Zero-diagonality as a linear structure

Jan R. Magnus[a], Enrique Sentana[b]

[a]*Department of Econometrics and Data Science,*
Vrije Universiteit Amsterdam and Tinbergen Institute
[b]*CEMFI, Madrid*

Abstract

A linear structure is a family of matrices that satisfy a given set of linear restrictions, such as symmetry or diagonality. We add to the literature on linear structures by studying the family of matrices where all diagonal elements are zero, and discuss econometric examples where these results can be fruitfully applied.

Keywords: Diagonality, Networks, Restricted matrices, Spatial econometric models, Structural vector autoregressions

JEL Classification: C65, C21, C32

Corresponding author: Enrique Sentana, CEMFI, Casado del Alisal 5, 28014 Madrid, Spain. E-mail: `sentana@cemfi.es`

---

1. Introduction

A linear structure (L-structure) is a family of matrices of given order that satisfy a specific set of linear restrictions. For square matrices (the most common case) examples are symmetry, skew-symmetry, (strict) lower triangularity, and diagonality. The general theory of L-structures was developed in Magnus (1988), hereafter M88, with applications to solving systems of equations and optimization involving patterned matrices. This theory also plays a role in the estimation of multivariate models in which linearly restricted matrices appear. The purpose of this note is to complement these results by investigating the properties of an L-structure not studied so far: zero-diagonal matrices.

Matrices with zero diagonals arise naturally in networks. Consider the following social interactions model for n individuals:

$$y = \rho W y + \beta x + \gamma W x + e, \tag{1}$$

where y and x contain the values of the endogenous and exogenous variables, W is the so-called adjacency matrix whose ij-th element indicates the strength of the link from unit i to unit j, e contains the structural residuals, and ρ, β, and γ are parameters (de Paula et al., 2018). The diagonal elements of an adjacency matrix are zero because loops are not allowed in directed networks.

Adjacency-type matrices also arise in spatial models (Anselin, 1988), some of which are formally identical to (1) except that observations correspond to geographically located variables rather than to individuals.

Matrices with restricted diagonals are also important in structural vector autoregressions. Consider the n-variate time-series process

$$y_t = \Phi y_{t-1} + J\Psi\xi_t, \tag{2}$$

where $\xi_t|(y_{t-1}, y_{t-2}, \dots) \sim i.i.d.\ (0, I_n)$, Ψ is a diagonal matrix whose elements contain the free scale of the structural shocks, and the columns of J, whose diagonal elements are normalized to 1, measure the relative effects of each of the structural shocks on all the observed variables. The pattern of the impulse response functions is completely determined by Φ and J, while the diagonal matrix Ψ only affects their scale. Although J is not a linear structure, the matrix $J - I_n$ is because its diagonal elements are all zero. Lanne et al. (2017) provide sufficient conditions for the identification of the shocks and the free elements of J and Ψ in non-Gaussian models.

In this note we discuss in Section 2 some general features of an L-structure and define the operator and basis for the vector space associated with zero-diagonality. In Section 3 we briefly review the L-structure of diagonal matrices, and in Section 4 we develop the new L-structure of zero-diagonal matrices. All proofs are in the appendix.

2. Linear structures and basis matrices

Consider a real $n \times n$ matrix $A = (a_{ij})$, restricted by some linear constraints, for example $a_{ij} = a_{ji}$ (symmetry), $a_{ij} = -a_{ji}$ (skew-symmetry), or

$a_{ii} = 0$ (zero-diagonality). The collection of matrices A of a given order that satisfy a specific set of linear restrictions constitutes an L-structure.

Let $\operatorname{vec} A$ denote the $n^2 \times 1$ vector containing the columns of A, one underneath the other, and let $\psi(A)$ denote the vector containing only the 'essential' elements of the L-structure. For example, for a symmetric or lower triangular matrix, $\psi(A)$ is the $n(n+1)/2 \times 1$ vector containing the lower triangular elements of A, ordered as $\operatorname{vec} A$ but with some elements removed; a vector commonly denoted by $\operatorname{vech}(A)$.

In this note two L-structures and their corresponding ψ-vectors will be considered, depending on whether A is

- diagonal ($a_{ij} = 0$ for $i \neq j$): $\psi_d(A)$ of dimension $m_d = n$, or
- zero-diagonal ($a_{ii} = 0$): $\psi_o(A)$ of dimension $m_o = n(n-1)$.

When $n = 3$ the relevant ψ-vectors are

$$\psi_d(A) = \begin{pmatrix} a_{11} \\ a_{22} \\ a_{33} \end{pmatrix}, \qquad \psi_o(A) = \begin{pmatrix} a_{21} \\ a_{31} \\ a_{12} \\ a_{32} \\ a_{13} \\ a_{23} \end{pmatrix}.$$

Given $\psi(A)$, the corresponding basis matrices (denoted by Δ) are defined implicitly by

- $\Delta_d \psi_d(A) = \operatorname{vec} A$ for any diagonal matrix A, and
- $\Delta_o \psi_o(A) = \operatorname{vec} A$ for any zero-diagonal matrix A.

The basis matrices Δ_d and Δ_o are of order $n^2 \times n$ and $n^2 \times n(n-1)$, respectively, and for $n = 3$ they take the form

$$\Delta_d = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \Delta_o = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The two basis matrices are 'elimination' matrices because the columns are linearly independent and each column contains one single 1 and zeros elsewhere. They eliminate certain elements of the matrix: for diagonal matrices all non-diagonal elements are eliminated; for zero-diagonal matrices all diagonal elements. Elimination matrices thus correspond to exclusion restrictions and therefore play a fundamental role in deriving the score and asymptotic variance matrix of the estimated free parameters.

The following properties hold for both L-structures $L(\Delta_i)$, in fact for all eliminations matrices.

Proposition 1. *We have for all elimination matrices Δ_i, including $i = o$ and $i = d$,*
(a) Δ_i has full column rank m_i,
(b) $\Delta_i'\Delta_i = I_{m_i}$,
(c) the Moore-Penrose inverse is $\Delta_i^+ = \Delta_i'$,
(d) $\Delta_i' \operatorname{vec} A = \psi_i(A)$ for all A,
(e) $\Delta_i\Delta_i' \operatorname{vec} A = \operatorname{vec} A$ for all $A \in L(\Delta_i)$,
(f) $\partial \operatorname{vec} A/\partial(\psi_i(A))' = \Delta_i$ for all $A \in L(\Delta_i)$.

Note that *(d)* is valid for all A. In contrast, *(e)* and *(f)* are only valid within the chosen L-structure. In Sections 3 and 4 we also show the effect of $\Delta_d\Delta_d'$ and $\Delta_o\Delta_o'$ on a general matrix A.

Property *(f)* is important in statistical inference where derivatives are required. Suppose, for example, that $d\phi(A) = Q\, d \operatorname{vec} A$ for some scalar function ϕ. If there are exclusion restrictions on A associated with a basis matrix Δ_i, then the derivative of $\phi(A)$ is not Q but $Q\Delta_i$.

Diagonality was discussed in M88 (Chapter 6), but the zero-diagonal L-structure has not been discussed and will be our main interest. Since skew-symmetric matrices are zero-diagonal, all results on zero-diagonal matrices apply also to them.

3. Diagonality

We first review some properties of the diagonal L-structure and present some generalizations. The relevant Δ-matrix can be written as

$$\Delta_d' = (E_{11}, E_{22}, \dots, E_{nn}),$$

where E_{ij} is the $n \times n$ matrix with 1 in the ijth position and zeros elsewhere. We know that $\Delta_d \psi_d(A) = \operatorname{vec} A$ when A is diagonal (the implicit definition), and that for any A, diagonal or not,

$$\Delta_d \Delta'_d \operatorname{vec} A = \Delta_d \psi_d(A) = \operatorname{vec}(\operatorname{dg}(A));$$

see M88, Theorem 7.3(ii), where the matrix $\operatorname{dg}(A)$ is a transformation of the *matrix* A containing only its diagonal elements. In contrast, the matrix $\operatorname{diag}(a)$ is a function of the *vector* a and contains the components of a on its diagonal. Thus,

$$\psi_d\left(\operatorname{diag}(a)\right) = a, \qquad \operatorname{diag}\left(\psi_d(A)\right) = \operatorname{dg}(A).$$

Since the ψ_d operator only affects the diagonal elements of A we have $\psi_d(A) = \psi_d(A')$ and hence $\Delta'_d K = \Delta'_d$, where K is the $n^2 \times n^2$ commutation matrix; see M88, Theorem 7.4(i).

The next result links the ψ_d operator to the Hadamard (or element-by-element) product.

Proposition 2. *Given two matrices A and B both of order $n \times n$, we have*

$$\Delta'_d (A \otimes B) \Delta_d = \Delta'_d (B \otimes A) \Delta_d = A \odot B = B \odot A,$$

and in particular, for any two diagonal matrices Λ_1 and Λ_2,

$$\Delta'_d (\Lambda_1 \otimes \Lambda_2) \Delta_d = \Delta'_d (\Lambda_2 \otimes \Lambda_1) \Delta_d = \Lambda_1 \Lambda_2 = \Lambda_2 \Lambda_1.$$

The first expression in Proposition 2 applies also to rectangular matrices A and B, as long as A and B have the same order, but the Δ_d matrices that pre- and postmultiply will then have different orders.

We conclude this short review by presenting a generalization of Proposition 2 and Theorem 7.7(i) of M88.

Proposition 3. *Let*

$$M = \sum_{i=1}^{n} \sum_{j=1}^{n} \mu_{ij} (E_{ii} \otimes E_{jj})$$

be a diagonal $n^2 \times n^2$ matrix with diagonal elements μ_{ij}. Then,

$$\Delta'_d M \Delta_d = \operatorname{diag}(\mu_{11}, \mu_{22}, \ldots, \mu_{nn})$$

and

$$\Delta_d \Delta'_d M \Delta_d = M \Delta_d.$$

4. Zero-diagonality

Our main interest is the class of zero-diagonal matrices which is complementary to the class of diagonal matrices.

Proposition 4. *The $n^2 \times n^2$ matrix (Δ_o, Δ_d) is orthogonal.*

This implies that $\Delta_o' \Delta_d = 0$ and

$$\Delta_o \Delta_o' = I_{n^2} - \Delta_d \Delta_d',$$

a diagonal idempotent matrix of rank $n(n-1)$.

From the implicit definition we have $\Delta_o \Delta_o' \operatorname{vec} A = \operatorname{vec} A$ for all zero-diagonal A, while for any $n \times n$ matrix A,

$$\Delta_o \Delta_o' \operatorname{vec} A = \operatorname{vec}(A - \operatorname{dg}(A)).$$

Next we study the matrix $\Delta_o' K \Delta_o$, where K is the commutation matrix. While K has the effect $K \operatorname{vec} A = \operatorname{vec} A'$ for any A, the matrix $\Delta_o' K \Delta_o$ has the effect $\Delta_o' K \Delta_o \psi_o(A) = \psi_o(A')$ for any zero-diagonal A. It thus plays the role of the commutation matrix for zero-diagonal matrices.

Proposition 5. *The matrix $\Delta_o' K \Delta_o$ is a symmetric permutation matrix of order $n(n-1)$, and hence orthogonal.*

The next two results are used to obtain the asymptotic variance matrix of the ML estimators of the free elements of J and Ψ in the structural vector autoregression model (2) discussed in the Introduction. Proposition 6 is needed to prove Proposition 7, and more generally, in the derivation of the information matrix.

Proposition 6. *Let M be the diagonal $n^2 \times n^2$ matrix defined in Proposition 3. Then,*

$$\Delta_o' M \Delta_o = \operatorname{diag}(\mu_{ij} | i \neq j) = \operatorname{diag}(\mu_{12}, \mu_{13}, \ldots, \mu_{n,n-1}), \qquad \Delta_o' M \Delta_d = 0,$$

and

$$\Delta_o \Delta_o' M \Delta_o = M \Delta_o.$$

In particular, for any $n \times n$ diagonal matrix Λ,

$$\Delta_o' (\Lambda \otimes I_n) \Delta_o = \Lambda \otimes I_{n-1}.$$

In the last expression, the order of the two diagonal matrices cannot be reversed in general. So it is not true that $\Delta_o'(I_n \otimes \Lambda)\Delta_o = I_{n-1} \otimes \Lambda$, unless $\Lambda = I_n$.

Proposition 7. *Let M be the diagonal $n^2 \times n^2$ matrix defined in Proposition 3. If $K + M$ is nonsingular, then*

$$(\Delta_o, \Delta_d)'(K+M)^{-1}(\Delta_o, \Delta_d) = \begin{pmatrix} (\Delta_o'(K+M)\Delta_o)^{-1} & 0 \\ 0 & (I + \Delta_d' M \Delta_d)^{-1} \end{pmatrix}$$

and

$$(K+M)^{-1} = \Delta_o \left(\Delta_o'(K+M)\Delta_o\right)^{-1} \Delta_o' + \Delta_d (I + \Delta_d' M \Delta_d)^{-1} \Delta_d'.$$

We can be more precise about the nonsingularity of $K + M$. For any scalar α_{ij}, let $A_{ij} = \alpha_{ij}E_{ij} + E_{ji}$. The set of vectors $\{\operatorname{vec} A_{ij}\}$ spans $\mathbb{R}^{n^2}$ provided $\alpha_{ii} \neq -1$ and $\alpha_{ij} + \alpha_{ji} = 0$ $(i \neq j)$. Also,

$$(K+M)\operatorname{vec} A_{ij} = (\alpha_{ij}\mu_{ji} + 1)\operatorname{vec} E_{ij} + (\alpha_{ij} + \mu_{ij})\operatorname{vec} E_{ji},$$

and hence $(K+M)\operatorname{vec} A_{ij} = \lambda_{ij}\operatorname{vec} A_{ij}$ if and only if

$$\alpha_{ij}\mu_{ji} + 1 = \alpha_{ij}\lambda_{ij}, \qquad \alpha_{ij} + \mu_{ij} = \lambda_{ij}.$$

The eigenvalues of $K + M$ are $\lambda_{ii} = 1 + \mu_{ii}$ $(i = 1, \ldots, n)$, $\lambda_{ij} = p_{ij} + q_{ij}$ $(i < j)$, and $\lambda_{ij} = p_{ij} - q_{ij}$ $(i > j)$, where

$$p_{ij} = \frac{\mu_{ij} + \mu_{ji}}{2}, \qquad q_{ij} = \sqrt{\left(\frac{\mu_{ij} - \mu_{ji}}{2}\right)^2 + 1}.$$

The matrix is nonsingular if and only if $\mu_{ii} \neq -1$ $(i = 1, \ldots, n)$ and $\mu_{ij}\mu_{ji} \neq 1$ for $i \neq j$.

Appendix: Proofs

Proof of Proposition 1: This follows from M88, Theorems 2.3 and 2.4. Properties (a), (e), and (f) are valid for all basis matrices. Properties (b) and (d) follow from the fact that the columns of an elimination matrix are selected columns of the identity matrix. Property (c) follows from (b).

Proof of Proposition 2: See M88, Theorem 7.7(ii).

Proof of Proposition 3: Let $\Delta_d' = (E_{11}, E_{22}, \dots, E_{nn})$, e_i the ith elementary vector so that $E_{ii} = e_i e_i'$, and M_i the ith diagonal block (of order $n \times n$) of M. Then,

$$\begin{aligned}\Delta_d' M \Delta_d &= \sum_{i=1}^{n} E_{ii} M_i E_{ii} = \sum_{i=1}^{n} e_i e_i' M_i e_i e_i' = \sum_{i=1}^{n} \mu_{ii} E_{ii} \\ &= \operatorname{diag}(\mu_{11}, \mu_{22}, \dots, \mu_{nn}).\end{aligned}$$

To prove the second equation, write $M = \sum_{ij} \mu_{ij}(E_{ii} \otimes E_{jj})$ and let A be diagonal. Then,

$$\begin{aligned}\Delta_d \Delta_d' M \Delta_d \psi_d(A) &= \sum_{ij} \mu_{ij} \Delta_d \Delta_d' (E_{ii} \otimes E_{jj}) \Delta_d \psi_d(A) \\ &= \sum_{ij} \mu_{ij} \Delta_d \Delta_d' (E_{ii} \otimes E_{jj}) \operatorname{vec} A = \sum_{ij} \mu_{ij} \Delta_d \Delta_d' \operatorname{vec}(E_{jj} A E_{ii}) \\ &= \sum_{ij} \mu_{ij} \operatorname{vec}(E_{jj} A E_{ii}) = \sum_{ij} \mu_{ij} (E_{ii} \otimes E_{jj}) \operatorname{vec} A \\ &= \sum_{ij} \mu_{ij} (E_{ii} \otimes E_{jj}) \Delta_d \psi_d(A) = M \Delta_d \psi_d(A),\end{aligned}$$

where we have used the diagonality of $E_{jj} A E_{ii}$. Since this holds for all $\psi_d(A)$, the proof is complete.

Proof of Proposition 4: Since the matrices Δ_d and Δ_o select the diagonal and nondiagonal elements of $\operatorname{vec} A$, respectively, the matrix (Δ_o, Δ_d) is a permutation matrix, hence orthogonal.

Proof of Proposition 5: Since $\Delta_o' K \Delta_o \psi_o(A) = \psi_o(A')$ for any zero-diagonal A, the matrix $\Delta_o' K \Delta_o$ is a permutation matrix. It is symmetric because K is symmetric. Since all permutation matrices are orthogonal, the result follows.

Proof of Proposition 6: We have

$$\Delta_o' M \Delta_o = \sum_{ij} \mu_{ij} \Delta_o' (E_{ii} \otimes E_{jj}) \Delta_o = \sum_{ij} \mu_{ij} \Delta_o' (e_i \otimes e_j)(e_i \otimes e_j)' \Delta_o$$
$$= \sum_{ij} \mu_{ij} \psi_o(E_{ji}) \psi_o(E_{ji})' = \operatorname{diag}(\mu_{12}, \mu_{13}, \ldots, \mu_{n,n-1}).$$

Similarly,

$$\Delta_o' M \Delta_d = \sum_{ij} \mu_{ij} \psi_o(E_{ji}) \psi_d(E_{ji})' = 0,$$

because either $\psi_o(E_{ji}) = 0$ or $\psi_d(E_{ji}) = 0$.

The proof of the third statement is the mirror image of the corresponding statement in Proposition 3. The special case $M = \Lambda \otimes I_n$ follows by noting that $\mu_{ij} = \lambda_i$ for all j, so that

$$\mu_{12}, \mu_{13}, \ldots, \mu_{n,n-1} = \underbrace{\lambda_1, \ldots, \lambda_1}_{n-1 \text{ times}}, \underbrace{\lambda_2, \ldots, \lambda_2}_{n-1 \text{ times}}, \ldots, \underbrace{\lambda_n, \ldots, \lambda_n}_{n-1 \text{ times}}.$$

Proof of Proposition 7: We have

$$(\Delta_o, \Delta_d)'(K+M)(\Delta_o, \Delta_d) = \begin{pmatrix} \Delta_o'(K+M)\Delta_o & \Delta_o' K \Delta_d + \Delta_o' M \Delta_d \\ \Delta_d' K \Delta_o + \Delta_d' M \Delta_o & \Delta_d' K \Delta_d + \Delta_d' M \Delta_d \end{pmatrix}$$
$$= \begin{pmatrix} \Delta_o'(K+M)\Delta_o & 0 \\ 0 & I + \Delta_d' M \Delta_d \end{pmatrix},$$

using the facts that $K\Delta_d = \Delta_d$ (M88, Theorem 7.4(i)), $\Delta_d'\Delta_d = I$ (Proposition 1), $\Delta_d'\Delta_o = 0$ (Proposition 4), and $\Delta_d' M \Delta_o = 0$ (Proposition 6). Since (Δ_o, Δ_d) is orthogonal, the results follow.

Acknowledgements

We are grateful to the editor, associate editor, and two referees for constructive comments; and to Gabriele Fiorentini for useful discussions. Sentana acknowledges financial support from the Spanish Government through grant ECO 2017-89689.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Springer.

de Paula, A., Rasul, I. and Souza, P. (2018). Recovering social networks from panel data: identification, simulations and an application. Working paper CWP17/18, The Institute for Fiscal Studies, UCL.

Lanne, M., Meitz, M. and Saikkonen, P. (2017). Identification and estimation of non-Gaussian structural vector autoregressions. *Journal of Econometrics*, 196, 288–304.

Magnus, J. R. (1988). *Linear structures.* Oxford University Press.