{ij_1}, a_{i_1 j_1}, a_{i_1 j_2}, a_{i_2 j_2}, a_{i_2 j_3}, \ldots, a_{i_{t-1} j_t}, a_{kj_t}.$$

Thus, by (40), $\arg(w_i) = \arg(w_k)$, so w is a constant multiple of a positive vector, and similarly so is z.

The following lemma is a generalization of part of the proof of theorem 2 of Stoer & Witzgall (1962).

Lemma 3.16. If $\|.\|_1 = l_p^m(.)$ and $\|.\|_2 = l_p^n(.)$, $1 < p < \infty$, then, for any non-negative $m \times n$ matrix A with $K(A)$ connected, there exists, up to positive multiples, a unique pair of positive vectors y^T, x satisfying (38).

Proof. The norms $\|.\|_1$ and $\|.\|_2$ satisfy the conditions of lemma 3.15, so A has a positive maximizing pair y^T, x which, by (13), satisfies (38). By using (7) this implies that, for some $\alpha, \beta > 0$,

$$Ax = \alpha y^{q/p}, \quad A^T y = \beta x^{p/q}, \tag{41}$$

exponentiation of vectors being elementwise. The replacement of x and y by some multiples of themselves yields (41) with two new constants, $\hat{\alpha}$ and $\hat{\beta}$, which satisfy

$$\hat{\alpha}^{1/q}\hat{\beta}^{1/p} = \alpha^{1/q}\beta^{1/p} = \lambda^{1/q},$$

where $\lambda^{1/q} > 0$ is a constant independent of the multiples chosen. Now replace x by a suitable multiple so that $\hat{\alpha} = \lambda$, $\hat{\beta} = 1$, i.e.

$$Ax = \lambda y^{q/p}, \quad A^T y = x^{p/q}. \tag{42}$$

Scaling x and y simultaneously, it is seen that, for any $\gamma > 0$, $\gamma^q x$ and $\gamma^p y$ also satisfy (42).

Now suppose that (42) has two solutions, $x > \mathbf{0}$, $y > \mathbf{0}$, $\lambda > 0$ and $\bar{x} > \mathbf{0}$, $\bar{y} > \mathbf{0}$, $\bar{\lambda} > 0$, where $\lambda \leqslant \bar{\lambda}$. Scaling x and y with

$$\gamma = \left(\frac{\bar{\lambda}}{\lambda}\right)^{1/q} \times \max_i \left(\frac{\bar{y}_i}{y_i}\right)^{1/p}$$

the new x and y satisfy

$$A(x - \bar{x}) = \lambda y^{q/p} - \bar{\lambda}\bar{y}^{q/p} \geqslant \mathbf{0}, \tag{43}$$

with equality for at least one component. Since $\lambda \leqslant \bar{\lambda}$, it follows from (43) that $y \geqslant \bar{y}$, with $y > \bar{y}$ if $\lambda < \bar{\lambda}$, and thus, from (42), that $x \geqslant \bar{x}$. If $y > \bar{y}$ then, as $K(A)$ is connected, $A^T(y - \bar{y}) > \mathbf{0}$ and thus, by (42), $x > \bar{x}$. Similarly, if $x > \bar{x}$ then $A(x - \bar{x}) > \mathbf{0}$ which contradicts (43). Therefore $x \ngtr \bar{x}$, $y \ngtr \bar{y}$, so $\lambda = \bar{\lambda}$.

Now, letting $\boldsymbol{r} = \boldsymbol{y}-\bar{\boldsymbol{y}}$, $\boldsymbol{s} = \boldsymbol{x}-\bar{\boldsymbol{x}}$, the conditions of lemma 3.11 apply, by (42), so $\boldsymbol{r}$ and $\boldsymbol{s}$ must both be zero. Thus the solution of (42) is unique up to positive multiples.

Corollary 3.17. Let $\|.\|_1$, $\|.\|_2$ and A be defined as in the lemma. Then, if $\boldsymbol{y}^{\mathrm{T}}$ and $\boldsymbol{x}$ are positive and satisfy (38), they are a maximizing pair for $\mathrm{lub}_{12}(A)$.

Proof. This follows directly from lemma 3.16 on using lemma 3.15 and (13).

Before obtaining property $\bar{\mathrm{S}}_{12}$ for the l_p-norms for general p, properties S_{12} and $\bar{\mathrm{S}}_{12}$ are established for $p = 1$ and ∞.

Lemma 3.18. Property S_{12} holds for $\|.\|_1 = l^m_\infty(.)$, $\|.\|_2 = l^n_\infty(.)$.

Proof. Let A be any positive $m\times n$ matrix, and $\boldsymbol{y}^{\mathrm{T}}$, $\boldsymbol{x}$ any pair of non-zero non-negative vectors satisfying (38). Then, from (9), for all i,

$$y_i(A\boldsymbol{x})_i = y_i \max_k (A\boldsymbol{x})_k, \tag{44}$$

$$(\boldsymbol{y}^{\mathrm{T}}A)_i x_i = (\boldsymbol{y}^{\mathrm{T}}A)_i \max_k x_k. \tag{45}$$

Now, since $A > 0$, $\boldsymbol{y}^{\mathrm{T}}A > \mathbf{0}$, and thus (45) implies that $\boldsymbol{x} = \xi\boldsymbol{e}$, where $\xi = \max_k x_k = \|\boldsymbol{x}\|_2$. Therefore, by using (44),

$$\boldsymbol{y}^{\mathrm{T}}A\boldsymbol{x} = \sum_i y_i \max_k (A\boldsymbol{x})_k = \|\boldsymbol{y}^{\mathrm{T}}\|_1^{\mathrm{D}}\,\|\boldsymbol{x}\|_2 \max_k (A\boldsymbol{e})_k.$$

The result then follows since $\max_k (A\boldsymbol{e})_k = \mathrm{lub}_{12}(A)$.

Corollary 3.19. Property S_{12} holds for $\|.\|_1 = l^m_1(.)$, $\|.\|_2 = l^n_1(.)$.

Lemma 3.20. Property $\bar{\mathrm{S}}_{12}$ holds for $\|.\|_1 = l^m_\infty(.)$, $\|.\|_2 = l^n_\infty(.)$.

Proof. Let A be any non-negative $m\times n$ matrix, and $\boldsymbol{y}^{\mathrm{T}}$, $\boldsymbol{x}$ any pair of positive vectors satisfying (38). Then (44) and (45) hold for all i, by (9). Now $\boldsymbol{y}^{\mathrm{T}} > \mathbf{0}$, so $A\boldsymbol{x} = \zeta\boldsymbol{e}$, where $\zeta = \max_k (A\boldsymbol{x})_k$, by (44). Therefore

$$\boldsymbol{y}^{\mathrm{T}}A\boldsymbol{x} = \zeta\boldsymbol{y}^{\mathrm{T}}\boldsymbol{e} = \zeta\|\boldsymbol{y}^{\mathrm{T}}\|_1^{\mathrm{D}}. \tag{46}$$

If $A\boldsymbol{e}_j$, the jth column of A, is non-zero then $(\boldsymbol{y}^{\mathrm{T}}A)_j > 0$, so $x_j = \max_k x_k = \|\boldsymbol{x}\|_2$ by (45). Thus

$$\|\boldsymbol{x}\|_2 A\boldsymbol{e} = \|\boldsymbol{x}\|_2 \sum_j A\boldsymbol{e}_j = \sum_j x_j A\boldsymbol{e}_j = A\boldsymbol{x} = \zeta\boldsymbol{e}.$$

Now $\mathrm{lub}_{12}(A) = \max_k (A\boldsymbol{e})_k$, so $\zeta = \mathrm{lub}_{12}(A)\,\|\boldsymbol{x}\|_2$, from which the result follows by (46).

Corollary 3.21. Property $\bar{\mathrm{S}}_{12}$ holds for $\|.\|_1 = l^m_1(.)$, $\|.\|_2 = l^n_1(.)$.

Lemma 3.22. Property $\bar{\mathrm{S}}_{12}$ holds for $\|.\|_1 = l^m_p(.)$, $\|.\|_2 = l^n_p(.)$, $1 \leqslant p \leqslant \infty$.

Proof. If $p = 1$ or ∞ then the result follows from lemma 3.20, so it is assumed that $1 < p < \infty$. Let A be any $m\times n$ non-negative matrix. If A has zero rows or columns it follows from (7) that there is no pair of positive vectors satisfying (38), so the result holds trivially. So assume that A has no zero rows or columns, and let P and Q be such that

$$P^{\mathrm{T}}AQ = A_1 \oplus A_2 \oplus \ldots \oplus A_t, \tag{47}$$

where $K(A_i)$ is connected for each i. Let $\boldsymbol{y}^{\mathrm{T}}$ and $\boldsymbol{x}$ be any pair of positive vectors satisfying (38), and let $\boldsymbol{y}^{\mathrm{T}}P$ and $Q^{\mathrm{T}}\boldsymbol{x}$ be partitioned conformably with (47) so that

$$\boldsymbol{y}^{\mathrm{T}}P = [\boldsymbol{y}_1^{\mathrm{T}}\boldsymbol{y}_2^{\mathrm{T}}\ldots\boldsymbol{y}_t^{\mathrm{T}}], \quad Q^{\mathrm{T}}\boldsymbol{x} = \begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\\ \vdots\\ \boldsymbol{x}_t\end{bmatrix}.$$

Now, by (7), the dualities given by (38) imply the existence of $\alpha, \beta > 0$ such that (41) is satisfied. Since exponentiation is componentwise, this yields

$$A_i x_i = \alpha y_i^{q/p}, \quad A_i^T y_i = \beta x_i^{p/q} \quad (1 \leqslant i \leqslant t). \tag{48}$$

From (48) it immediately follows that

$$y_i^T \| A_i x_i \quad \text{and} \quad y_i^T A_i \| x_i \quad (1 \leqslant i \leqslant t),$$

where the dualities are with respect to $l_p(.)$ norms of the appropriate dimensions. Hence by corollary 3.17 y_i^T, x_i is a maximizing pair for A_i with respect to $l_p(.)$ norms of the appropriate dimensions.

Replacing x by a suitable multiple yields (42), as before, which, by (48), will also hold for each of the x_i, y_i pairs. From (42) it can be deduced that

$$y^T A x = \lambda(\|y^T\|_1^D)^q = \|x\|_2^p,$$

so

$$y^T A x = (y^T A x)^{1/q} (y^T A x)^{1/p} = \lambda^{1/q} \|y^T\|_1^D \|x\|_2. \tag{49}$$

Now, similarly, (49) must also hold for each maximizing pair y_i^T, x_i, which yields

$$\text{lub}\,(A_i) = \lambda^{1/q} \quad (1 \leqslant i \leqslant t), \tag{50}$$

where the lub norms are, again, with respect to the appropriate $l_p(.)$ norms. However, from corollary 3.7 and lemma 3.3, it follows that

$$\text{lub}_{12}(P^T A Q) = \max_{1 \leqslant i \leqslant t} \{\text{lub}\,(A_i)\},$$

since it is easy to see that any pair of norms $l_p^k(.)$, $l_p^l(.)$ has property L. [See §6.] Thus $\text{lub}_{12}(A) = \lambda^{1/q}$, by (11) and (50). It then follows from (49) that y^T, x is a maximizing pair for $\text{lub}_{12}(A)$, which completes the proof.

THEOREM 3.23. Property S_{12}^* holds for $\|.\|_1 = l_p^m(.)$, $\|.\|_2 = l_p^n(.)$, $1 \leqslant p \leqslant \infty$.

Proof. Let $A = P(A_1 \oplus 0)\, Q^T$ be a non-negative $m \times n$ matrix, with A_1 being $k \times l$, $1 \leqslant k \leqslant m$, $1 \leqslant l \leqslant n$, and let y_1 and x_1 be positive vectors in X_k and X_l, respectively, such that

$$y^T = (y_1^T \oplus 0)\, P^T \quad \text{and} \quad x = Q(x_1 \oplus 0)$$

satisfy (38). Now $Ax = P(A_1 x_1 \oplus 0)$ and $y^T A = (y_1^T A_1 \oplus 0)\, Q^T$, thus, by using (7), (8) or (9), the dualities (38) imply that $y_1^T \|_3\, A_1 x_1$ and $y_1^T A_1 \|_4\, x_1$ where $\|.\|_3 = l_p^k(.)$ and $\|.\|_4 = l_p^l(.)$. Therefore, by lemma 3.22, y_1^T, x_1 is a maximizing pair for $\text{lub}_{34}(A_1)$, so

$$\text{lub}_{34}(A_1) = \frac{y_1^T A_1 x_1}{\|y_1^T\|_3^D \|x_1\|_4} = \frac{y^T A x}{\|y^T\|_1^D \|x\|_2}. \tag{51}$$

However, since $\|.\|_2$ is subspace monotonic, $\text{lub}_{12}(A) = \text{lub}_{34}(A_1)$ by lemma 3.3. This, together with (51), implies that y^T, x is a maximizing pair for $\text{lub}_{12}(A)$, which completes the proof.

Finally, in this section, the following theorems, used in the sequel, are stated.

THEOREM 3.24. For any absolute norm $\|.\|$, given two positive vectors y and x in X_n, there exists $D \in \mathscr{D}_n$, unique up to positive multiples, such that $y^T D \| D^{-1} x$.

The proof of theorem 3.24 can be found in Stoer & Witzgall (1962). Several generalizations of this result have subsequently been obtained, in particular the following theorem (Gries & Stoer 1967):

THEOREM 3.25. For any absolute norm $\|.\|$, given a pair of non-zero non-negative vectors $\boldsymbol{y}$ and $\boldsymbol{x}$ in X_n with coinciding zeros, there exists $D \in \mathscr{D}_n$ such that $\boldsymbol{y}^T D \| D^{-1}\boldsymbol{x}$. Moreover, the diagonal entries of D, d_{ii}, for which $y_i, x_i = 0$ are arbitrary positive numbers, and those for which $y_i, x_i > 0$ are uniquely determined to within a positive multiple.

This theorem can also be obtained by using theorem 2 of Zenger (1968), for an absolute norm. In fact, Gries & Stoer prove theorem 3.25 for a somewhat larger class of norms, namely *orthant–monotonic* norms, and they also extend theorem 3.24 to *arbitrary* norms on $\mathbb{R}^n$ and *orthant–monotonic* norms on $\mathbb{C}^n$.

THEOREM 3.26. For any absolute norm $\|.\|$ having property S′, given a square matrix $A > 0$, there exists $D \in \mathscr{D}$ such that $\operatorname{lub}(D^{-1}AD) = \rho(A)$.

Actually, Stoer & Witzgall (1962) prove theorem 3.26 only for the l_p-norms, but the only use made of these norms is the fact that they have property S′. In fact, they show that these norms have property S, and thus, obviously, property S′.

THEOREM 3.27. For any pair of absolute norms $\|.\|_1$, $\|.\|_2$ having property S′, given a pair of *positive* matrices B and C, $m \times n$ and $n \times m$, respectively, there exist $D_1 \in \mathscr{D}_m$, $D_2 \in \mathscr{D}_n$ such that

$$\operatorname{lub}_{12}(D_1 B D_2)\operatorname{lub}_{21}(D_2^{-1} C D_1^{-1}) = \rho(BC).$$

Theorem 3.27 is part (i) of lemma I of Bauer (1963). Although Bauer obtains the result only for square matrices, B and C, the proof is also valid for co-conformable rectangular matrices.

4. COMBINATORIAL RESULTS

This section begins with a number of preliminary lemmas, which are subsequently used to obtain characterizations of the combinatorial and spectral structure of non-negative matrices having any of properties C^l, C^r and C. These characterizations are then extended to co-conformable pairs of non-negative matrices whose products possess any of these properties. It will be seen in §5 that these results can be used to characterize the structure of various classes of optimally scalable matrices.

The proof of the following lemma follows directly from the definition of matrix multiplication.

LEMMA 4.1. Let B and C be conformable non-negative matrices such that $BC = 0$, then

(i) if B has no zero columns, $C = 0$,
(ii) if C has no zero rows, $B = 0$.

LEMMA 4.2. Let A be a non-singular matrix, then the following conditions are equivalent:

(i) A is fully indecomposable,
(ii) $|A|\,|A^{-1}|$ is fully indecomposable,
(iii) $|A|\,|A^{-1}|$ is irreducible,
(iv) A^{-1} is fully indecomposable,
(v) $|A^{-1}|\,|A|$ is fully indecomposable,
(vi) $|A^{-1}|\,|A|$ is irreducible.

Proof. It can easily be seen that, if A is partly decomposable with P^TAQ in reduced form, A^{-1} is also partly decomposable with $Q^TA^{-1}P$ in reduced form. (See proof of corollary 1 to theorem 2 in Fenner & Loizou (1971).) Thus A is fully indecomposable if and only if A^{-1} is fully indecomposable, showing the equivalence of (i) and (iv). As (iv)–(vi) are the same as (i)–(iii) with A replaced by A^{-1} and vice versa, it is sufficient to prove the equivalence of (i), (ii) and (iii). Now, since the product of two non-negative fully indecomposable matrices is fully indecomposable (Fenner & Loizou 1971; Lewin 1971), (ii) follows from (i) and (iv). Since a fully indecomposable matrix is necessarily irreducible, it only remains to be shown that (iii) implies (i). However, if A is partly decomposable then, by the aforementioned result, $P^T|A|\,|A^{-1}|\,P$ will be in reduced form, contradicting (iii).

Lemma 4.3. Let B and C be conformable non-negative matrices such that

$$BC = \begin{bmatrix} F_1 & Z \\ 0 & F_2 \end{bmatrix}, \tag{52}$$

where F_1 and F_2 are non-vacuous, F_1 is square, and either C has no zero columns or F_1 is non-zero. Then there exists Q such that

$$BQ = \begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix}, \quad Q^TC = \begin{bmatrix} C_{11} & C_{12} \\ 0 & C_{22} \end{bmatrix}, \tag{53}$$

where B_{11} and C_{11} are co-conformable and non-vacuous, B_{11} has the same number of rows as F_1, and C_{11} has no zero rows. If, in addition, B has no zero rows then B_{22} and C_{22} are also non-vacuous.

Proof. Let

$$B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad C = [C_1 \quad C_2], \tag{54}$$

with B_1 having the same number of rows as F_1, and C_1 the same number of columns. Now $C_1 \neq 0$, since either C has no zero columns or, by (52) and (54), $B_1 C_1 = F_1 \neq 0$. Therefore, let Q be such that

$$Q^T C_1 = \begin{bmatrix} C_{11} \\ 0 \end{bmatrix}, \tag{55}$$

where C_{11} has no zero rows and is non-vacuous, but the zero submatrix may be vacuous. Letting

$$BQ = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, \quad Q^TC = \begin{bmatrix} C_{11} & C_{12} \\ 0 & C_{22} \end{bmatrix},$$

where these two matrices are partitioned so that B_{11} and C_{11} are co-conformable, it follows from (52) that $B_{21}C_{11} = 0$, whence lemma 4.1 implies that $B_{21} = 0$ as C_{11} has no zero rows. Thus the desired form (53) is obtained. If the zero submatrix in (55) is vacuous then so are B_{22} and C_{22}, which implies that B has zero rows since $B_{21} = 0$. Thus, if B has no zero rows, B_{22} and therefore C_{22} must be non-vacuous.

Lemma 4.4. Let B and C be co-conformable non-negative matrices with B having no zero columns, C having no zero rows, and BC irreducible, then CB is also irreducible.

Proof. Suppose that CB is reducible, then there exists Q such that

$$Q^TCBQ = \begin{bmatrix} F_1 & Z \\ 0 & F_2 \end{bmatrix},$$

where F_1 and F_2 are square and non-vacuous. Now C has no zero rows and B no zero columns so, by applying lemma 4.3 to Q^TC and BQ, it follows that there exists P such that

$$P^TBQ = \begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix}, \quad Q^TCP = \begin{bmatrix} C_{11} & C_{12} \\ 0 & C_{22} \end{bmatrix}, \tag{56}$$

where B_{11} and C_{11} are co-conformable and non-vacuous, and so are B_{22} and C_{22}. From (56) it is seen that P^TBCP is in reduced form which contradicts the irreducibility of BC, therefore CB is irreducible.

LEMMA 4.5. Let B and C be co-conformable non-negative matrices, with no zero rows or columns, such that

$$BC = F_1 \oplus F_2 \oplus \ldots \oplus F_t, \tag{57}$$

where the F_i are all square and non-vacuous. Then there exists Q such that

$$BQ = B_1 \oplus B_2 \oplus \ldots \oplus B_t, \tag{58}$$

$$Q^{\mathrm{T}}C = C_1 \oplus C_2 \oplus \ldots \oplus C_t, \tag{59}$$

where, for each i, B_i and C_i are co-conformable and non-vacuous with $B_i C_i = F_i$.

Proof. Suppose first that $t = 2$. Now B has no zero rows and C no zero columns, so by lemma 4.3 there exists Q such that (53) holds with B_{ii} and C_{ii} non-vacuous, and $B_{ii}C_{ii} = F_i$ for $i = 1, 2$. In the present case $Z = 0$ in (52), hence (53) implies that $B_{11}C_{12} = 0$ and $B_{12}C_{22} = 0$. As B has no zero columns neither does B_{11}, so $C_{12} = 0$ by lemma 4.1. Similarly, $B_{12} = 0$, which yields the result for $t = 2$. The result follows for all t by induction.

LEMMA 4.6. If A is a square irreducible non-negative matrix and $\boldsymbol{x}$ is a *positive* vector such that either (i) $A\boldsymbol{x} \leqslant \rho(A)\,\boldsymbol{x}$, or (ii) $A\boldsymbol{x} \geqslant \rho(A)\,\boldsymbol{x}$, then $A\boldsymbol{x} = \rho(A)\,\boldsymbol{x}$.

Proof. This follows directly from theorem 2.2 in Varga (1962).

LEMMA 4.7. If A is a square irreducible non-negative matrix and $\boldsymbol{x}$ is a *non-negative* vector such that, for some $\alpha > 0$, $A\boldsymbol{x} \leqslant \alpha\boldsymbol{x}$, then either $\boldsymbol{x} = \boldsymbol{0}$, or $\boldsymbol{x} > \boldsymbol{0}$ and $\alpha \geqslant \rho(A)$.

Proof. Suppose that A, $\boldsymbol{x}$ and α satisfy the conditions of the lemma. Then, since A is irreducible, $A+I$ is fully indecomposable (Fenner & Loizou 1971), so by theorem 2 of Lewin (1971), if $\boldsymbol{x} \neq \boldsymbol{0}$ and $\boldsymbol{x} \ngtr \boldsymbol{0}$, the number of non-zero components in $(A+I)\,\boldsymbol{x}$ exceeds the number in $\boldsymbol{x}$. However, this contradicts $(A+I)\,\boldsymbol{x} \leqslant (\alpha+1)\,\boldsymbol{x}$, so $\boldsymbol{x} = \boldsymbol{0}$ or $\boldsymbol{x} > \boldsymbol{0}$. If $\boldsymbol{x} > \boldsymbol{0}$ then, by lemma 4.6, $\alpha \geqslant \rho(A)$.

Before proceeding to characterize the structure of non-negative matrices which possess properties C^{l}, C^{r} and C, the following theorem, used subsequently, is stated:

THEOREM 4.8. If A is a square non-negative matrix and $\alpha > \rho(A)$, then

$$\alpha I - A \quad \text{is non-singular} \quad \text{and} \quad (\alpha I - A)^{-1} \geqslant 0.$$

The proof of this theorem can be found on page 83 of Varga (1962).

4.1. *The structure of matrices with properties* C^{l}, C^{r} *and* C

(i) *The matrix A*

LEMMA 4.9. Let $A \neq 0$ be a square non-negative matrix. Then there exists a positive vector $\boldsymbol{y}^{\mathrm{T}}$ such that

$$\boldsymbol{y}^{\mathrm{T}}A \leqslant \rho(A)\,\boldsymbol{y}^{\mathrm{T}} \tag{60}$$

if and only if there exists P such that

$$P^{\mathrm{T}}AP = \left[\begin{array}{ccccc|c} A_1 & & & & & Z_1 \\ & A_2 & & \mathrm{O} & & Z_2 \\ & & \cdot & & & \cdot \\ & \mathrm{O} & & \cdot & & \cdot \\ & & & & \cdot & \cdot \\ & & & & A_t & Z_t \\ \hline & & \mathrm{O} & & & \tilde{A} \end{array}\right], \tag{61}$$

where, for each i, A_i is a square irreducible matrix with $\rho(A_i) = \rho(A) > 0$, and $\rho(\tilde{A}) < \rho(A)$ if $\tilde{A}$ is non-vacuous.

Proof. Suppose that there exists P such that $P^{\mathrm{T}}AP$ has the form (61), and, for $1 \leqslant i \leqslant t$, let $\boldsymbol{y}_i^{\mathrm{T}} > \boldsymbol{0}$ be a left P-vector of A_i. If $\tilde{A}$ is vacuous then $\boldsymbol{y}^{\mathrm{T}} = (\boldsymbol{y}_1^{\mathrm{T}} \boldsymbol{y}_2^{\mathrm{T}} \ldots \boldsymbol{y}_t^{\mathrm{T}})\, P^{\mathrm{T}}$ satisfies (60) as $\boldsymbol{y}^{\mathrm{T}}A = \rho(A)\,\boldsymbol{y}^{\mathrm{T}}$. If $\tilde{A}$ is non-vacuous then $\rho(\tilde{A}) < \rho(A)$, so, by theorem 4.8, $(\rho(A)\,I - \tilde{A})^{-1}$ exists and is non-negative. If $\tilde{\boldsymbol{y}}^{\mathrm{T}}$ is defined by

$$\tilde{\boldsymbol{y}}^{\mathrm{T}} = \left(\boldsymbol{e}^{\mathrm{T}} + \sum_{i=1}^{t} \boldsymbol{y}_i^{\mathrm{T}} Z_i\right)(\rho(A)\,I - \tilde{A})^{-1}$$

then $\tilde{\boldsymbol{y}}^{\mathrm{T}}$ is positive and satisfies

$$\tilde{\boldsymbol{y}}^{\mathrm{T}}\tilde{A} + \sum_{i=1}^{t} \boldsymbol{y}_i^{\mathrm{T}} Z_i < \rho(A)\,\tilde{\boldsymbol{y}}^{\mathrm{T}}.$$

The vector $\boldsymbol{y}^{\mathrm{T}} = (\boldsymbol{y}_1^{\mathrm{T}} \boldsymbol{y}_2^{\mathrm{T}} \ldots \boldsymbol{y}_t^{\mathrm{T}}\, \tilde{\boldsymbol{y}}^{\mathrm{T}})\, P^{\mathrm{T}}$ thus satisfies (60).

Now suppose, conversely, the existence of $\boldsymbol{y}^{\mathrm{T}} > \boldsymbol{0}$ satisfying (60). Since $A \neq 0$, $\rho(A) > 0$ by (60). Let P_0 be such that $P_0^{\mathrm{T}} A P_0$ is in *normal form* (Varga 1962, p. 46), and let A_{kk} be the last diagonal block having spectral radius $\rho(A)$, so A_{kk} is irreducible as $\rho(A) > 0$. Then $P_0^{\mathrm{T}} A P_0$ may be partitioned as

$$P_0^{\mathrm{T}} A P_0 = \begin{bmatrix} U_{11} & U_{12} & U_{13} \\ 0 & A_{kk} & U_{23} \\ 0 & 0 & U_{33} \end{bmatrix}, \tag{62}$$

where the diagonal blocks are square, and if U_{33} is non-vacuous $\rho(U_{33}) < \rho(A)$. If U_{11} is vacuous then (62) is already in the form (61) with $t = 1$ and $A_1 = A_{kk}$, so assume that U_{11} is non-vacuous and let $\boldsymbol{y}^{\mathrm{T}} P_0$ be partitioned conformably with (62) as $(\boldsymbol{y}_1^{\mathrm{T}} \boldsymbol{y}_2^{\mathrm{T}} \boldsymbol{y}_3^{\mathrm{T}})$. Inequality (60) thus implies that

$$\boldsymbol{y}_1^{\mathrm{T}} U_{12} + \boldsymbol{y}_2^{\mathrm{T}} A_{kk} \leqslant \rho(A)\,\boldsymbol{y}_2^{\mathrm{T}}. \tag{63}$$

By using lemma 4.6, (63) implies that $\boldsymbol{y}_2^{\mathrm{T}} A_{kk} = \rho(A)\,\boldsymbol{y}_2^{\mathrm{T}}$, therefore $\boldsymbol{y}_1^{\mathrm{T}} U_{12} = \boldsymbol{0}$, so $U_{12} = 0$. Thus, if P_1 interchanges the first and second row and column blocks of (62),

$$P_1^{\mathrm{T}} P_0^{\mathrm{T}} A P_0 P_1 = \begin{bmatrix} A_{kk} & 0 & U_{23} \\ 0 & U_{11} & U_{13} \\ 0 & 0 & U_{33} \end{bmatrix}. \tag{64}$$

If $\rho(U_{11}) < \rho(A)$ then (64) is already in the form (61) with $t = 1$ and $A_1 = A_{kk}$, otherwise the above process of decomposition and permutation can be applied to U_{11} since U_{11} and $\boldsymbol{y}_1^{\mathrm{T}}$ satisfy the inequality (60). Repetition of this process must ultimately yield a matrix in the required form (61).

Corollary 4.10. Let $A \neq 0$ be a square non-negative matrix. Then there exists a positive vector $\boldsymbol{x}$ such that $A\boldsymbol{x} \leqslant \rho(A)\,\boldsymbol{x}$ if and only if there exists Q such that

$$Q^{\mathrm{T}} A Q = \begin{bmatrix} \hat{A} & 0 \\ W & \tilde{\tilde{A}} \end{bmatrix}, \tag{65}$$

where $\hat{A}$ is the direct sum of irreducible matrices all having spectral radius $\rho(A) > 0$, and $\rho(\tilde{\tilde{A}}) < \rho(A)$ if $\tilde{\tilde{A}}$ is non-vacuous.

Proof. This is immediate on applying lemma 4.9. to A^{T}.

Corollary 4.11. Let $A \neq 0$ be a square non-negative matrix. Then there exists a pair of positive vectors $\boldsymbol{y}^{\mathrm{T}}$, $\boldsymbol{x}$ such that

$$\boldsymbol{y}^{\mathrm{T}} A \leqslant \rho(A)\,\boldsymbol{y}^{\mathrm{T}}, \quad A\boldsymbol{x} \leqslant \rho(A)\,\boldsymbol{x}, \tag{66}$$

if and only if there exists P such that

$$P^{\mathrm{T}}AP = \left[\begin{array}{cccc:c} A_1 & & & & \\ & A_2 & & \mathrm{O} & \\ & & \cdot & & \mathrm{O} \\ & \mathrm{O} & & \cdot & \\ & & & A_t & \\ \hdashline & & \mathrm{O} & & \tilde{A} \end{array}\right], \tag{67}$$

where, for each i, A_i is a square irreducible matrix with $\rho(A_i) = \rho(A) > 0$, and $\rho(\tilde{A}) < \rho(A)$ if $\tilde{A}$ is non-vacuous.

Proof. If $P^{\mathrm{T}}AP$ is in the form (67) then (66) follows from lemma 4.9 and corollary 4.10.

If, conversely, there exist positive vectors $\boldsymbol{y}^{\mathrm{T}}$ and $\boldsymbol{x}$ satisfying (66), then, by lemma 4.9 and corollary 4.10, there exist P and Q such that $P^{\mathrm{T}}AP$ and $Q^{\mathrm{T}}AQ$ are in the forms (61) and (65), respectively. Now the strong components of $G(A)$ are invariant under permutation and must, therefore, consist of the $G(A_i)$, $1 \leqslant i \leqslant t$, and the strong components of $G(\tilde{A})$, in some order. Moreover, $G(A_1 \oplus A_2 \oplus \ldots \oplus A_t)$ and $G(\hat{A})$ must both consist of just those strong components which correspond to submatrices having spectral radius $\rho(A)$. Thus the rows and columns of A occurring in $\hat{A}$ are precisely those occurring in $A_1 \oplus A_2 \oplus \ldots \oplus A_t$, so $Z_1, Z_2, \ldots, Z_t$ and W are all zero, which yields the result (67).

Lemma 4.12. Let $A \neq 0$ be a square non-negative matrix. Then A has property C^{l} if and only if there exists P such that (61) holds, with the A_i and $\tilde{A}$ satisfying the same conditions as in the statement of lemma 4.9.

Proof. Suppose that there exists P such that $P^{\mathrm{T}}AP$ has the form (61), and let

$$\boldsymbol{x} = P(\boldsymbol{x}_1^{\mathrm{T}}\,\boldsymbol{x}_2^{\mathrm{T}} \ldots \boldsymbol{x}_t^{\mathrm{T}}\,\tilde{\boldsymbol{x}}^{\mathrm{T}})^{\mathrm{T}}$$

be any right P-vector of A, partitioned conformably with (61). Then

$$\tilde{A}\tilde{\boldsymbol{x}} = \rho(A)\,\tilde{\boldsymbol{x}}. \tag{68}$$

If $\tilde{A}$ is non-vacuous then, as $\rho(\tilde{A}) < \rho(A)$, (68) implies that $\tilde{\boldsymbol{x}} = \mathbf{0}$. Now, for each i, let $\boldsymbol{y}_i^{\mathrm{T}}$ be a left P-vector of A_i, which must be positive since A_i is irreducible, and if $\tilde{A}$ is non-vacuous let $\tilde{\boldsymbol{y}}^{\mathrm{T}} = \sum_{i=1}^{t} \boldsymbol{y}_i^{\mathrm{T}} Z_i(\rho(A)\,I - \tilde{A})^{-1}$, which must be non-negative since $(\rho(A)\,I - \tilde{A})^{-1} \geqslant 0$ by theorem 4.8. Then $\boldsymbol{y}^{\mathrm{T}} = (\boldsymbol{y}_1^{\mathrm{T}}\boldsymbol{y}_2^{\mathrm{T}} \ldots \boldsymbol{y}_t^{\mathrm{T}}\tilde{\boldsymbol{y}}^{\mathrm{T}})\,P^{\mathrm{T}}$ is a left P-vector of A, any zero element of which must lie in $\tilde{\boldsymbol{y}}^{\mathrm{T}}$. Hence A has property C^{l}.

Now assume, conversely, that A has property C^{l}. Suppose $\rho(A) = 0$, then A must have at least one zero column. Let Q_0 be such that

$$Q_0^{\mathrm{T}}AQ_0 = \begin{bmatrix} A_{11} & 0 \\ A_{21} & 0 \end{bmatrix}, \tag{69}$$

where the first column block in (69) has no zero columns (it must be non-vacuous since $A \neq 0$), and A_{11} is square. Since $\rho(A_{11}) = 0$, A_{11} has at least one zero column, so $A_{21} \neq 0$. Now $(\mathbf{0}\,\boldsymbol{e}^{\mathrm{T}})^{\mathrm{T}}$, partitioned conformably with (69), is a right P-vector of $Q_0^{\mathrm{T}}AQ_0$, so by property C^{l} there exists a left P-vector $\boldsymbol{y}^{\mathrm{T}} = (\boldsymbol{y}_1^{\mathrm{T}}\boldsymbol{y}_2^{\mathrm{T}})$ of $Q_0^{\mathrm{T}}AQ_0$ with $\boldsymbol{y}_2^{\mathrm{T}} > \mathbf{0}$. This contradicts $A_{21} \neq 0$, therefore $\rho(A) > 0$. Let P_0 be such that $P_0^{\mathrm{T}}AP_0$ is in the form (62), where again A_{kk} is irreducible, and $\rho(U_{33}) < \rho(A)$

if U_{33} is non-vacuous. If U_{11} is vacuous then $P_0^T A P_0$ is already in the form (61) with $t = 1$ and $A_1 = A_{kk}$, so assume that U_{11} is non-vacuous. Let $\boldsymbol{x}$ be a right P-vector of A having the *maximum* number of non-zero components, and, by using property C^l, let $\boldsymbol{y}^T$ be a left P-vector of A such that $y_i = 0$ implies $x_i = 0$. Also let $\boldsymbol{y}^T P_0$ and $P_0^T \boldsymbol{x}$ be partitioned conformably with (62) as $(\boldsymbol{y}_1^T \boldsymbol{y}_2^T \boldsymbol{y}_3^T)$ and $(\boldsymbol{x}_1^T \boldsymbol{x}_2^T \boldsymbol{x}_3^T)^T$, respectively. If U_{33} is non-vacuous then $\boldsymbol{x}_3 = \boldsymbol{0}$ since $\rho(U_{33}) < \rho(A)$. Now choose R_1 such that $\boldsymbol{y}_1^T R_1 = (\hat{\boldsymbol{y}}_0^T \boldsymbol{0})$ where $\hat{\boldsymbol{y}}_0^T > \boldsymbol{0}$ if it is non-vacuous. If $R_1^T \boldsymbol{x}_1$ is partitioned correspondingly as $(\hat{\boldsymbol{x}}_0^T \hat{\boldsymbol{x}}_1^T)^T$ then, by hypothesis, $\hat{\boldsymbol{x}}_1 = \boldsymbol{0}$ if it is non-vacuous.

Now letting $Q_1 = P_0(R_1 \oplus I)$, correspondingly partition the first row block and column block of $Q_1^T A Q_1$, so that

$$Q_1^T A Q_1 = \begin{bmatrix} V_{00} & V_{01} & V_{02} & V_{03} \\ V_{10} & V_{11} & V_{12} & V_{13} \\ 0 & 0 & A_{kk} & U_{23} \\ 0 & 0 & 0 & U_{33} \end{bmatrix}. \tag{70}$$

Similarly, let $\boldsymbol{y}^T Q_1 = (\hat{\boldsymbol{y}}_0^T \boldsymbol{0}\, \boldsymbol{y}_2^T \boldsymbol{y}_3^T)$ and $Q_1^T \boldsymbol{x} = (\hat{\boldsymbol{x}}_0^T \boldsymbol{0}\, \boldsymbol{x}_2^T \boldsymbol{0})^T$. It is assumed here that each of the submatrices of (70) is non-vacuous; however, the proof which follows is still valid, with minor modifications, if this is not the case. It is now shown that

$$V_{02} = 0, \quad V_{12} = 0. \tag{71}$$

From (70),

$$\hat{\boldsymbol{y}}_0^T V_{01} = \boldsymbol{0}, \tag{72}$$

$$\hat{\boldsymbol{y}}_0^T V_{02} + \boldsymbol{y}_2^T A_{kk} = \rho(A)\, \boldsymbol{y}_2^T, \tag{73}$$

$$V_{10} \hat{\boldsymbol{x}}_0 + V_{12} \boldsymbol{x}_2 = \boldsymbol{0}, \tag{74}$$

$$A_{kk} \boldsymbol{x}_2 = \rho(A)\, \boldsymbol{x}_2. \tag{75}$$

From (73),

$$\boldsymbol{y}_2^T A_{kk} \leqslant \rho(A)\, \boldsymbol{y}_2^T. \tag{76}$$

As A_{kk} is irreducible, $\boldsymbol{y}_2^T$ is either zero or positive by lemma 4.7. Similarly, from (75), $\boldsymbol{x}_2$ is either zero or positive. If $\boldsymbol{y}_2^T > \boldsymbol{0}$ then equality holds in (76) by lemma 4.6, and trivially it also holds if $\boldsymbol{y}_2^T = \boldsymbol{0}$. Therefore, by (73), $\hat{\boldsymbol{y}}_0^T V_{02} = \boldsymbol{0}$. This and (72) imply that $V_{01} = 0$ and $V_{02} = 0$, since $\hat{\boldsymbol{y}}_0^T > \boldsymbol{0}$. If $\boldsymbol{x}_2 > \boldsymbol{0}$ then (74) implies that $V_{12} = 0$, yielding (71). If $\boldsymbol{x}_2 = \boldsymbol{0}$ let $(\bar{\boldsymbol{x}}_1^T \bar{\boldsymbol{x}}_2^T)^T$ be a right P-vector of

$$\begin{bmatrix} V_{11} & V_{12} \\ 0 & A_{kk} \end{bmatrix},$$

and let $\bar{\boldsymbol{x}} = \boldsymbol{x} + Q_1(\boldsymbol{0}\, \bar{\boldsymbol{x}}_1^T \bar{\boldsymbol{x}}_2^T \boldsymbol{0})^T$. Then $\bar{\boldsymbol{x}}$ is also a right P-vector of A, since $V_{01} = 0$ and $V_{02} = 0$, which contradicts the 'maximality' of $\boldsymbol{x}$. Thus $\boldsymbol{x}_2$ is non-zero, establishing (71), which implies that $U_{12} = 0$ in (62). Now, on defining P_1 as in lemma 4.9, (64) holds, so if $\rho(U_{11}) < \rho(A)$ (64) is already in the form (61) with $t = 1$ and $A_1 = A_{kk}$. If this is not the case then consider U_{11}. It is easy to see that this matrix has property C^l, since A does, so the above process of decomposition and permutation can be applied to U_{11}. Repeating this ultimately yields a matrix in the required form (61).

Corollary 4.13. Let $A \neq 0$ be a square non-negative matrix. Then A has property C^r if and only if there exists Q such that (65) holds with $\hat{A}$ and $\hat{\tilde{A}}$ as in corollary 4.10.

Corollary 4.14. Let $A \neq 0$ be a square non-negative matrix. Then A has property C if and only if there exists P such that (67) holds with the A_i and $\tilde{A}$ satisfying the same conditions as in the statement of corollary 4.11.

Proof. Suppose that there exists P such that $P^{T}AP$ has the form (67), and let

$$\boldsymbol{x} = P(\boldsymbol{x}_1^{T}\,\boldsymbol{x}_2^{T}\,\dots\,\boldsymbol{x}_t^{T}\,\tilde{\boldsymbol{x}}^{T})^{T}$$

be any right P-vector of A, partitioned conformably with (67). Then

$$A_i\boldsymbol{x}_i = \rho(A)\,\boldsymbol{x}_i \quad (1 \leqslant i \leqslant t), \tag{77}$$

$$\tilde{A}\tilde{\boldsymbol{x}} = \rho(A)\,\tilde{\boldsymbol{x}}. \tag{78}$$

If $\tilde{A}$ is non-vacuous then, as $\rho(\tilde{A}) < \rho(A)$, (78) implies that $\tilde{\boldsymbol{x}} = \boldsymbol{0}$. Similarly, (77) implies that each $\boldsymbol{x}_i$ $(1 \leqslant i \leqslant t)$ is either zero or a right P-vector of A_i. For each i, let $\boldsymbol{y}_i^{T}$ correspondingly be either zero or a left P-vector of A_i. In the latter case $\boldsymbol{x}_i > \boldsymbol{0}$ and $\boldsymbol{y}_i^{T} > \boldsymbol{0}$ as A_i is irreducible. Now, letting $\tilde{\boldsymbol{y}}^{T} = \boldsymbol{0}$, $\boldsymbol{y}^{T} = (\boldsymbol{y}_1^{T}\boldsymbol{y}_2^{T}\dots\boldsymbol{y}_t^{T}\tilde{\boldsymbol{y}}^{T})\,P^{T}$ is a left P-vector of A, the zeros of which coincide with those of $\boldsymbol{x}$. Hence A has property C.

Now suppose, conversely, that A has property C. Certainly A has property C^{l} so, by lemma 4.12, there exists P such that (61) holds, with the A_i and $\tilde{A}$ satisfying the same conditions as in the statement of lemma 4.9. For $1 \leqslant i \leqslant t$, let $\boldsymbol{x}_i$ be a right P-vector of A_i, so $\boldsymbol{x}_i > \boldsymbol{0}$. Thus

$$\boldsymbol{x} = P(\boldsymbol{x}_1^{T}\,\boldsymbol{x}_2^{T}\,\dots\,\boldsymbol{x}_t^{T}\,\boldsymbol{0})^{T}$$

is a right P-vector of A, so let $\boldsymbol{y}^{T} = (\boldsymbol{y}_1^{T}\boldsymbol{y}_2^{T}\dots\boldsymbol{y}_t^{T}\,\boldsymbol{0})P^{T}$ be a corresponding left P-vector of A with $\boldsymbol{y}_i^{T} > \boldsymbol{0}$, $1 \leqslant i \leqslant t$. From (61) it follows that $\sum_{i=1}^{t}\boldsymbol{y}_i^{T}Z_i = \boldsymbol{0}$, so $Z_i = 0$, $1 \leqslant i \leqslant t$, which yields the required form (67).

Combining now lemmas 4.9 and 4.12, and similarly for their corresponding corollaries, yields the following theorems characterizing properties C^{l}, C^{r} and C.

THEOREM 4.15. For a square non-negative matrix $A \neq 0$, the following are equivalent:
(i) A has property C^{l}.
(ii) There exists $\boldsymbol{y}^{T} > \boldsymbol{0}$ such that $\boldsymbol{y}^{T}A \leqslant \rho(A)\,\boldsymbol{y}^{T}$.
(iii) There exists P such that (61) holds with the A_i and $\tilde{A}$ as in lemma 4.9.

THEOREM 4.16. For a square non-negative matrix $A \neq 0$, the following are equivalent:
(i) A has property C^{r}.
(ii) There exists $\boldsymbol{x} > \boldsymbol{0}$ such that $A\boldsymbol{x} \leqslant \rho(A)\,\boldsymbol{x}$.
(iii) There exists Q such that (65) holds with $\hat{A}$ and $\tilde{\tilde{A}}$ as in corollary 4.10.

THEOREM 4.17. For a square non-negative matrix $A \neq 0$, the following are equivalent:
(i) A has property C.
(ii) There exist $\boldsymbol{y}^{T}, \boldsymbol{x} > \boldsymbol{0}$ such that $\boldsymbol{y}^{T}A \leqslant \rho(A)\,\boldsymbol{y}^{T}$ and $A\boldsymbol{x} \leqslant \rho(A)\,\boldsymbol{x}$.
(iii) There exists P such that (67) holds with the A_i and $\tilde{A}$ as in corollary 4.11.

COROLLARY 4.18. A square non-negative matrix A has property C if and only if it has properties C^{l} and C^{r}.

Either from theorem 4.17 because of the form of (67), or from corollary 4.18, it follows that A has property C if and only if A^{T} also has property C, thus, in the definition of property C the roles of left and right P-vectors could be reversed.

COROLLARY 4.19. If a square non-negative matrix A has a pair of positive left and right P-vectors then A has property C.

COROLLARY 4.20. Let $A \neq 0$ be a square non-negative matrix. Then A has a pair of positive left and right P-vectors if and only if there exists P such that $P^{T}AP$ is the direct sum of irreducible matrices all having spectral radius $\rho(A)$.

Proof. If A has a pair of positive P-vectors then, by theorem 4.17, (67) holds, and it follows from the proof of corollary 4.14 that $\tilde{A}$ must be vacuous. The proof in the opposite direction is immediate.

(ii) *The product BC*

In this subsection the structure of pairs of co-conformable non-negative matrices, B and C, whose products BC and CB both have property C^l, C^r or C, is examined. It is first shown that, in certain circumstances, if one of the products has one of these properties then so does the other.

LEMMA 4.21. Let B and C be co-conformable non-negative matrices with B having no zero columns, C having no zero rows and BC having property C^l, then CB also has property C^l.

Proof. Since B has no zero columns and $C \neq 0$, $BC \neq 0$ by lemma 4.1. Thus BC satisfies the conditions of theorem 4.15 (i), so there exists P such that

$$P^{\mathrm{T}}BCP = \begin{bmatrix} \hat{F} & Z \\ 0 & \tilde{F} \end{bmatrix},$$

where $\hat{F} = F_1 \oplus F_2 \oplus \ldots \oplus F_t$ with the F_i irreducible matrices all having spectral radius $\rho(BC) > 0$, and $\rho(\tilde{F}) < \rho(BC)$ if $\tilde{F}$ is non-vacuous. Now $\hat{F} \neq 0$ so, if $\tilde{F}$ is non-vacuous, applying lemma 4.3 to $P^{\mathrm{T}}B$ and CP implies the existence of Q such that (56) holds with B_{11} and C_{11} co-conformable and non-vacuous, $B_{11}C_{11} = \hat{F}$, and C_{11} having no zero rows. Since B has no zero columns neither does B_{11}, and as $\hat{F}$ is the direct sum of irreducible matrices it has no zero rows or columns, so B_{11} has no zero rows and C_{11} no zero columns.

If $\tilde{F}$ is vacuous then let $B_{11} = P^{\mathrm{T}}B$, $C_{11} = CP$, which again can have no zero rows or columns. Therefore, in all cases, B_{11} and C_{11} satisfy the conditions of lemma 4.5, so there exists R such that

$$B_{11}R = B_1 \oplus B_2 \oplus \ldots \oplus B_t, \tag{79}$$

$$R^{\mathrm{T}}C_{11} = C_1 \oplus C_2 \oplus \ldots \oplus C_t, \tag{80}$$

where B_iC_i is irreducible and $\rho(B_iC_i) = \rho(BC)$ $(1 \leqslant i \leqslant t)$. Now, for each i, B_i has no zero columns and C_i no zero rows, so C_iB_i is irreducible by lemma 4.4. Thus, from (79) and (80), it is seen that $R^{\mathrm{T}}C_{11}B_{11}R$ is the direct sum of irreducible matrices all having spectral radius $\rho(BC) > 0$. If $\tilde{F}$ is non-vacuous, by using (56), it is now clear that $(R^{\mathrm{T}} \oplus I)\, Q^{\mathrm{T}}CBQ(R \oplus I)$ has the form (61); similarly, if $\tilde{F}$ is vacuous, $R^{\mathrm{T}}CBR$ has this form, and thus in either case CB satisfies the conditions of theorem 4.15 (iii). Therefore CB has property C^l.

COROLLARY 4.22. Let B and C be co-conformable non-negative matrices with B having no zero columns, C having no zero rows and BC having property C^r, then CB also has property C^r.

Proof. Since $C^{\mathrm{T}}B^{\mathrm{T}}$ has property C^l, lemma 4.21 implies that $B^{\mathrm{T}}C^{\mathrm{T}}$ also has property C^l; this yields the result.

COROLLARY 4.23. Let B and C be co-conformable non-negative matrices with B having no zero columns, C having no zero rows and BC having property C, then CB also has property C.

Proof. This is immediate from lemma 4.21 and corollary 4.22 on using corollary 4.18.

LEMMA 4.24. Let B and C be co-conformable non-negative matrices for which BC and CB are both non-zero and have property C^l. Now let P and Q be permutation matrices (whose existence is guaranteed by theorem 4.15) such that

$$P^{\mathrm{T}}BCP = \begin{bmatrix} \hat{F} & Z \\ 0 & \tilde{F} \end{bmatrix}, \quad Q^{\mathrm{T}}CBQ = \begin{bmatrix} \hat{H} & W \\ 0 & \tilde{H} \end{bmatrix}, \tag{81}$$

where
$$\hat{F} = F_1 \oplus F_2 \oplus \ldots \oplus F_s, \quad \hat{H} = H_1 \oplus H_2 \oplus \ldots \oplus H_t,$$
with F_i $(1 \leqslant i \leqslant s)$ and H_j $(1 \leqslant j \leqslant t)$ all being irreducible matrices having spectral radius $\rho(BC)$; $\rho(\tilde{F}) < \rho(BC)$ if $\tilde{F}$ is non-vacuous, and $\rho(\tilde{H}) < \rho(BC)$ if $\tilde{H}$ is non-vacuous. Then
$$P^{\mathrm{T}}BQ = \begin{bmatrix} \hat{B} & U_1 \\ 0 & \tilde{B} \end{bmatrix}, \quad Q^{\mathrm{T}}CP = \begin{bmatrix} \hat{C} & V_1 \\ 0 & \tilde{C} \end{bmatrix}, \tag{82}$$
where $\hat{B}$ has the same number of rows as $\hat{F}$ and the same number of columns as $\hat{H}$, and conversely for $\hat{C}$. In addition, $\hat{B}$ and $\hat{C}$ have no zero rows or columns.

Proof. Let
$$P^{\mathrm{T}}BQ = \begin{bmatrix} \hat{B} & U_1 \\ B_1 & \tilde{B} \end{bmatrix}, \quad Q^{\mathrm{T}}CP = \begin{bmatrix} \hat{C} & V_1 \\ C_1 & \tilde{C} \end{bmatrix}, \tag{83}$$
where $\hat{B}$ has the same number of rows as $\hat{F}$ and the same number of columns as $\hat{H}$, and conversely for $\hat{C}$, so possibly U_1 or B_1, or both, could be vacuous. Note that B_1 and V_1 are non-vacuous if and only if $\tilde{F}$ is non-vacuous, and C_1 and U_1 are non-vacuous if and only if $\tilde{H}$ is non-vacuous. Since all the matrices involved are non-negative it follows from (81) and (83) that
$$\hat{B}\hat{C} + U_1 C_1 = \hat{F}, \tag{84a}$$
$$\hat{C}\hat{B} + V_1 B_1 = \hat{H}, \tag{84b}$$
$$B_1 V_1 + \tilde{B}\tilde{C} = \tilde{F}, \quad \text{if } \tilde{F} \text{ is non-vacuous}, \tag{85a}$$
$$C_1 U_1 + \tilde{C}\tilde{B} = \tilde{H}, \quad \text{if } \tilde{H} \text{ is non-vacuous}, \tag{85b}$$
$$B_1 \hat{C} = 0, \quad \text{if } \tilde{F} \text{ is non-vacuous}, \tag{86a}$$
$$C_1 \hat{B} = 0, \quad \text{if } \tilde{H} \text{ is non-vacuous}. \tag{86b}$$
If $\tilde{F}$ is vacuous then $V_1 B_1$ and $\tilde{C}\tilde{B}$ in (84*b*) and (85*b*), respectively, should be taken as zero, and similarly for $U_1 C_1$ and $\tilde{B}\tilde{C}$ in (84*a*) and (85*a*), respectively, if $\tilde{H}$ is vacuous.

Now, if $\tilde{H}$ is non-vacuous, by (85*b*) and theorem 2.8 in Varga (1962),
$$\rho(\hat{F}) = \rho(BC) > \rho(\tilde{H}) \geqslant \rho(C_1 U_1) = \rho(U_1 C_1).$$
It therefore follows from (84*a*) that $\hat{B}\hat{C} \neq 0$, and obviously this also holds if $\tilde{H}$ is vacuous.

Suppose first that $\tilde{F}$ is non-vacuous; so by using (86*a*)
$$\begin{bmatrix} \hat{B} \\ B_1 \end{bmatrix} [\hat{C} \quad V_1] = \begin{bmatrix} \hat{B}\hat{C} & \hat{B}V_1 \\ 0 & B_1 V_1 \end{bmatrix}.$$
Thus, since $\hat{B}\hat{C} \neq 0$, by lemma 4.3 there exists R such that
$$\begin{bmatrix} \hat{B} \\ B_1 \end{bmatrix} R = \begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix}, \quad R^{\mathrm{T}}[\hat{C} \quad V_1] = \begin{bmatrix} C_{11} & C_{12} \\ 0 & C_{22} \end{bmatrix}, \tag{87}$$
where B_{11} and C_{11} are co-conformable and non-vacuous, and C_{11} has no zero rows. Suppose that B_{22} and C_{22} are non-vacuous, then from (87), by using (84*b*),
$$R^{\mathrm{T}}\hat{H}R = \begin{bmatrix} C_{11}B_{11} & C_{11}B_{12} + C_{12}B_{22} \\ 0 & C_{22}B_{22} \end{bmatrix}.$$
Since the H_j are all irreducible and have spectral radius $\rho(BC)$, it follows that
$$\rho(C_{11}B_{11}) = \rho(C_{22}B_{22}) = \rho(BC). \tag{88}$$

Now, from (85*a*) and (87), $\tilde{F} \geqslant B_1 V_1 = B_{22} C_{22}$, so again by theorem 2.8 in Varga (1962),

$$\rho(BC) > \rho(\tilde{F}) \geqslant \rho(B_{22} C_{22}) = \rho(C_{22} B_{22}),$$

which contradicts (88). Accordingly B_{22} and C_{22} must be vacuous, and (87) reduces to

$$\begin{bmatrix} \hat{B} \\ B_1 \end{bmatrix} R = \begin{bmatrix} B_{11} \\ 0 \end{bmatrix}, \quad R^{\mathrm{T}}[\hat{C} \quad V_1] = [C_{11} \quad C_{12}].$$

Thus $\hat{C}$ has no zero rows and $B_1 = 0$,whence (84*b*) implies that $\hat{B}$ has no zero columns.

Now consider the case when $\tilde{F}$ is vacuous. Equation (84*b*) reduces to $\hat{C}\hat{B} = \hat{H}$, so again $\hat{C}$ has no zero rows, $\hat{B}$ no zero columns, and B_1 is, vacuously, zero.

By symmetry between B and C it follows similarly that $\hat{B}$ has no zero rows, $\hat{C}$ no zero columns, and $C_1 = 0$ if it is non-vacuous, which completes the proof.

A corresponding result holds if BC and CB both have property C^{r}, and the analogous result for property C is given by the following lemma.

LEMMA 4.25. Let B and C be co-conformable non-negative matrices for which BC and CB are both non-zero and have property C. Now let P and Q be permutation matrices (whose existence is guaranteed by theorem 4.17) such that

$$P^{\mathrm{T}}BCP = \begin{bmatrix} \hat{F} & 0 \\ 0 & \tilde{F} \end{bmatrix}, \quad Q^{\mathrm{T}}CBQ = \begin{bmatrix} \hat{H} & 0 \\ 0 & \tilde{H} \end{bmatrix},$$

where $\hat{F}$, $\hat{H}$, $\tilde{F}$ and $\tilde{H}$ satisfy the same conditions as in lemma 4.24. Then

$$P^{\mathrm{T}}BQ = \begin{bmatrix} \hat{B} & 0 \\ 0 & \tilde{B} \end{bmatrix}, \quad Q^{\mathrm{T}}CP = \begin{bmatrix} \hat{C} & 0 \\ 0 & \tilde{C} \end{bmatrix}, \tag{89}$$

where $\hat{B}$ and $\hat{C}$ satisfy the same conditions as in lemma 4.24.

Proof. By lemma 4.24, $P^{\mathrm{T}}BQ$ and $Q^{\mathrm{T}}CP$ satisfy (82), and $\hat{B}$ and $\hat{C}$ have no zero rows or columns. Moreover, theorem 4.17 implies that W and Z are zero in (81) if they are non-vacuous. Then, from (82), $\hat{B}V_1 = 0$ if V_1 is non-vacuous, and $\hat{C}U_1 = 0$ if U_1 is non-vacuous. It thus follows from lemma 4.1 that U_1 and V_1 are zero if they are non-vacuous, which completes the proof.

It is noted that if BC is non-zero and has any of properties C^{l}, C^{r} and C then $\rho(BC) > 0$ by theorems 4.15, 4.16 and 4.17, respectively, so CB is also non-zero. Thus the condition that CB be non-zero in lemmas 4.24 and 4.25 is redundant. These lemmas are now used to characterize the structure of B and C given that BC and CB both have property C or C^{l}.

THEOREM 4.26. Let B and C be co-conformable non-negative matrices such that $BC \neq 0$. Then BC and CB both have property C if and only if there exist P and Q such that

$$P^{\mathrm{T}}BQ = \left[\begin{array}{cccc:c} B_1 & & & \mathrm{O} & \\ & B_2 & & & \\ & & \ddots & & \mathrm{O} \\ \mathrm{O} & & & B_t & \\ \hdashline & & \mathrm{O} & & \tilde{B} \end{array}\right], \quad Q^{\mathrm{T}}CP = \left[\begin{array}{cccc:c} C_1 & & & \mathrm{O} & \\ & C_2 & & & \\ & & \ddots & & \mathrm{O} \\ \mathrm{O} & & & C_t & \\ \hdashline & & \mathrm{O} & & \tilde{C} \end{array}\right], \tag{90}$$

where, for each i, B_i and C_i are co-conformable, $B_i C_i$ and $C_i B_i$ are irreducible matrices having spectral radius $\rho(BC)$, and $\rho(\tilde{B}\tilde{C}) < \rho(BC)$ if $\tilde{B}$ and $\tilde{C}$ are non-vacuous.

Proof. Suppose B and C can be put in the form (90) satisfying all the required conditions. If $\tilde{B}$ and $\tilde{C}$ are non-vacuous then by (90) $P^{T}BCP$ has the form of (67), and theorem 4.17 implies that BC has property C. If $\tilde{B}$ and $\tilde{C}$ are vacuous then

$$P^{T}BCP = \begin{cases} B_1C_1 \oplus B_2C_2 \oplus \dots \oplus B_tC_t, & \text{if } \tilde{B} \text{ has no rows,} \\ B_1C_1 \oplus B_2C_2 \oplus \dots \oplus B_tC_t \oplus 0, & \text{if } \tilde{B} \text{ has no columns.} \end{cases}$$

In both cases the conditions of theorem 4.17 (iii) are satisfied so BC has property C. Since B_iC_i is irreducible, $\rho(B_iC_i) = \rho(BC) = \rho(CB) > 0$, so $CB \neq 0$. It thus follows similarly, by theorem 4.17, that CB also has property C.

Now suppose, conversely, that BC and CB have property C. Then, by lemma 4.25, there exist P_1 and Q_1 such that (89) holds (with P, Q replaced by P_1, Q_1), where $\hat{B}$ and $\hat{C}$ are co-conformable and have no zero rows or columns. Furthermore,

$$\hat{B}\hat{C} = F_1 \oplus F_2 \oplus \dots \oplus F_s,$$

where each F_i is irreducible and $\rho(F_i) = \rho(BC)$. Also, if $\tilde{B}$ and $\tilde{C}$ are non-vacuous $\rho(\tilde{B}\tilde{C}) < \rho(BC)$. Now $\hat{B}$ and $\hat{C}$ satisfy the conditions of lemma 4.5, so there exists R such that

$$\hat{B}R = B_1 \oplus B_2 \oplus \dots \oplus B_s, \quad R^{T}\hat{C} = C_1 \oplus C_2 \oplus \dots \oplus C_s,$$

where B_i and C_i are co-conformable, and $B_iC_i = F_i$. Now B_i and C_i have no zero rows or columns, so C_iB_i is irreducible by lemma 4.4. Letting $P = P_1$, $Q = Q_1(R \oplus I)$, the result then follows.

If B and C are square and *essentially non-singular* (Fenner & Loizou 1977) then all the B_i and all the C_i will also be square and essentially non-singular. This follows from the following corollary.

Corollary 4.27. If B and C are as in theorem 4.26, and, furthermore, they are both square and at least one of them is essentially non-singular, then B_i and C_i are square, for each i.

Proof. Without loss of generality it may be supposed that B is essentially non-singular. Let B be $n \times n$ then, by the Frobenius–König theorem (Mirsky & Perfect 1966), B has no $k \times (n-k+1)$ zero submatrix, $1 \leqslant k \leqslant n$. From the form of (90) it, therefore, follows that, for each i, $B_1 \oplus B_2 \oplus \dots \oplus B_i$ must be square, which yields the result.

For the special case $B = |A|$, $C = |A^{-1}|$, where A is an arbitrary square *non-singular* matrix, a more specific structure for A, than that given by theorem 4.26, can be obtained by using the fact that any non-singular matrix is necessarily essentially non-singular.

Theorem 4.28. Let A be a non-singular matrix, then $|A|\,|A^{-1}|$ has property C if and only if there exist P and Q such that

$$P^{T}AQ = \left[\begin{array}{cccc|c} A_1 & & & & \\ & A_2 & & \mathrm{O} & \\ & & \ddots & & \mathrm{O} \\ & \mathrm{O} & & A_t & \\ \hline & & \mathrm{O} & & \tilde{A} \end{array}\right], \qquad (91)$$

where, for each i, A_i is a square fully indecomposable matrix, $\rho(|A_i|\,|A_i^{-1}|) = \rho(|A|\,|A^{-1}|)$, and $\rho(|\tilde{A}|\,|\tilde{A}^{-1}|) < \rho(|A|\,|A^{-1}|)$ if $\tilde{A}$ is non-vacuous.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Proof. Suppose there exist P and Q such that $P^{\mathrm{T}}AQ$ has the form (91). Then $|A_i|\,|A_i^{-1}|$ and $|A_i^{-1}|\,|A_i|$ are irreducible by lemma 4.2, so theorem 4.26 implies that $|A|\,|A^{-1}|$ has property C.

Conversely, suppose that $|A|\,|A^{-1}|$ has property C. Then, by corollary 4.23, $|A^{-1}|\,|A|$ also has property C, so $|A|$ and $|A^{-1}|$ satisfy the conditions of theorem 4.26. Thus there exist P and Q such that (90) holds with $B = |A|$ and $C = |A^{-1}|$. Furthermore, for each i, B_i and C_i are square by corollary 4.27. Therefore, $P^{\mathrm{T}}AQ$ has the form (91) where $B_i = |A_i|$ for each i, and $\tilde{B} = |\tilde{A}|$. Each A_i, and $\tilde{A}$, must be non-singular, so $C_i = |A_i^{-1}|$ for each i, and $\tilde{C} = |\tilde{A}^{-1}|$. For each i, $|A_i|\,|A_i^{-1}|$ is irreducible so, by lemma 4.2, A_i is fully indecomposable, which completes the proof.

Results for property C^1, analogous to those of theorems 4.26 and 4.28 for property C, are now obtained.

Theorem 4.29. Let B and C be co-conformable non-negative matrices such that $BC \neq 0$. Then BC and CB both have property C^1 if and only if there exist P and Q such that

$$P^{\mathrm{T}}BQ = \left[\begin{array}{ccccc|c} B_1 & & & & & U_1 \\ & B_2 & & \mathrm{O} & & U_2 \\ & & \cdot & & & \cdot \\ & & & \cdot & & \cdot \\ & \mathrm{O} & & & \cdot & \cdot \\ & & & & B_t & U_t \\ \hline & & \mathrm{O} & & & \tilde{B} \end{array}\right], \quad Q^{\mathrm{T}}CP = \left[\begin{array}{ccccc|c} C_1 & & & & & V_1 \\ & C_2 & & \mathrm{O} & & V_2 \\ & & \cdot & & & \cdot \\ & & & \cdot & & \cdot \\ & \mathrm{O} & & & \cdot & \cdot \\ & & & & C_t & V_t \\ \hline & & \mathrm{O} & & & \tilde{C} \end{array}\right], \tag{92}$$

where the B_i, the C_i, $\tilde{B}$ and $\tilde{C}$ satisfy the same conditions as in theorem 4.26.

Proof. The proof is exactly analogous to that of theorem 4.26 but uses theorem 4.15 and lemma 4.24, where theorem 4.26 relies upon theorem 4.17 and lemma 4.25, respectively.

Corollary 4.30. If B and C are as in theorem 4.29 and are, in addition, square and essentially non-singular, then B_i and C_i are square and essentially non-singular, for each i.

Proof. Let B and C be $n \times n$, then neither B nor C can have a $k \times (n-k+1)$ zero submatrix, $1 \leqslant k \leqslant n$. Therefore, since $B_1 \oplus B_2 \oplus \ldots \oplus B_i$ and $C_1 \oplus C_2 \oplus \ldots \oplus C_i$ are co-conformable for each i, it follows from the form of (92) that they must both be square, which yields the result.

Theorem 4.31. Let A be a non-singular matrix, then $|A|\,|A^{-1}|$ has property C^1 if and only if there exist P and Q such that

$$P^{\mathrm{T}}AQ = \left[\begin{array}{ccccc|c} A_1 & & & & & Z_1 \\ & A_2 & & \mathrm{O} & & Z_2 \\ & & \cdot & & & \cdot \\ & & & \cdot & & \cdot \\ & \mathrm{O} & & & \cdot & \cdot \\ & & & & A_t & Z_t \\ \hline & & \mathrm{O} & & & \tilde{A} \end{array}\right],$$

where the A_i and $\tilde{A}$ satisfy the same conditions as in theorem 4.28.

Proof. The proof is exactly analogous to that of theorem 4.28 but uses theorem 4.29, corollary 4.30 and lemma 4.21, where theorem 4.28 relies upon theorem 4.26, corollary 4.27 and corollary 4.23, respectively.

Corresponding results to those of theorems 4.29 and 4.31 hold for BC and CB, or $|A|\,|A^{-1}|$, having property C^r.

5. Main results

This section begins with some simple results which give the values of the minimal condition numbers $\psi(.)$ and $\kappa(.)$ in certain special cases, and bounds on them in other cases.

5.1. *Minimal condition numbers*

Theorem 5.1. If A is a square matrix and has c.b.s.d. then, for any absolute norm $\|.\|$ with property S′,

$$\psi(A) = \rho(|A|).$$

Proof. Since $\|.\|$ is absolute and A has c.b.s.d., by (24), letting $B = |A|$, it is seen that it is sufficient to prove that $\psi(B) = \rho(B)$ for $B \geqslant 0$. Let $\Delta > 0$ have the same dimensions as B, then by (19), for any D,

$$\psi(B) \leqslant \operatorname{lub}(D^{-1}BD) \leqslant \operatorname{lub}(D^{-1}(B+\Delta)\,D).$$

From theorem 3.26 it follows that there exists D such that $\operatorname{lub}(D^{-1}(B+\Delta)\,D) = \rho(B+\Delta)$, so $\psi(B) \leqslant \rho(B+\Delta)$. Thus, letting $\Delta \to 0$ yields the result by (17) and the continuity of the eigenvalues (Ostrowski 1966).

Corollary 5.2. For any square matrix A, and any lub-absolute norm with property S′, $\psi(A) = \rho(|A|)$.

By using (21), theorem 5.1 yields the following:

Corollary 5.3. For any square matrix A, and any absolute norm with property S′,

$$\rho(|A|) \geqslant \psi(A) \geqslant \rho(A).$$

The following results for the two-sided case are slight generalizations of results in Bauer (1963).

Theorem 5.4. If B and C are co-conformable matrices and have c.b.s.d. then, for any pair of absolute norms with property S′,

$$\kappa_{12}(B;\, C) = \rho(|B|\,|C|).$$

Proof. By (25), it is again sufficient to prove the result for $B, C \geqslant 0$. Let $\Delta > 0$ have the same dimensions as B, then by using theorem 3.27 and (19) it follows, as in the proof of theorem 5.1, that $\kappa_{12}(B;\, C) \leqslant \rho((B+\Delta)\,(C+\Delta^{T}))$. So, letting $\Delta \to 0$, the result follows immediately by (18) and the continuity of the eigenvalues.

Corollary 5.5. For any co-conformable matrices B and C, and any lub-absolute pair of norms with property S′, $\kappa_{12}(B;\, C) = \rho(|B|\,|C|)$.

By using (22), theorem 5.4 yields the following corollary:

Corollary 5.6. For any co-conformable matrices B and C, and any pair of absolute norms with property S′,

$$\rho(|B|\,|C|) \geqslant \kappa_{12}(B;\, C) \geqslant \rho(BC).$$

5.2. *Optimal scalability for similarity scaling*

Various necessary and sufficient conditions for optimal scalability in sense (I) are now established. Certain results of Ström (1972) are then shown to be special cases of these, and some of his results are extended.

THEOREM 5.7. Let $\|.\|$ be a norm having properties L and S*, and let A be an $n\times n$ non-negative matrix having property C, then there exists $D\in\mathscr{D}_n$ such that

$$\mathrm{lub}\,(D^{-1}AD) = \rho(A).$$

Proof. If $A = 0$ the result holds trivially, so assume $A \neq 0$. By theorem 4.17 there exists P such that

$$P^{\mathrm{T}}AP = \begin{bmatrix}\hat{A} & 0\\ 0 & \tilde{A}\end{bmatrix},$$

where $\hat{A}$ is a $k\times k$ matrix which is the direct sum of irreducible matrices all having spectral radius $\rho(A)$, and $\rho(\tilde{A}) < \rho(A)$ if $\tilde{A}$ is non-vacuous. So, by corollary 4.20, let $\hat{\boldsymbol{y}}^{\mathrm{T}}$ and $\hat{\boldsymbol{x}}$ be a pair of positive P-vectors of $\hat{A}$, and let

$$\boldsymbol{y}^{\mathrm{T}} = [\hat{\boldsymbol{y}}^{\mathrm{T}}\,\boldsymbol{0}]\,P^{\mathrm{T}},\quad \boldsymbol{x} = P\begin{bmatrix}\hat{\boldsymbol{x}}\\ \boldsymbol{0}\end{bmatrix}. \tag{93}$$

By theorem 3.25 there exists $D = P(\hat{D}\oplus\tilde{D})\,P^{\mathrm{T}}\in\mathscr{D}_n$, where $\tilde{D}$ is an arbitrary matrix in $\mathscr{D}_{n-k}$, such that

$$\boldsymbol{y}^{\mathrm{T}}D\,\|\,D^{-1}\boldsymbol{x}. \tag{94}$$

Now let $\boldsymbol{v}^{\mathrm{T}} = \boldsymbol{y}^{\mathrm{T}}D$, $\boldsymbol{u} = D^{-1}\boldsymbol{x}$, $B = U(\hat{A}\oplus 0)\,U^{-1}$, where $U = D^{-1}P$, then

$$B\boldsymbol{u} = \rho(A)\,\boldsymbol{u},\quad \boldsymbol{v}^{\mathrm{T}}B = \rho(A)\,\boldsymbol{v}^{\mathrm{T}}. \tag{95}$$

Hence, (94) and (95) imply

$$\boldsymbol{v}^{\mathrm{T}}\,\|\,\boldsymbol{u},\quad \boldsymbol{v}^{\mathrm{T}}\,\|\,B\boldsymbol{u},\quad \boldsymbol{v}^{\mathrm{T}}B\,\|\,\boldsymbol{u}. \tag{96}$$

By using property S* it can be seen that $\boldsymbol{v}^{\mathrm{T}}, \boldsymbol{u}$ is a maximizing pair for B, i.e.

$$\mathrm{lub}\,(B) = \frac{\boldsymbol{v}^{\mathrm{T}}B\boldsymbol{u}}{\|\boldsymbol{v}^{\mathrm{T}}\|^{\mathrm{D}}\,\|\boldsymbol{u}\|}.$$

On using (95) and (96), this yields

$$\mathrm{lub}^{UU}(\hat{A}\oplus 0) = \rho(A). \tag{97}$$

If $\tilde{A}$ is vacuous then (97) yields the desired result, otherwise by theorem 5.1 $\psi(P(0\oplus\tilde{A})\,P^{\mathrm{T}}) = \rho(P(0\oplus\tilde{A})\,P^{\mathrm{T}}) = \rho(\tilde{A})$, so there exists $\tilde{D}\in\mathscr{D}_{n-k}$ such that

$$\mathrm{lub}(D^{-1}P(0\oplus\tilde{A})\,P^{\mathrm{T}}D) = \mathrm{lub}^{UU}(0\oplus\tilde{A}) < \rho(A). \tag{98}$$

Thus, by using corollary 3.7 and lemma 3.1, it follows from (97) and (98) that $\mathrm{lub}^{UU}(\hat{A}\oplus\tilde{A}) = \mathrm{lub}(D^{-1}AD) = \rho(A)$. This completes the proof.

By using corollary 4.20 the above proof yields the following corollary:

COROLLARY 5.8. Let $\|.\|$ be an absolute norm having property $\bar{\mathrm{S}}$, and let A be an $n\times n$ non-negative matrix having a pair of positive left and right P-vectors, then there exists $D\in\mathscr{D}_n$ such that $\mathrm{lub}(D^{-1}AD) = \rho(A)$.

It is noted that theorem 5.7 and corollary 5.8 hold for all l_p-norms by virtue of theorem 3.23 and lemma 3.22.

THEOREM 5.9. Let $\|.\|$ be a strongly monotonic norm, and let A be an $n\times n$ non-negative matrix for which there exists $D\in\mathscr{D}_n$ such that

$$\mathrm{lub}(D^{-1}AD) = \rho(A), \tag{99}$$

then A has property C^{l}.

Proof. If $\rho(A) = 0$ then (99) implies that $A = 0$, so A has property C^{l}, trivially. Hence it is assumed that $\rho(A) > 0$.

Let Q be such that $Q^{T}AQ$ is in normal form (Varga 1962), so the diagonal blocks of $Q^{T}AQ$ are either irreducible or 1×1 zero matrices, and all have spectral radius not exceeding $\rho(A)$. Now let V be any diagonal block which is irreducible and has spectral radius $\rho(A) > 0$. Then $Q^{T}AQ$ may be partitioned as

$$Q^{T}AQ = \begin{bmatrix} U_{11} & U_{12} & U_{13} \\ 0 & V & U_{23} \\ 0 & 0 & U_{33} \end{bmatrix}, \tag{100}$$

where U_{11} is either a square submatrix or is vacuous, and similarly for U_{33}.

Now let $D \in \mathscr{D}_n$ satisfy (99), and let $Q^{T}DQ$ be partitioned conformably with (100) as $D_1 \oplus D_2 \oplus D_3$. It is now shown that if U_{12} is non-vacuous then it must be zero. For suppose that $U_{12} \neq 0$, and let $\boldsymbol{x}_2 > \mathbf{0}$ be a right P-vector of V, with $\boldsymbol{x}$, partitioned conformably with (100), defined by $\boldsymbol{x} = Q(\mathbf{0} \oplus \boldsymbol{x}_2 \oplus \mathbf{0})$. Then

$$D^{-1}A\boldsymbol{x} = Q\begin{bmatrix} D_1^{-1}U_{12}\boldsymbol{x}_2 \\ D_2^{-1}V\boldsymbol{x}_2 \\ \mathbf{0} \end{bmatrix} \geqslant Q\begin{bmatrix} \mathbf{0} \\ \rho(A)\, D_2^{-1}\boldsymbol{x}_2 \\ \mathbf{0} \end{bmatrix} = \rho(A)\, D^{-1}\boldsymbol{x}, \tag{101}$$

but the two sides of (101) are not equal as $U_{12} \neq 0$ and $\boldsymbol{x}_2 > \mathbf{0}$. Thus, by the strong monotonicity of the norm, $\|D^{-1}A\boldsymbol{x}\| > \rho(A)\,\|D^{-1}\boldsymbol{x}\|$, which implies that $\operatorname{lub}(D^{-1}AD) > \rho(A)$, contradicting (99).

Therefore U_{12} is either vacuous or zero. Since this is true for every diagonal block of $Q^{T}AQ$ having spectral radius $\rho(A)$, it follows easily that, by permuting the row and column blocks of $Q^{T}AQ$, this matrix can be put into the form (61) with the A_i and $\tilde{A}$ therein satisfying the same conditions as in lemma 4.9. Hence, by theorem 4.15, A has property C^{l}, which completes the proof.

Corollary 5.10. Let $\|.\|$ be a norm such that $\|.\|^{D}$ is strongly monotonic, and let A be an $n \times n$ non-negative matrix for which there exists $D \in \mathscr{D}_n$ such that $\operatorname{lub}(D^{-1}AD) = \rho(A)$, then A has property C^{r}.

Corollary 5.11. Let $\|.\|$ be a norm such that $\|.\|$ and $\|.\|^{D}$ are both strongly monotonic, and let A be an $n \times n$ non-negative matrix for which there exists $D \in \mathscr{D}_n$ such that $\operatorname{lub}(D^{-1}AD) = \rho(A)$, then A has property C.

It is noted that theorem 5.9, and corollaries 5.10 and 5.11 hold for the l_p-norms with $1 \leqslant p < \infty$, $1 < p \leqslant \infty$ and $1 < p < \infty$, respectively.

By virtue of theorem 5.1, combining theorem 5.7 and corollary 5.11 yields a necessary and sufficient condition for a non-negative matrix to be optimally scalable in the similarity sense, (I), with respect to a strongly monotonic norm having properties L and S*, and whose dual norm is also strongly monotonic.

Theorem 5.12. Let $\|.\|$ be a norm having properties L and S*, and such that $\|.\|$ and $\|.\|^{D}$ are both strongly monotonic. A square non-negative matrix is optimally scalable in sense (I) with respect to $\|.\|$ if and only if it has property C.

Corollary 5.13. Let $\|.\|$ be as in theorem 5.12. A square matrix A having c.b.s.d. is optimally scalable in sense (I) with respect to $\|.\|$ if and only if $|A|$ has property C.

Proof. This follows directly from theorem 5.12 by using (23) and (24).

These characterizations for optimal scalability in sense (I) hold for the l_p-norms, $1 < p < \infty$. For $p = 1$ and ∞, optimal scalability for *arbitrary* matrices is characterized by the following theorem.

Theorem 5.14. Let $\|.\| = l_p(.)$, $p = 1$ or ∞. A square matrix A is optimally scalable in sense (I) if and only if $|A|$ has property C^l for $p = 1$, or property C^r for $p = \infty$.

Proof. Suppose first that $p = \infty$. It has been shown by Ström (1972) that there exists $D \in \mathscr{D}$ such that

$$\operatorname{lub}(D^{-1}AD) = \rho(|A|)$$

if and only if there exists $x > 0$ such that $|A|x \leqslant \rho(|A|)\,x$. From corollary 5.2, $\psi(A) = \rho(|A|)$, so the result follows by theorem 4.16. For $p = 1$, the result follows by considering A^H with $p = \infty$. This completes the proof.

It is easy to see that the characterizations of theorem 5.14 also hold for any norm of the form $\|x\| = l_p(D_0 x)$, $p = 1$ or ∞, $D_0 \in \mathscr{D}$. Similarly, for norms of this form with $1 < p < \infty$, corollary 5.13 holds.

Theorems 4.15, 4.16 and 4.17 furnish alternative characterizations to those of theorems 5.14 and 5.12, and, in particular, exhibit the structure of those matrices having properties C^l, C^r and C, respectively.

It has been shown by Ström (1972) that, for $\|.\| = l_\infty(.)$, a matrix A is optimally scalable in the similarity sense if A is the direct sum of irreducible matrices. This result follows immediately from theorem 5.14, since, by theorem 4.17, $|A|$ has property C and therefore property C^r. Obviously this result also holds for $\|.\| = l_1(.)$, and, if A has c.b.s.d., for any l_p-norm, $1 \leqslant p \leqslant \infty$. Ström also showed that a companion matrix,

$$B = \begin{bmatrix} b_1 & b_2 & \cdots & b_{n-1} & b_n \\ -1 & 0 & \cdots & 0 & 0 \\ & -1 & \ddots & & \\ & & \ddots & \ddots & \mathrm{O} \\ & & & \ddots & \ddots \\ & \mathrm{O} & & \ddots & \ddots \\ & & & -1 & 0 \end{bmatrix}, \tag{102}$$

is optimally scalable in sense (I) for $\|.\| = l_\infty^n(.)$. This again follows from theorem 5.14, since $|B|$ has property C^r provided the b_i are not all zero. This is proved in the following lemma.

Lemma 5.15. If B is defined as in (102) then

(i) $|B|$ has property C^l if and only if $b_n \neq 0$.
(ii) $|B|$ has property C^r if and only if $b_i \neq 0$ for some i.
(iii) $|B|$ has property C if and only if $b_n \neq 0$.

Proof. Suppose first that $b_n \neq 0$. By considering $G(B)$ it follows immediately that B is irreducible, so $|B|$ has property C, and thus also properties C^l and C^r.

Now suppose that $b_n = b_{n-1} = \ldots = b_{k+1} = 0$, but $b_k \neq 0$. Partitioning off the first k rows and k columns of B yields

$$|B| = \begin{bmatrix} B_{11} & 0 \\ B_{21} & B_{22} \end{bmatrix},$$

where B_{11} is irreducible, $B_{21} \neq 0$, and $\rho(B_{22}) = 0$. It follows from theorems 4.15, 4.16 and 4.17 that $|B|$ has property C^r, but neither property C^l nor property C.

Lastly, suppose that $b_i = 0$ for all i $(1 \leqslant i \leqslant n)$. Now $\rho(|B|) = 0$, so theorems 4.15, 4.16 and 4.17 imply that $|B|$ has none of the properties C^l, C^r and C. This completes the proof.

Furthermore, in theorem 2 of Ström (1972) it is shown that there exists D such that $\operatorname{lub}(D^{-1}BD) = \rho(|B|)$, for $\|.\| = l_2^m(.)$. In fact this is only true if $b_n \neq 0$, as otherwise the matrix D obtained therein is singular. This result also follows from corollary 5.13 and theorem 5.1 as B has c.b.s.d., and is irreducible for $b_n \neq 0$. The following theorem generalizes the above results.

THEOREM 5.16. Let $\|.\| = l_p^m(.)$ $(1 \leqslant p \leqslant \infty)$. The companion matrix B as defined in (102) is optimally scalable in sense (I) if and only if $b_n \neq 0$ for $p \neq \infty$, or $b_i \neq 0$ for some i $(1 \leqslant i \leqslant n)$ for $p = \infty$.

Proof. This follows directly from lemma 5.15 and either corollary 5.13 or theorem 5.14, as appropriate.

A partial generalization of theorem 5.16 to absolute norms having property $\bar{\mathrm{S}}$ is given by the following theorem.

THEOREM 5.17. Let $\|.\|$ be an absolute norm having property $\bar{\mathrm{S}}$. If $b_n \neq 0$ the companion matrix B is optimally scalable in sense (I) with respect to $\|.\|$.

Proof. Since B has c.b.s.d., and is irreducible if $b_n \neq 0$, the result follows immediately by corollary 5.8 and theorem 5.1.

5.3. *Optimal scalability for two-sided scaling*

In this subsection necessary and sufficient conditions for optimal scalability in sense (IV A) are obtained, and these are then utilized to obtain corresponding results for sense (IV).

The first part of the proof of the following theorem parallels that of lemma I (i) of Bauer (1963).

THEOREM 5.18. Let the pair of norms $\|.\|_1$, $\|.\|_2$ have properties L and S*, and let B and C be $m \times n$ and $n \times m$ non-negative matrices, respectively, for which BC and CB are non-zero and both have property C. Then there exist $D_1 \in \mathscr{D}_m$ and $D_2 \in \mathscr{D}_n$ such that

$$\operatorname{lub}_{12}(D_1 B D_2)\operatorname{lub}_{21}(D_2^{-1} C D_1^{-1}) = \rho(BC). \tag{103}$$

Proof. By lemma 4.25, there exist P and Q such that

$$P^{\mathrm{T}}BQ = \begin{bmatrix} \hat{B} & 0 \\ 0 & \tilde{B} \end{bmatrix}, \quad Q^{\mathrm{T}}CP = \begin{bmatrix} \hat{C} & 0 \\ 0 & \tilde{C} \end{bmatrix},$$

where $\hat{B}$ and $\hat{C}$ are $k \times l$ and $l \times k$ matrices, respectively, $1 \leqslant k \leqslant m$, $1 \leqslant l \leqslant n$, with $\hat{B}\hat{C}$ and $\hat{C}\hat{B}$ both being direct sums of irreducible matrices all having spectral radius $\rho(BC)$, and $\rho(\tilde{B}\tilde{C})$ and $\rho(\tilde{C}\tilde{B})$ less than $\rho(BC)$ if $\tilde{B}$ and $\tilde{C}$ are non-vacuous. So, by corollary 4.20, let $\hat{\boldsymbol{y}}^{\mathrm{T}}$ and $\hat{\boldsymbol{x}}$ be a pair of positive left and right P-vectors of $\hat{B}\hat{C}$, then $\boldsymbol{y}^{\mathrm{T}}$ and $\boldsymbol{x}$ defined by (93) are left and right P-vectors of BC. Now, letting ρ denote $\rho(BC)$, define

$$\boldsymbol{w}^{\mathrm{T}} = (1/\rho)\,\boldsymbol{y}^{\mathrm{T}}B, \quad \boldsymbol{z} = C\boldsymbol{x}, \tag{104}$$

so the zeros of $\boldsymbol{w}$ and $\boldsymbol{z}$ coincide as $\hat{B}$ has no zero columns and $\hat{C}$ no zero rows, by lemma 4.25. It is then easily verified that $\boldsymbol{w}^{\mathrm{T}}$ and $\boldsymbol{z}$ are left and right P-vectors of CB, respectively.

By lemma 3.2 the norms $\|.\|_1$ and $\|.\|_2$ both have property L and are thus absolute. Therefore, by theorem 3.25, there exist $D_1 = P(\hat{D}_1 \oplus \tilde{D}_1)P^{\mathrm{T}} \in \mathscr{D}_m$ and $D_2 = Q(\hat{D}_2 \oplus \tilde{D}_2)Q^{\mathrm{T}} \in \mathscr{D}_n$ – where $\tilde{D}_1$ and $\tilde{D}_2$ are arbitrary matrices in $\mathscr{D}_{m-k}$ and $\mathscr{D}_{n-l}$, respectively – such that

$$\boldsymbol{y}^{\mathrm{T}}D_1^{-1}\|_1 D_1\boldsymbol{x}, \quad \boldsymbol{w}^{\mathrm{T}}D_2\|_2 D_2^{-1}\boldsymbol{z}. \tag{105}$$

Now let

$$B_1 = U(\hat{B} \oplus 0)\, V^{-1}, \quad C_1 = V(\hat{C} \oplus 0)\, U^{-1},$$

where $U = D_1 P$ and $V = D_2^{-1} Q$, then from (104)

$$B_1(D_2^{-1}\boldsymbol{z}) = \rho D_1 \boldsymbol{x}, \quad (\boldsymbol{y}^{\mathrm{T}} D_1^{-1})\, B_1 = \rho \boldsymbol{w}^{\mathrm{T}} D_2,$$

so by (105)

$$\boldsymbol{y}^{\mathrm{T}} D_1^{-1} \parallel_1 B_1(D_2^{-1}\boldsymbol{z}), \quad (\boldsymbol{y}^{\mathrm{T}} D_1^{-1})\, B_1 \parallel_2 D_2^{-1}\boldsymbol{z}.$$

Therefore, property S_{12}^* yields

$$\mathrm{lub}_{12}(B_1) = \frac{(\boldsymbol{y}^{\mathrm{T}} D_1^{-1})\, B_1 (D_2^{-1}\boldsymbol{z})}{\|\boldsymbol{y}^{\mathrm{T}} D_1^{-1}\|_1^{\mathrm{D}}\, \|D_2^{-1}\boldsymbol{z}\|_2} = \frac{\rho \boldsymbol{y}^{\mathrm{T}} \boldsymbol{x}}{\|\boldsymbol{y}^{\mathrm{T}} D_1^{-1}\|_1^{\mathrm{D}}\, \|D_2^{-1}\boldsymbol{z}\|_2}, \tag{106}$$

as (104) implies that $P(\hat{B} \oplus 0)\, Q^{\mathrm{T}}\boldsymbol{z} = \rho \boldsymbol{x}$. Similarly, it follows from (104) and (105) that

$$\boldsymbol{w}^{\mathrm{T}} D_2 \parallel_2 C_1(D_1 \boldsymbol{x}), \quad (\boldsymbol{w}^{\mathrm{T}} D_2)\, C_1 \parallel_1 D_1 \boldsymbol{x},$$

so property S_{21}^* yields

$$\mathrm{lub}_{21}(C_1) = \frac{(\boldsymbol{w}^{\mathrm{T}} D_2)\, C_1 (D_1 \boldsymbol{x})}{\|\boldsymbol{w}^{\mathrm{T}} D_2\|_2^{\mathrm{D}}\, \|D_1 \boldsymbol{x}\|_1} = \frac{\boldsymbol{w}^{\mathrm{T}} \boldsymbol{z}}{\|\boldsymbol{w}^{\mathrm{T}} D_2\|_2^{\mathrm{D}}\, \|D_1 \boldsymbol{x}\|_1}. \tag{107}$$

Thus, by multiplying (106) and (107), and by using (105), it follows that

$$\mathrm{lub}_{12}(B_1)\, \mathrm{lub}_{21}(C_1) = \mathrm{lub}_{12}^{UV}(\hat{B} \oplus 0)\, \mathrm{lub}_{21}^{VU}(\hat{C} \oplus 0) = \rho(BC). \tag{108}$$

If $\tilde{B}$ and $\tilde{C}$ are vacuous then (108) yields the desired result, otherwise, by theorem 5.4,

$$\kappa_{12}(P(0 \oplus \tilde{B})\, Q^{\mathrm{T}};\quad Q(0 \oplus \tilde{C})\, P^{\mathrm{T}}) = \rho(P(0 \oplus \tilde{B}\tilde{C})\, P^{\mathrm{T}}) = \rho(\tilde{B}\tilde{C}),$$

so there exist $\tilde{D}_1 \in \mathscr{D}_{m-k}$ and $\tilde{D}_2 \in \mathscr{D}_{n-l}$ such that

$$\mathrm{lub}_{12}(D_1 P(0 \oplus \tilde{B})\, Q^{\mathrm{T}} D_2)\, \mathrm{lub}_{21}(D_2^{-1} Q(0 \oplus \tilde{C})\, P^{\mathrm{T}} D_1^{-1}) = \mathrm{lub}_{12}^{UV}(0 \oplus \tilde{B})\, \mathrm{lub}_{21}^{VU}(0 \oplus \tilde{C}) < \rho(BC). \tag{109}$$

Now, by using lemma 3.1 and corollary 3.9, it follows from (108) and (109) that there exists $D \in \mathscr{D}_m$ such that

$$\mathrm{lub}_{12}^{UV}(D(\hat{B} \oplus \tilde{B}))\, \mathrm{lub}_{21}^{VU}((\hat{C} \oplus \tilde{C})\, D^{-1}) = \rho(BC).$$

Thus, letting $D_1^* = D_1 P D P^{\mathrm{T}}$,

$$\mathrm{lub}_{12}(D_1^* B D_2)\, \mathrm{lub}_{21}(D_2^{-1} C D_1^{*-1}) = \rho(BC),$$

which yields the result.

Corollary 5.19. Let the pair of norms $\|.\|_1$, $\|.\|_2$ be absolute and have property S^*, and let B and C be as in theorem 5.18. Furthermore, let either BC or CB have a pair of positive left and right P-vectors, then there exist $D_1 \in \mathscr{D}_m$ and $D_2 \in \mathscr{D}_n$ such that

$$\mathrm{lub}_{12}(D_1 B D_2)\, \mathrm{lub}_{21}(D_2^{-1} C D_1^{-1}) = \rho(BC).$$

Corollary 5.20. Let the pair of norms $\|.\|_1$, $\|.\|_2$ be absolute and have property $\bar{\mathrm{S}}$, and let B and C be as in theorem 5.18. Furthermore, let BC and CB both have pairs of positive left and right P-vectors, then there exist $D_1 \in \mathscr{D}_m$ and $D_2 \in \mathscr{D}_n$ such that

$$\mathrm{lub}_{12}(D_1 B D_2)\, \mathrm{lub}_{21}(D_2^{-1} C D_1^{-1}) = \rho(BC).$$

The above corollaries follow from the proof of theorem 5.18, since if BC has a pair of positive P-vectors then $k = m$, by corollary 4.20, or if CB has a pair of positive P-vectors then $l = n$; in either case $\tilde{B}$ and $\tilde{C}$ are both vacuous. From theorem 3.23 and lemma 3.22 it follows that theorem 5.18 and its corollaries hold for the l_p-norms, $\|.\|_1 = l_p^m(.)$, $\|.\|_2 = l_p^n(.)$, for all p.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Theorem 5.21. Let B and C be $m \times n$ and $n \times m$ non-negative matrices, respectively, and let $\|.\|_1$, $\|.\|_2$ be a pair of norms for which there exist $D_1 \in \mathscr{D}_m$ and $D_2 \in \mathscr{D}_n$ such that

$$\text{lub}_{12}(D_1 BD_2)\,\text{lub}_{21}(D_2^{-1} CD_1^{-1}) = \rho(BC), \tag{110}$$

then

(i) if $\|.\|_1$ is strongly monotonic, BC has property C^l,
(ii) if $\|.\|_1^P$ is strongly monotonic, BC has property C^r,
(iii) if $\|.\|_2$ is strongly monotonic, CB has property C^l,
(iv) if $\|.\|_2^P$ is strongly monotonic, CB has property C^r.

Proof. Let $D_1 \in \mathscr{D}_m$ and $D_2 \in \mathscr{D}_n$ be such that (110) holds. By the multiplicativity of lub norms and (16),

$$\text{lub}_{12}(D_1 BD_2)\,\text{lub}_{21}(D_2^{-1} CD_1^{-1}) \geqslant \text{lub}_{11}(D_1 BCD_1^{-1}) \geqslant \rho(BC), \tag{111}$$

$$\text{lub}_{21}(D_2^{-1} CD_1^{-1})\,\text{lub}_{12}(D_1 BD_2) \geqslant \text{lub}_{22}(D_2^{-1} CBD_2) \geqslant \rho(BC), \tag{112}$$

so equation (110) implies equality throughout (111) and (112). Therefore, if $\|.\|_1$ is strongly monotonic, from (111) it follows that BC satisfies the conditions of theorem 5.9 and thus has property C^l. Similarly, if $\|.\|_1^P$ is strongly monotonic, BC has property C^r by corollary 5.10. The corresponding results for CB follow from (112) in a similar manner.

In order to obtain necessary and sufficient conditions for optimal scalability in sense (IV A), it is necessary to look at the requirement of theorem 5.18 that BC and CB be non-zero. If BC and CB both have property C^l then either both are zero or both are non-zero, since if BC is non-zero then $\rho(BC) = \rho(CB) > 0$ by theorem 4.15. This remark is obviously equally valid if BC and CB both have property C^r or, trivially, property C. Furthermore, if $BC = 0$ then it is clear that equation (103) holds if and only if $B = 0$ or $C = 0$. If (103) holds and $BC \neq 0$ then obviously $\rho(BC) > 0$, so $CB \neq 0$. Thus the following necessary and sufficient conditions for optimal scalability in the two-sided sense are obtained.

Theorem 5.22. Let the pair of norms $\|.\|_1$, $\|.\|_2$ have properties L and S*, and both be strongly monotonic with strongly monotonic dual norms. A pair of co-conformable $m \times n$ and $n \times m$ non-negative matrices B and C, respectively, is optimally scalable in sense (IV A) if and only if BC and CB both have property C, and $BC = 0$ only if $B = 0$ or $C = 0$.

Proof. This follows directly from theorems 5.18 and 5.21, and the above remarks, by virtue of theorem 5.4.

Corollary 5.23. Let $\|.\|_1$, $\|.\|_2$ be as in theorem 5.22. A pair of co-conformable $m \times n$ and $n \times m$ matrices B and C, respectively, which have c.b.s.d., is optimally scalable in sense (IV A) if and only if $|B|\,|C|$ and $|C|\,|B|$ both have property C, and $|B|\,|C| = 0$ only if $B = 0$ or $C = 0$.

Proof. This follows directly from theorem 5.22 on using (23) and (25).

It is noted that theorem 5.22 and corollary 5.23 hold for the pair of l_p-norms $l_p^m(.)$, $l_p^n(.)$ for $1 < p < \infty$. The following results deal with the cases $p = 1$ and $p = \infty$.

Theorem 5.24. Let $\|.\|_1 = l_p^m(.)$, $\|.\|_2 = l_p^n(.)$, $p = 1$ or ∞, and let B and C be $m \times n$ and $n \times m$ non-negative matrices, respectively, for which BC and CB are non-zero and both have property C^l if $p = 1$, or property C^r if $p = \infty$. Then there exist $D_1 \in \mathscr{D}_m$ and $D_2 \in \mathscr{D}_n$ such that

$$\text{lub}_{12}(D_1 BD_2)\,\text{lub}_{21}(D_2^{-1} CD_1^{-1}) = \rho(BC).$$

Proof. Suppose first that $p = 1$, so BC and CB both have property C^l. By lemma 4.24, there exist P and Q such that (82) holds, where $\hat{B}$ and $\hat{C}$ are $k \times l$ and $l \times k$ matrices, respectively, with

$\hat{B}\hat{C}$ being the direct sum of irreducible matrices all having spectral radius $\rho(BC)$, and $\rho(\tilde{B}\tilde{C}) < \rho(BC)$ if $\tilde{B}$ and $\tilde{C}$ are non-vacuous. Now let $B^* = P(\hat{B} \oplus \tilde{B})\, Q^T$, $C^* = Q(\hat{C} \oplus \tilde{C})\, P^T$ then, by theorem 4.17, B^* and C^* satisfy the conditions of theorem 5.18 so, as in the proof of theorem 5.18, there exist $D_1 \in \mathscr{D}_m$ and $D_2 \in \mathscr{D}_n$ such that (108) holds, and (109) also holds if $\tilde{B}$ and $\tilde{C}$ are non-vacuous. Since the norms are symmetric, by virtue of (11), equation (108) can be written in the form

$$\text{lub}_{12}(\hat{D}_1 \hat{B} \hat{D}_2 \oplus 0)\, \text{lub}_{21}(\hat{D}_2^{-1} \hat{C} \hat{D}_1^{-1} \oplus 0) = \rho(BC),$$

and similarly for (109) if $\tilde{B}$ and $\tilde{C}$ are non-vacuous.

Now, applying lemma 3.10 to $P^T D_1 B D_2 Q$ and $Q^T D_2^{-1} C D_1^{-1} P$, there exist $\bar{D}_1 \in \mathscr{D}_m$ and $\bar{D}_2 \in \mathscr{D}_n$ such that

$$\text{lub}_{12}(\bar{D}_1 P^T D_1 B D_2 Q \bar{D}_2)\, \text{lub}_{21}(\bar{D}_2^{-1} Q^T D_2^{-1} C D_1^{-1} P \bar{D}_1^{-1}) = \rho(BC). \tag{113}$$

Letting $D_1^* = P\bar{D}_1 P^T D_1$ and $D_2^* = D_2 Q \bar{D}_2 Q^T$, (113) becomes

$$\text{lub}_{12}(P^T D_1^* B D_2^* Q)\, \text{lub}_{21}(Q^T D_2^{*-1} C\, D_1^{*-1} P) = \rho(BC),$$

and the result follows by (11).

For $p = \infty$ the result follows in a precisely analogous manner. This completes the proof.

Corollary 5.25. Let $\|.\|_1 = l_p^m(.)$, $\|.\|_2 = l_p^n(.)$, $p = 1$ or ∞. A pair of co-conformable $m \times n$ and $n \times m$ matrices B and C, respectively, is optimally scalable in sense (IVA) if and only if $|B|\,|C|$ and $|C|\,|B|$ both have property C^l if $p = 1$, or property C^r if $p = \infty$, and $|B|\,|C| = 0$ only if $B = 0$ or $C = 0$.

Proof. This is immediate from theorems 5.21 and 5.24, and the remarks preceding theorem 5.22, by virtue of corollary 5.5, since the pair of norms is lub-absolute.

Clearly, theorem 5.24 and corollary 5.25 extend to pairs of norms of the form

$$\|\boldsymbol{x}\|_1 = l_p^m(D_0 \boldsymbol{x}), \quad \|\boldsymbol{y}\|_2 = l_p^n(D_0^* \boldsymbol{y}), \quad D_0 \in \mathscr{D}_m, \quad D_0^* \in \mathscr{D}_n, \tag{114}$$

for $p = 1$ or ∞. Correspondingly, for these norms with $1 < p < \infty$ theorem 5.22 and corollary 5.23 hold.

The structure of pairs of matrices satisfying corollary 5.23 or corollary 5.25 is exhibited by theorem 4.26 or theorem 4.29, respectively.

If it is assumed in the statement of corollary 5.23 that B and C have no zero rows or columns, then optimal scalability may be characterized in terms of *either* $|B|\,|C|$ *or* $|C|\,|B|$ having property C, by virtue of corollary 4.23. Lemma 4.21 and corollary 4.22 imply a corresponding result for corollary 5.25. Under the same assumption, corollary 4.23 further implies that corollary 5.23 also holds if $\|.\|_1$ and $\|.\|_1^D$ are strongly monotonic, but not necessarily $\|.\|_2$ and $\|.\|_2^D$, or vice versa. Lemma 4.21 and corollary 4.22 imply a similar result if just $\|.\|_1$ and $\|.\|_2^D$ are strongly monotonic.

Now, turning to problem (IV), the two-sided scaling of a non-singular matrix, corollaries 5.23 and 5.25 yield necessary and sufficient conditions for the minimal condition number to be attainable.

Theorem 5.26. Let $\|.\|_1$, $\|.\|_2$ be as in theorem 5.22 with $m = n$. A non-singular matrix A, for which A and A^{-1} have c.b.s.d., is optimally scalable in sense (IV) if and only if $|A|\,|A^{-1}|$ has property C.

Proof. This follows directly from corollaries 5.23 and 4.23 as discussed above.

Corollary 5.27. Let the pair of norms $\|.\|_1$, $\|.\|_2$ be absolute and have property $\bar{\text{S}}$, with $m = n$. Let A be a non-singular matrix for which both A and A^{-1} have c.b.s.d., and $|A|\,|A^{-1}|$ has a pair of positive left and right P-vectors, then A is optimally scalable in sense (IV).

Proof. Since $|A|\,|A^{-1}|$ has a pair of positive left and right P-vectors it has property C by corollary 4.19, and, furthermore, it is easily seen that $|A^{-1}|\,|A|$ also has a pair of positive left and right P-vectors, and thus also has property C. The result then follows by using corollary 5.20.

Again, theorem 5.26 holds for the l_p-norms $\|.\|_1 = \|.\|_2 = l_p^n(.)$, $1 < p < \infty$. For $p = 1$ or ∞ the corresponding result is given by the following theorem.

THEOREM 5.28. Let $\|.\|_1 = \|.\|_2 = l_p(.)$, $p = 1$ or ∞. A non-singular matrix A is optimally scalable in sense (IV) if and only if $|A|\,|A^{-1}|$ has property C^l if $p = 1$, or property C^r if $p = \infty$.

Proof. This follows from corollary 5.25 and lemma 4.21 or corollary 4.22.

It is noted that theorem 2 of Businger (1968) is a special case of theorem 5.28, with his characterization of the optimal D_1 and D_2 following from theorem 5.18 and (9).

Again, theorems 5.26 and 5.28 also hold for pairs of norms of the form (114), with $m = n$.

The structure of those non-singular matrices satisfying theorem 5.26 or 5.28 is given by theorem 4.28 or 4.31, respectively.

6. CHARACTERIZATION OF PROPERTY L

This section, which is devoted to characterizing property L, begins by considering pairs of l_p-norms which have property L_{12}.

LEMMA 6.1. The pair of norms $\|.\|_1 = l_{p_1}^m(.)$, $\|.\|_2 = l_{p_2}^n(.)$ has property L_{12} if and only if $p_1 \geqslant p_2$.

Proof. Since the norms are symmetric, the permutation matrices P and Q may be omitted in the definition of property L_{12}. From the definition of the l_p-norms,

$$\|\boldsymbol{x}_1 \oplus \boldsymbol{x}_2\|_1 = l_{p_1}^m(\boldsymbol{x}_1 \oplus \boldsymbol{x}_2) = l_{p_1}^2((\xi_1\,\xi_2)^{\mathrm{T}}), \tag{115}$$

where $\xi_1 = \|\boldsymbol{x}_1 \oplus \mathbf{0}\|_1 = l_{p_1}^m(\boldsymbol{x}_1 \oplus \mathbf{0})$, $\quad \xi_2 = \|\mathbf{0} \oplus \boldsymbol{x}_2\|_1 = l_{p_1}^m(\mathbf{0} \oplus \boldsymbol{x}_2)$ and, similarly,

$$\|\boldsymbol{y}_1 \oplus \boldsymbol{y}_2\|_2 = l_{p_2}^n(\boldsymbol{y}_1 \oplus \boldsymbol{y}_2) = l_{p_2}^2((\eta_1\eta_2)^{\mathrm{T}}), \tag{116}$$

where $\eta_1 = \|\boldsymbol{y}_1 \oplus \mathbf{0}\|_2$, $\eta_2 = \|\mathbf{0} \oplus \boldsymbol{y}_2\|_2$.

Now suppose first that $\theta = p_1/p_2 \geqslant 1$ and let $\boldsymbol{x}_1 \oplus \boldsymbol{x}_2 \in \mathrm{X}_m$, $\boldsymbol{y}_1 \oplus \boldsymbol{y}_2 \in \mathrm{X}_n$ satisfy

$$\|\boldsymbol{x}_1 \oplus \mathbf{0}\|_1 \leqslant \|\boldsymbol{y}_1 \oplus \mathbf{0}\|_2, \quad \|\mathbf{0} \oplus \boldsymbol{x}_2\|_1 \leqslant \|\mathbf{0} \oplus \boldsymbol{y}_2\|_2. \tag{117}$$

Letting $\boldsymbol{\xi} = (\xi_1\,\xi_2)^{\mathrm{T}}$ and $\boldsymbol{\eta} = (\eta_1\eta_2)^{\mathrm{T}}$, the triangle inequality and (117) imply

$$l_\theta^2(\boldsymbol{\xi}^{p_2}) \leqslant \xi_1^{p_2} + \xi_2^{p_2} \leqslant \eta_1^{p_2} + \eta_2^{p_2}.$$

Hence, by using (115) and (116), it immediately follows that $\|\boldsymbol{x}_1 \oplus \boldsymbol{x}_2\|_1 \leqslant \|\boldsymbol{y}_1 \oplus \boldsymbol{y}_2\|_2$, thus property L_{12} holds.

Supposing conversely that property L_{12} holds, choose $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, with $\boldsymbol{x}_1 \oplus \boldsymbol{x}_2 \in \mathrm{X}_m$ and $\boldsymbol{y}_1 \oplus \boldsymbol{y}_2 \in \mathrm{X}_n$, such that

$$\|\boldsymbol{x}_1 \oplus \mathbf{0}\|_1 = \|\mathbf{0} \oplus \boldsymbol{x}_2\|_1 = \|\boldsymbol{y}_1 \oplus \mathbf{0}\|_2 = \|\mathbf{0} \oplus \boldsymbol{y}_2\|_2 = 1.$$

Hence, by property L_{12}, $\|\boldsymbol{x}_1 \oplus \boldsymbol{x}_2\|_1 \leqslant \|\boldsymbol{y}_1 \oplus \boldsymbol{y}_2\|_2$, which, on using (115) and (116), implies that $2^{1/p_1} \leqslant 2^{1/p_2}$, which yields the result.

COROLLARY 6.2. The pair of norms $l_{p_1}^m(.)$, $l_{p_2}^n(.)$ has property L if and only if $p_1 = p_2$.

Lemma 3.6 suggests a connection between pairs of norms having property L_{12} and a class of norms first introduced by Fiedler & Pták (1960), and investigated subsequently by Stoer

(1964*b*). This connection is examined in this section with the object of characterizing those pairs of norms having property L_{12}.

An *L-norm* (Stoer 1964*b*) is a norm defined on X_n satisfying the following two conditions:

(A 1) for any k, $1 \leqslant k \leqslant n$, and any permutation matrix P,

$$\operatorname{lub}(P(I_k \oplus 0)P^{\mathrm{T}}) \leqslant 1,$$

(A 2) for any $n \times n$ matrix $A_1 \oplus A_2 \oplus \ldots \oplus A_t$, with the A_i all *square*, and any P,

$$\operatorname{lub}(P(A_1 \oplus A_2 \oplus \ldots \oplus A_t)P^{\mathrm{T}}) \leqslant \max_{1 \leqslant j \leqslant t}\{\operatorname{lub}(P(0 \oplus \ldots 0 \oplus A_j \oplus 0 \ldots \oplus 0)P^{\mathrm{T}})\}.$$

(These conditions are expressed in a different notation from that used by Stoer 1964*b*.) It is easy to see that every L-norm is absolute (see Stoer 1964*b*). Conditions (A 1) and (A 2) are now examined.

Remark 6.3. A norm $\|\cdot\|$ satisfies (A 1) if and only if it is subspace monotonic.

Proof. Now, for any P and k, $1 \leqslant k \leqslant n$,

$$\operatorname{lub}(P(I_k \oplus 0)P^{\mathrm{T}}) = \sup_{x_1, x_2} \frac{\|P(x_1 \oplus \mathbf{0})\|}{\|P(x_1 \oplus x_2)\|},$$

where the supremum is taken over all non-zero vectors $x_1 \oplus x_2 \in X_n$, with $x_1 \in X_k$. Hence the result follows immediately from definition 2.10 and (A 1).

Turning now to condition (A 2), it is clear from corollary 3.7, by using lemma 3.1, that a norm having property L satisfies (A 2). As condition (A 2) only involves a single norm, in order to obtain a connection with property L_{12} it is convenient to introduce an extension of condition (A 2) for a pair of norms $\|\cdot\|_1$, $\|\cdot\|_2$:

(A 3) for any $m \times n$ matrix $A_1 \oplus A_2 \oplus \ldots \oplus A_t$, with the A_i all non-vacuous, and any P and Q,

$$\operatorname{lub}_{1\,2}^{PQ}(A_1 \oplus A_2 \oplus \ldots \oplus A_t) \leqslant \max_{1 \leqslant j \leqslant t}\{\operatorname{lub}_{1\,2}^{PQ}(0 \oplus \ldots 0 \oplus A_j \oplus 0 \ldots \oplus 0)\}.$$

It is similarly clear, again from corollary 3.7 and lemma 3.1, that a pair of norms having property L_{12} satisfies (A 3). Property L_{12} is now characterized by the following lemma.

Lemma 6.4. The pair of norms $\|\cdot\|_1$, $\|\cdot\|_2$ has property L_{12} if and only if condition (A 3) is satisfied and $\|\cdot\|_2$ is subspace monotonic.

Proof. The necessity of the conditions follows from the above comments together with remark 2.13. Suppose then that the pair of norms satisfies (A 3) and $\|\cdot\|_2$ is subspace monotonic. Let $x = P(x_1 \oplus x_2) \in X_m$, $y = Q(y_1 \oplus y_2) \in X_n$ satisfy

$$\|x_1 \oplus \mathbf{0}\|_{1P} \leqslant \|y_1 \oplus \mathbf{0}\|_{2Q}, \quad \|\mathbf{0} \oplus x_2\|_{1P} \leqslant \|\mathbf{0} \oplus y_2\|_{2Q}, \tag{118}$$

for some P and Q. If y_1 or y_2 is zero then, correspondingly, x_1 or x_2 is also zero, and it is immediate that $\|x\|_1 \leqslant \|y\|_2$. Suppose, therefore, that y_1 and y_2 are both non-zero and, using remark 2.11, let $(w_1 \oplus \mathbf{0})^{\mathrm{H}}$ and $(\mathbf{0} \oplus w_2)^{\mathrm{H}}$ be dual, with respect to $\|\cdot\|_{2Q}$, to $y_1 \oplus \mathbf{0}$ and $\mathbf{0} \oplus y_2$, respectively, where w_1 and y_1 have the same dimension and so do w_2 and y_2. Furthermore, by suitable scaling, w_1 and w_2 may be chosen such that

$$w_1^{\mathrm{H}} y_1 = \|y_1 \oplus \mathbf{0}\|_{2Q}, \quad w_2^{\mathrm{H}} y_2 = \|\mathbf{0} \oplus y_2\|_{2Q}, \tag{119}$$

and thus $\|(w_1 \oplus \mathbf{0})^{\mathrm{H}}\|_{2Q}^{\mathrm{D}} = \|(\mathbf{0} \oplus w_2)^{\mathrm{H}}\|_{2Q}^{\mathrm{D}} = 1$. Now, letting

$$A_1 = \frac{x_1 w_1^{\mathrm{H}}}{\|y_1 \oplus \mathbf{0}\|_{2Q}}, \quad A_2 = \frac{x_2 w_2^{\mathrm{H}}}{\|\mathbf{0} \oplus y_2\|_{2Q}},$$

it follows from (119) that

$$A_1 \boldsymbol{y}_1 = \boldsymbol{x}_1, \quad A_2 \boldsymbol{y}_2 = \boldsymbol{x}_2. \tag{120}$$

Furthermore, since

$$\|(A_1 \oplus 0)\, \boldsymbol{z}\|_{1P} = \frac{|(\boldsymbol{w}_1 \oplus \mathbf{0})^{\mathrm{H}} \boldsymbol{z}|\, \|\boldsymbol{x}_1 \oplus \mathbf{0}\|_{1P}}{\|\boldsymbol{y}_1 \oplus \mathbf{0}\|_{2Q}},$$

it follows that

$$\operatorname{lub}_{1\,2}^{PQ}(A_1 \oplus 0) = \frac{\|\boldsymbol{x}_1 \oplus \mathbf{0}\|_{1P}}{\|\boldsymbol{y}_1 \oplus \mathbf{0}\|_{2Q}} \sup_{\boldsymbol{z} \neq \mathbf{0}} \frac{|(\boldsymbol{w}_1 \oplus \mathbf{0})^{\mathrm{H}} \boldsymbol{z}|}{\|\boldsymbol{z}\|_{2Q}} \leqslant \|(\boldsymbol{w}_1 \oplus \mathbf{0})^{\mathrm{H}}\|_{2Q}^{\mathrm{D}} = 1, \tag{121}$$

by (118). Similarly,

$$\operatorname{lub}_{1\,2}^{PQ}(0 \oplus A_2) \leqslant 1. \tag{122}$$

Now, by (120),

$$\|\boldsymbol{x}\|_1 = \|A_1 \boldsymbol{y}_1 \oplus A_2 \boldsymbol{y}_2\|_{1P} \leqslant \operatorname{lub}_{1\,2}^{PQ}(A_1 \oplus A_2)\, \|\boldsymbol{y}\|_2,$$

and it then follows from condition (A 3), (121) and (122) that $\|\boldsymbol{x}\|_1 \leqslant \|\boldsymbol{y}\|_2$. Thus property L_{12} holds, which proves the sufficiency of the conditions.

Corollary 6.5. If the pair of norms $\|\cdot\|_1$, $\|\cdot\|_2$ has property L then they are both L-norms.

Proof. By lemma 3.2 each of the norms has property L so, using remark 6.3, the result follows directly from lemma 6.4, since, if $\|\cdot\|_1 = \|\cdot\|_2$, condition (A 3) obviously implies condition (A 2).

It is noted that lemma 6.4 is essentially a generalization and extension of theorem 2 of Lancaster & Farahat (1972). [This theorem is itself quite similar to theorem 2 of Stoer (1964*b*).] From the proof of lemma 6.4, it is clear that the corresponding result for a subspace monotonic norm satisfying condition (A 2) is obtained by imposing the restrictions that $P = Q$, and that $\boldsymbol{x}_1$ and $\boldsymbol{y}_1$ have the same dimension, in the definition of property L. Thus, with remark 6.3, this yields

Corollary 6.6. A norm $\|\cdot\|$ is an L-norm if and only if for any P, and any $\boldsymbol{x} = P(\boldsymbol{x}_1 \oplus \boldsymbol{x}_2)$, $\boldsymbol{y} = P(\boldsymbol{y}_1 \oplus \boldsymbol{y}_2) \in \mathrm{X}_n$ with $\boldsymbol{x}_1, \boldsymbol{y}_1 \in \mathrm{X}_k$, for some k, $1 \leqslant k \leqslant n-1$,

$$\|\boldsymbol{x}_1 \oplus \mathbf{0}\|_P \leqslant \|\boldsymbol{y}_1 \oplus \mathbf{0}\|_P \quad \text{and} \quad \|\mathbf{0} \oplus \boldsymbol{x}_2\|_P \leqslant \|\mathbf{0} \oplus \boldsymbol{y}_2\|_P \quad \Rightarrow \quad \|\boldsymbol{x}\| \leqslant \|\boldsymbol{y}\|.$$

By using a result of Stoer (1964*b*), a complete characterization of property L is now presented.

Theorem 6.7. The pair of norms $\|\cdot\|_1$, $\|\cdot\|_2$ has property L if and only if there exist $D_1 \in \mathscr{D}_m$, $D_2 \in \mathscr{D}_n$ such that

(i) if $m = n = 2$, $\|\cdot\|_{1D_1} = \|\cdot\|_{2D_2}$, and this norm is absolute and symmetric,

(ii) if m or $n > 2$, $\|\cdot\|_{1D_1} = l_p^m(\cdot)$, $\|\cdot\|_{2D_2} = l_p^n(\cdot)$ for some p, $1 \leqslant p \leqslant \infty$.

Proof. Suppose first that the pair of norms has property L, then by corollary 6.5 they are both L-norms. It is proved in Stoer (1964*b*) that a norm $\|\cdot\|$ on X_n is an L-norm if and only if, for $n = 2$, it is absolute or, for $n > 2$, there exists $D \in \mathscr{D}_n$ such that $\|\cdot\|_D = l_p^n(\cdot)$ for some p, $1 \leqslant p \leqslant \infty$.

From this result, if $m > 2$ and $n > 2$, (ii) follows immediately by corollary 6.2, on using lemma 3.1.

Consider now the case when $m = 2$ and $n > 2$. Then $\|\cdot\|_{2D_2} = l_p^n(\cdot)$, for some p and some D_2. Now define D_1 so that $\|\boldsymbol{e}_j\|_{1D_1} = 1$ for $j = 1, 2$. Let $\boldsymbol{x} = (x_1 x_2)^{\mathrm{T}}$ be any vector in X_2, and choose $\boldsymbol{y} = \boldsymbol{y}_1 \oplus \boldsymbol{y}_2 \in \mathrm{X}_n$ such that

$$\|\boldsymbol{y}_1 \oplus \mathbf{0}\|_{2D_2} = l_p^n(\boldsymbol{y}_1 \oplus \mathbf{0}) = |x_1|, \quad \|\mathbf{0} \oplus \boldsymbol{y}_2\|_{2D_2} = l_p^n(\mathbf{0} \oplus \boldsymbol{y}_2) = |x_2|. \tag{123}$$

By the definition of D_1, $\|(x_1 0)^{\mathrm{T}}\|_{1D_1} = |x_1|$, $\|(0 x_2)^{\mathrm{T}}\|_{1D_1} = |x_2|$, and thus by remark 2.14, on using lemma 3.1, equations (123) imply that $\|\boldsymbol{x}\|_{1D_1} = \|\boldsymbol{y}\|_{2D_2}$. Hence, from the definition of the l_p-norms and (123),

$$\|\boldsymbol{x}\|_{1D_1} = \|\boldsymbol{y}\|_{2D_2} = l_p^n(\boldsymbol{y}) = l_p^2(|\boldsymbol{x}|) = l_p^2(\boldsymbol{x}),$$

yielding the desired result for this case. By symmetry the case $m > 2$, $n = 2$ follows in a similar manner.

Next consider the case $m = n = 2$. By the quoted result of Stoer, the norms $\|.\|_1$ and $\|.\|_2$ are both absolute. Let D_1 and D_2 be such that

$$\|e_j\|_{1D_1} = \|e_j\|_{2D_2} = 1, \quad \text{for} \quad j = 1, 2.$$

Then, for any $x = (x_1 x_2)^T \in X_2$, and any 2×2 permutation matrix P,

$$\|(x_1 0)^T\|_{1D_1} = \|P(x_1 0)^T\|_{2D_2} = |x_1|, \quad \|(0 x_2)^T\|_{1D_1} = \|P(0 x_2)^T\|_{2D_2} = |x_2|.$$

So, by remark 2.14 and lemma 3.1, $\|x\|_{1D_1} = \|Px\|_{2D_2}$, whence the norm $\|.\|_{1D_1}$ is equal to the norm $\|.\|_{2D_2}$ and is symmetric.

Now suppose conversely that there exist D_1 and D_2 such that (i) or (ii) holds. If m or $n > 2$ then, by using corollary 6.2 and lemma 3.1, (ii) implies that the pair $\|.\|_1$, $\|.\|_2$ has property L. If $m = n = 2$, for any 2×2 permutation matrices P and Q, consider any two vectors, $x = P(x_1 x_2)^T$, $y = Q(y_1 y_2)^T$, satisfying

$$\|P(x_1 0)^T\|_{1D_1} \leqslant \|Q(y_1 0)^T\|_{2D_2}, \quad \|P(0 x_2)^T\|_{1D_1} \leqslant \|Q(0 y_2)^T\|_{2D_2}.$$

By condition (i), these inequalities imply that $|x| \leqslant |y|$, so $\|x\|_{1D_1} \leqslant \|y\|_{2D_2}$ by the monotonicity of the norm. Hence, by lemma 3.1, the pair $\|.\|_1$, $\|.\|_2$ has property L, which completes the proof.

Corollary 6.8. The norm $\|.\|$ has property L if and only if there exists $D \in \mathscr{D}_n$ such that

(i) if $n = 2$, $\|.\|_D$ is absolute and symmetric,

(ii) if $n > 2$, $\|.\|_D = l_p^n(.)$ for some p, $1 \leqslant p \leqslant \infty$.

Corollary 6.9. A norm on X_n, $n > 2$, has property L if and only if it is an L-norm.

It is observed that, in the proof of theorem 5.7, instead of the requirement that the norm $\|.\|$ have property L, it would be sufficient if it were an L-norm, since, apart from the fact that the norm is absolute, the only use made of property L is when $P = Q$ and the submatrices involved are square. Thus condition (A 2) would suffice in place of condition (A 3). However, corollary 6.9 implies that, in fact, the only possible relaxation of the conditions in theorem 5.7 is for the case $n = 2$, when the condition that the norm $\|.\|_D$ be symmetric, for some $D \in \mathscr{D}_2$, may be dropped. These comments apply similarly to theorem 5.12 and corollary 5.13.

7. Concluding remarks

In §5 necessary and sufficient conditions for optimal scalability in senses (I), (IV A) and (IV) have been obtained for particular classes of matrices and norms. In the following discussion attention is mainly confined to case (I), the similarity scaling, however, it is easily seen that it is also applicable to the two-sided scaling, with appropriate minor modifications.

For suitable norms, corollary 5.13 indicates that the necessary and sufficient condition for a *c.b.s.d. matrix* A to be optimally scalable in sense (I) is that $|A|$ has *property* C. The relevant class of norms comprises those norms having properties L and S* which are strongly monotonic and have strongly monotonic duals. However, from theorem 5.1 and corollary 5.11, it is seen that property C is *necessary* for optimal scalability for a larger class of norms, since it is then only required that the norm have property S′ and satisfy the strong monotonicity conditions. Conversely, by theorem 5.7, $|A|$ having property C is a *sufficient* condition for optimal scalability if the norm just has properties L and S*. From corollary 6.9 it is seen that, for norms on X_n with $n > 2$, a norm $\|.\|$ has property L if and only if $\|x\| = l_p(D_0 x)$ for some p, $1 \leqslant p \leqslant \infty$, and some $D_0 \in \mathscr{D}$. Therefore, it is only for this class of norms, with $1 < p < \infty$, that the characterization of optimal scalability given by corollary 5.13 is applicable. It seems likely that, for $n > 2$, it is only for this class of norms with $1 \leqslant p \leqslant \infty$ that $|A|$ having property C is a sufficient condition for optimal scalability, since theorem 5.7 depends critically on theorem 3.25 and corollary 3.7. As noted previously, the conditions of theorem 3.25 can be relaxed so that it is only required that the norm be *orthant-monotonic* (Gries & Stoer 1967), and it is easy to show that orthant-monotonicity implies subspace monotonicity. Thus it seems reasonable to conjecture that subspace monotonicity is *necessary* for theorem 3.25 to hold. Additionally, the result of corollary 3.7 is equivalent to condition (A 3) of § 6, so by lemma 6.4 the dependence of theorem 5.7 on theorem 3.25 and corollary 3.7 would be essentially equivalent to requiring property L. Although it appears likely from the discussion so far that property L is necessary for $|A|$ having property C to be a sufficient condition for optimal scalability, it is possible that there may be norms without property L for which this still holds. The reasons for this are, first, that there might be norms which are *not* subspace monotonic, but for which theorem 3.25 still holds. Secondly, the use made of corollary 3.7 in theorem 5.7 is in somewhat special circumstances, in that the deduction that $\mathrm{lub}\,(\hat{A} \oplus \tilde{A}) = \max\{\mathrm{lub}\,(\hat{A} \oplus 0), \mathrm{lub}\,(0 \oplus \tilde{A})\}$ is only required when $\rho(\tilde{A}) < \rho(\hat{A})$ and $\hat{A}$ is the direct sum of irreducible matrices all having the same spectral radius.

Corresponding comments apply to the two-sided scaling cases, (IV A) and (IV), for c.b.s.d. matrices B and C, or A and A^{-1}, respectively. Again in these cases, for m or $n > 2$, it seems likely that it is only for the class of norms defined by (114), with $1 \leqslant p \leqslant \infty$, that $|B|\,|C|$ and $|C|\,|B|$, or $|A|\,|A^{-1}|$, having property C is a sufficient condition for optimal scalability. This is suggested by theorem 6.7 which implies that a pair of norms has property L if and only if they are of the form (114) with $1 \leqslant p \leqslant \infty$.

As mentioned above, it seems unlikely that property C is a sufficient condition for optimal scalability (in senses (I), (IV A) and (IV)), except for the classes of norms discussed above. However, it follows from corollaries 5.8, 5.20, and 5.27 that if $|A|$, $|B|\,|C|$ and $|C|\,|B|$, or $|A|\,|A^{-1}|$, have pairs of positive left and right P-vectors then optimal scalability in sense (I), (IV A) or (IV), respectively, is attainable, provided only that the relevant norm, or pair of norms, is absolute and has property $\bar{\mathrm{S}}$.

All the comments of this section so far are also applicable to *non*-c.b.s.d. matrices, or pairs of matrices, provided the norm, or pair of norms, is additionally *lub-absolute*.

It is seen that either property S′ or property $\bar{\mathrm{S}}$ is involved in all of the necessary and the sufficient conditions for optimal scalability discussed above. Moreover, theorems 5.1 and 5.4 only furnish the minimal condition numbers for norms or pairs of norms having property S′. Therefore, it would be interesting to have characterizations of the norms which have these properties (as was obtained for property L in § 6) or at least to establish whether there are norms for which these properties hold, other than the norms of the form (114).

In conclusion, in this paper, optimal scalability in senses (I), (IV A) and (IV) has been completely characterized for arbitrary matrices and norms of the form $l_p(D_0 x)$, $D_0 \in \mathscr{D}$, or pairs of norms of the form (114), respectively, with $p = 1$ or ∞. For $1 < p < \infty$, optimal scalability with respect to these norms, or pairs of norms, has been characterized for matrices, or pairs of matrices, having c.b.s.d. It remains an open problem to obtain corresponding characterizations of optimal scalability with respect to these norms for arbitrary matrices. In particular, it would be highly desirable to obtain such a characterization for the spectral norm, i.e. the case $p = 2$.

References

Bauer, F. L. & Householder, A. S. 1961 Absolute norms and characteristic roots. *Num. Math.* **3**, 241–246.

Bauer, F. L., Stoer, J. & Witzgall, C. 1961 Absolute and monotonic norms. *Num. Math.* **3**, 257–264.

Bauer, F. L. 1962 Optimal scaling of matrices and the importance of the minimal condition. *IFIP-Congress* (Munich), pp. 198–201.

Bauer, F. L. 1963 Optimally scaled matrices. *Num. Math.* **5**, 73–87.

Bauer, F. L. 1966 Genauigkeitsfragen bei der lösung linearer Gleichungssysteme. *Z. angew. Math. Mech.* **46**, 409–421.

Bauer, F. L. 1968 Some aspects of scaling invariance. *Colloq. Int. C.N.R.S.* no. 165, pp. 37–47.

Bauer, F. L. 1969 Remarks on optimally scaled matrices. *Num. Math.* **13**, 1–3.

Businger, P. A. 1968 Matrices which can be optimally scaled. *Num. Math.* **12**, 346–348.

Businger, P. A. 1969 Extremal properties of balanced tri-diagonal matrices. *Maths Comput.* **23**, 193–196.

Dulmage, A. L. & Mendelsohn, N. S. 1958 Coverings of bipartite graphs. *Can. J. Math.* **10**, 517–534.

Fadeeva, V. N. 1967 Some extremum problems for matrix norms. *Zh. vȳchisl. Mat. mat. Fiz.* **7**, 401–404.

Fenner, T. I. 1977 Some results on condition numbers and matrix scaling. Ph.D. Thesis, University of London (to be submitted).

Fenner, T. I. & Loizou, G. 1971 On fully indecomposable matrices. *J. comput. Syst. Sci.* **5**, 607–622.

Fenner, T. I. & Loizou, G. 1974 Some new bounds on the condition numbers of optimally scaled matrices. *J. Ass. comput. Mach.* **21**, 514–524.

Fenner, T. I. & Loizou, G. 1977 Combinatorial aspects of rectangular non-negative matrices. *Discrete Math.* (in the Press).

Fiedler, M. & Pták, V. 1960 Some inequalities for the spectrum of a matrix. *Mat.-fyz. Čas.* **10**, 148–166.

Forsythe, G. E. & Straus, E. G. 1955 On best conditioned matrices. *Proc. Am. math. Soc.* **6**, 340–345.

Golub, G. H. & Varah, J. M. 1974 On a characterization of the best l_2-scaling of a matrix. *SIAM J. numer. Anal.* **11**, 472–479.

Gries, D. 1967 Characterizations of certain classes of norms. *Num. Math.* **10**, 30–41.

Gries, D. & Stoer, J. 1967 Some results on fields of values of a matrix. *SIAM J. numer. Anal.* **4**, 283–300.

Harary, F. 1962 A graph theoretic approach to matrix inversion by partitioning. *Num. Math.* **4**, 128–135.

Harary, F., Norman, R. Z. & Cartwright, D. 1965 *Structural models: An introduction to the theory of directed graphs.* New York: Wiley.

Hartfiel, D. J. & Maxson, C. J. 1975 The chainable matrix, a special combinatorial matrix. *Discrete Math.* **12**, 245–256.

Householder, A. S. 1964 *The theory of matrices in numerical analysis.* New York: Blaisdell.

Lancaster, P. & Farahat, H. K. 1972 Norms on direct sums and tensor products. *Maths Comput.* **26**, 401–414.

Lewin, M. 1971 On non-negative matrices. *Pacif. J. Math.* **36**, 753–759.

Marcus, M. & Minc, H. 1963 Disjoint pairs of sets and incidence matrices. *Illinois J. Math.* **7**, 137–147.

Marcus, M. & Minc, H. 1964 *A survey of matrix theory and matrix inequalities.* Boston: Allyn and Bacon.

McCarthy, C. & Strang, G. 1973 Optimal conditioning of matrices. *SIAM J. numer. Anal.* **10**, 370–388.

Mirsky, L. & Perfect, H. 1966 Systems of representatives. *J. math. Analysis Applic.* **15**, 520–568.

Osborne, E. E. 1960 On pre-conditioning of matrices. *J. Ass. comput. Mach.* **7**, 338–345.

Ostrowski, A. M. 1966 *Solution of equations and systems of equations.* New York: Academic Press.

Stoer, J. & Witzgall, C. 1962 Transformations by diagonal matrices in a normed space. *Num. Math.* **4**, 158–171.

Stoer, J. 1964*a* On the characterization of least upper bound norms in matrix space. *Num. Math.* **6**, 302–314.

Stoer, J. 1964*b* A characterization of Hölder norms. *J. Soc. ind. appl. Math.* **12**, 634–648.

Sluis, A. van der 1969 Condition numbers and equilibration of matrices. *Num. Math.* **14**, 14–23.

Ström, T. 1972 Minimization of norms and logarithmic norms by diagonal similarities. *Computing* **10**, 1–7.

Todd, J. 1968 On condition numbers. *Colloq. Int. C.N.R.S.* no. 165, pp. 141–159.

Varga, R. S. 1962 *Matrix iterative analysis.* Englewood Cliffs, N.J.: Prentice-Hall.

Wilson, R. J. 1972 *Introduction to graph theory.* Edinburgh: Oliver & Boyd.

Zenger, C. 1968 On convexity properties of the Bauer field of values of a matrix. *Num. Math.* **12**, 96–105.

MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A