
Dissertations, Theses, and Masters Projects

Theses, Dissertations, & Master Projects

1999

Totally nonnegative matrices

Shaun M. Fallat

College of William & Mary - Arts & Sciences

Follow this and additional works at: <https://scholarworks.wm.edu/etd>



Part of the [Mathematics Commons](#)

Recommended Citation

Fallat, Shaun M., "Totally nonnegative matrices" (1999). *Dissertations, Theses, and Masters Projects*. Paper 1539623948.

<https://dx.doi.org/doi:10.21220/s2-khbp-m642>

This Dissertation is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Dissertations, Theses, and Masters Projects by an authorized administrator of W&M ScholarWorks. For more information, please contact scholarworks@wm.edu.

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI[®]

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

NOTE TO USERS

This reproduction is the best copy available

UMI

Totally Nonnegative Matrices

A Dissertation

Presented to

The Faculty of the Department of Applied Science

The College of William & Mary in Virginia

In Partial Fulfillment

Of the Requirements for the Degree of

Doctor of Philosophy

by

Shaun M. Fallat

April 1999

UMI Number: 9942549

**Copyright 1999 by
Fallat, Shaun Michael**

All rights reserved.

**UMI Microform 9942549
Copyright 1999, by UMI Company. All rights reserved.**

**This microform edition is protected against unauthorized
copying under Title 17, United States Code.**

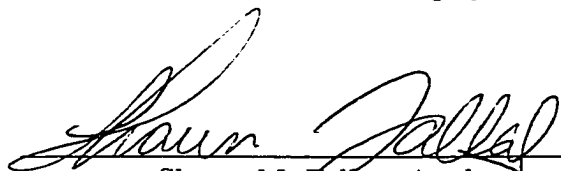
UMI
300 North Zeeb Road
Ann Arbor, MI 48103

APPROVAL SHEET

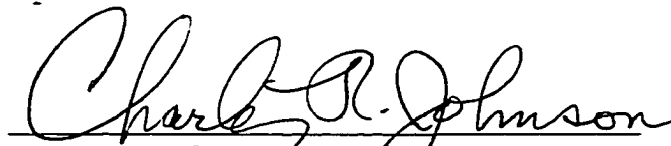
This Dissertation is submitted in partial fulfillment of

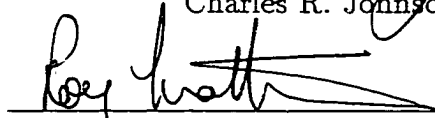
the requirements for the Degree of

Doctor of Philosophy

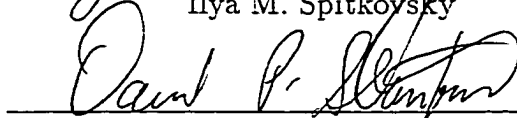

Shaun M. Fallat, Author

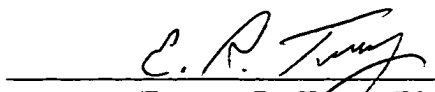
APPROVED. April 1999

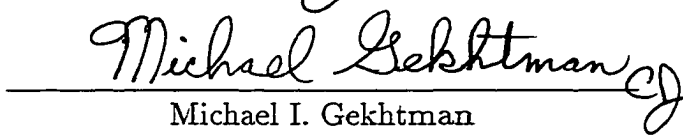

Charles R. Johnson


Roy C. Mathias


Ilya M. Spitkovsky


David P. Stanford


Eugene R. Tracy, Physics


Michael I. Gekhtman
Department of Mathematics
University of Notre Dame

This dissertation is lovingly dedicated to my wife, Darcie, for her unfailing support and encouragement throughout my academic career. You have stood proudly beside me every step of the way, and faithfully supported all of my decisions ...
Good or Bad!

Put simply: You are the secret of my success ...

Contents

Acknowledgments	vi
List of Figures	vii
Abstract	viii
1 Introduction	2
1.1 Notation and Matrix Theoretic Background	2
1.2 Certain Determinantal Identities	4
1.3 Overview	7
2 Preliminary Results and Discussion	14
2.1 Background Results	15
2.2 Factorizations of Totally Nonnegative Matrices	22
2.3 Classical Determinantal Inequalities	29
2.4 Altering Totally Nonnegative Matrices	31
2.5 Zero-Nonzero Patterns of Totally Nonnegative Matrices	36
2.6 Retractibility of Totally Nonnegative Matrices	44

2.7	Row and Column Inclusion	51
2.8	Sub-direct Sums	58
3	Spectral Properties and Jordan Structures	72
3.1	Main Result	76
3.2	Jordan Structures of Totally Nonnegative Matrices	83
4	Determinantal Inequalities	107
4.1	Preliminaries	108
4.2	Operators Preserving Bounded Ratios	120
4.3	Special Ratio Types	129
4.4	General Case for $n \leq 5$	146
4.5	Summary and Further Discussion	159
5	Hadamard Products of Totally Nonnegative Matrices	164
5.1	Notation and General Set-up	166
5.2	Elements of the Hadamard Dual	171
5.3	The 3-by- n Case	175
5.4	(0,+)-Patterns	184
5.5	Oppenheim's Inequality	191
	Bibliography	194
	<i>Vita</i>	202

Acknowledgments

I gratefully acknowledge the assistance of the entire Mathematics Department at the College of William and Mary for their willingness to treat and accept me as a colleague rather than a student. All of you have made this experience a pleasure and I thank you for that. In particular, I wish to thank my advisor, Charles Johnson, for introducing me to the wonderful world of totally nonnegative matrices and his enthusiasm in studying this topic. I would like to express my deepest thanks to my committee members: Roy, Ilya, David and Eugene Tracy for their careful reading of my thesis and subsequent recommendations. Finally, I am grateful to my external examiner, Michael Gekhtman, for his assistance (as some of this dissertation is joint work) and support throughout this entire process.

List of Figures

2.1	Elementary Diagram.	25
2.2	Second Elementary Diagram.	26
2.3	Diagonal Elementary Diagram.	27
2.4	General 3-by-3 Diagram.	27
2.5	General n -by- n diagram.	29
3.1	Fixed Principal Rank One.	87
3.2	Fixed Principal Rank Two.	88
3.3	Principal Rank Two.	98
4.1	General n -by- n diagram.	111
4.2	3-by-3 diagram.	113
4.3	3-by-3 example diagram.	145
4.4	Necessary diagrams for 3-by-3 case.	147
4.5	Particular Diagram	152
4.6	The 4-by-4 Diagrams.	161

Abstract

An m -by- n matrix A is called totally nonnegative (resp. totally positive) if the determinant of every square submatrix (i.e., minor) of A is nonnegative (resp. positive). The class of totally nonnegative matrices has been studied considerably, and this class arises in a variety of applications such as differential equations, statistics, mathematical biology, approximation theory, integral equations and combinatorics. The main purpose of this thesis is to investigate several aspects of totally nonnegative matrices such as spectral problems, determinantal inequalities, factorizations and entry-wise products. It is well-known that the eigenvalues of a totally nonnegative matrix are nonnegative. However, there are many open problems about what other properties exist for the eigenvalues of such matrices. In this thesis we extend classical results concerning the eigenvalues of a totally nonnegative matrix and prove that the positive eigenvalues of an irreducible totally nonnegative matrix are distinct. We also demonstrate various new relationships between the sizes and the number of Jordan blocks corresponding to the zero eigenvalue of an irreducible totally nonnegative matrix. These relationships are a necessary first step to characterizing all possible Jordan canonical forms of totally nonnegative matrices. Another notion investigated is determinantal inequalities among principal minors of totally nonnegative matrices. A characterization of all inequalities that hold among products of principal minors of totally nonnegative matrices up to at most 5 indices is proved, along with general conditions which guarantee when the product of two principal minors is less than another product of two principal minors. A third component of this thesis is a study of entry-wise products of totally nonnegative matrices. In particular, we consider such topics as: closure under this product, questions related to zero/non-zero patterns, and determinantal inequalities associated with this special product. Finally, a survey of classical results and recent developments, including: commonalities and differences among totally nonnegative matrices and other positivity classes of matrices; perturbations and factorizations of totally nonnegative matrices, are discussed.

Totally Nonnegative Matrices

Chapter 1

Introduction

The central theme in this dissertation is to investigate various properties of the class of *totally nonnegative matrices* (see definition to follow). We will begin by providing some notation and preliminary material followed by an overview of this dissertation including brief introductions to each of the (next four) chapters.

1.1 Notation and Matrix Theoretic Background

The set of all m -by- n matrices with entries from a field \mathbb{F} will be denoted by $M_{m,n}(\mathbb{F})$, and if $m = n$, $M_{n,n}(\mathbb{F})$ will be abbreviated to $M_n(\mathbb{F})$. If $\mathbb{F} = \mathbb{R}$, the set of all real numbers, then we may shorten this notation to $M_{m,n}$. For $A \in M_{m,n}(\mathbb{F})$ the notation $A = [a_{ij}]$ will indicate that the entries of A are $a_{ij} \in \mathbb{F}$, for $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$. The transpose of a given m -by- n matrix A will be denoted by A^T .

For $A \in M_{m,n}(\mathbb{F})$, $\alpha \subseteq \{1, 2, \dots, m\}$, and $\beta \subseteq \{1, 2, \dots, n\}$, the submatrix of A

lying in rows indexed by α and the columns indexed by β will be denoted by $A[\alpha|\beta]$. Similarly, $A(\alpha|\beta)$ is the submatrix obtained from A by deleting the rows indexed by α and columns indexed by β . If $A \in M_n(\mathbb{F})$ and $\alpha = \beta$, then the principal submatrix $A[\alpha|\alpha]$ is abbreviated to $A[\alpha]$, and the complementary principal submatrix is $A(\alpha)$. In the same manner, for an n -vector $x \in \mathbb{F}^n$, $x[\alpha]$ denotes the entries of x in the positions indexed by α and $x(\alpha)$ denotes the complementary vector. If $x = [x_i] \in \mathbb{F}^n$, then we let $\text{diag}(x_i)$ denote the n -by- n diagonal matrix with main diagonal entries x_i . For brevity, we often denote the sets $\{1, 2, \dots, m\}$, and $\{1, 2, \dots, n\}$ by M and N , respectively. A *minor* in a given matrix A is by definition the determinant of a (square) submatrix of $A \in M_{m,n}(\mathbb{F})$. For example, the (α, β) -minor of A will be denoted by $\det A[\alpha|\beta]$, and the α -principal minor is naturally denoted by $\det A[\alpha]$, in the case when A is square. For $\alpha = \{i_1, i_2, \dots, i_k\} \subseteq N$, with $i_1 < i_2 < \dots < i_k$, the *dispersion of α* , denoted by $d(\alpha)$, is defined to be $\sum_{j=1}^{k-1} (i_{j+1} - i_j - 1) = i_k - i_1 - (k - 1)$, with the convention that $d(\alpha) = 0$, when α is a singleton. The dispersion of a set α represents a measure of the “gaps” in the set α . In particular, observe that $d(\alpha) = 0$, whenever α is a contiguous subset of N . Also, for $\alpha \subseteq N$, we let α^c denote the set $N \setminus \alpha$, or the complement of α , and we let $|\alpha|$ denote the cardinality of the set α . Note that this means that $A(\alpha|\beta) = A[\alpha^c|\beta^c]$. The i^{th} standard basis vector, that is the n -vector whose only nonzero entry occurs in the i^{th} component and that entry is a one, is denoted by e_i , and the $(i, j)^{\text{th}}$ standard basis matrix, that is the m -by- n matrix whose only nonzero entry is in the $(i, j)^{\text{th}}$ position and this entry is a one, will be denoted by E_{ij} . Observe that if $m = n$, then $E_{ij} = e_i e_j^T$. We also let e denote

the n -vector consisting of all ones (the size of e will be determined from the context). Finally, we let $J_{m,n}$ ($J_{n,n} \equiv J_n$) and I_n denote the m -by- n matrix of all ones and the n -by- n identity matrix, respectively. The subscript is dropped when the sizes of these matrices is clear from the context.

1.2 Certain Determinantal Identities

In this section we list and briefly describe various classical determinantal identities that will be used throughout this dissertation. We begin the classical identity attributed to Jacobi (see, for example, [HJ1]).

Jacobi's Identity: If $A \in M_n(\mathbb{F})$ is nonsingular, then the minors of A^{-1} are related to those of A by Jacobi's identity. Jacobi's identity states that for $\alpha, \beta \subseteq N$, both nonempty, in which $|\alpha| = |\beta|$,

$$\det A^{-1}[\alpha|\beta] = (-1)^s \frac{\det A[\beta^c|\alpha^c]}{\det A}, \quad (1.1)$$

where $s = \sum_{i \in \alpha} i + \sum_{j \in \beta} j$. Observe that if α and β are both singletons, that is $\alpha = \{i\}, \beta = \{j\}$ ($1 \leq i, j \leq n$), then (1.1) becomes

$$a_{ij}^{-1} = (-1)^{i+j} \frac{\det A(\{j\}|\{i\})}{\det A},$$

in which a_{ij}^{-1} denotes the (i, j) entry of A^{-1} . This expression is the classical adjoint formula for the inverse of a matrix, so that Jacobi's identity may be viewed as a generalization of the adjoint formula. When $\alpha = \beta$, (1.1) takes the form

$$\det A^{-1}[\alpha] = \frac{\det A[\alpha^c]}{\det A}. \quad (1.2)$$

We now present some identities discovered by Sylvester (see [HJ1]).

Sylvester's identities: Let $A \in M_n(\mathbb{F})$, $\alpha \subseteq N$, and suppose $|\alpha| = k$. Define the $(n-k)$ -by- $(n-k)$ matrix $B = [b_{ij}]$, with $i, j \in \alpha^c$, by setting $b_{ij} = \det A[\alpha \cup \{i\} | \alpha \cup \{j\}]$, for every $i, j \in \alpha^c$. Then Sylvester's identity states that for each $\delta, \gamma \subset \alpha^c$, with $|\delta| = |\gamma| = l$,

$$\det B[\delta | \gamma] = (\det A[\alpha])^{l-1} \det A[\alpha \cup \delta | \alpha \cup \gamma]. \quad (1.3)$$

Observe that a special case of (1.3) is that $\det B = (\det A[\alpha])^{n-k-1} \det A$. Another very useful special case is the following. Let $A \in M_n(\mathbb{F})$ be the partitioned matrix

$$A = \begin{bmatrix} a_{11} & a_{12}^T & a_{13} \\ a_{21} & A_{22} & a_{23} \\ a_{31} & a_{32}^T & a_{33} \end{bmatrix},$$

where $A_{22} \in M_{n-2}(\mathbb{F})$ and a_{11}, a_{33} are scalars. Define the matrices

$$B = \begin{bmatrix} a_{11} & a_{12}^T \\ a_{21} & A_{22} \end{bmatrix}, \quad C = \begin{bmatrix} a_{12}^T & a_{13} \\ A_{22} & a_{23} \end{bmatrix},$$

$$D = \begin{bmatrix} a_{21} & A_{22} \\ a_{31} & a_{32}^T \end{bmatrix}, \quad E = \begin{bmatrix} A_{22} & a_{23} \\ a_{32}^T & a_{33} \end{bmatrix}.$$

If we let $\tilde{b} = \det B$, $\tilde{c} = \det C$, $\tilde{d} = \det D$, and $\tilde{e} = \det E$, then by (1.3) it follows that

$$\det \begin{bmatrix} \tilde{b} & \tilde{c} \\ \tilde{d} & \tilde{e} \end{bmatrix} = \det A_{22} \det A.$$

Hence, provided $\det A_{22} \neq 0$, we have

$$\det A = \frac{\det B \det E - \det C \det D}{\det A_{22}}.$$

The next determinantal identity relates the minors of a product to the minors of each of the factors.

Cauchy-Binet: Let $A \in M_{m,n}(\mathbb{F})$ and $B \in M_{n,p}(\mathbb{F})$, and suppose $\alpha \subseteq \{1, 2, \dots, m\}$ and $\beta \subseteq \{1, 2, \dots, p\}$ are two index sets each of cardinality k , where $1 \leq k \leq \min(m, n, p)$. Then an expression for the (α, β) -minor of AB is

$$\det AB[\alpha|\beta] = \sum_{\gamma} \det A[\alpha|\gamma] \det B[\gamma|\beta], \quad (1.4)$$

where the sum is taken over all index sets $\gamma \subseteq \{1, 2, \dots, n\}$ of cardinality k .

This useful identity bears a resemblance to the formula for matrix multiplication (and in fact can be thought of as a generalization of matrix multiplication). A consequence of the Cauchy-Binet identity is the multiplicativity of compound matrices

(see chapter 3). Finally, a special case of the above identity is the classical fact that the determinant is multiplicative, i.e., if $A, B \in M_n(\mathbb{F})$, then $\det AB = \det A \det B$.

The final notion we wish to discuss here is the so-called *Schur-complement* of a principal submatrix of a given matrix A (see [HJ1]).

Schur-complements: For $\phi \neq \alpha \subseteq N$ and $A \in M_n(\mathbb{F})$, if $A[\alpha]$ is nonsingular, then the *Schur-complement of $A[\alpha]$ in A* is the matrix $A[\alpha^c] - A[\alpha^c|\alpha](A[\alpha])^{-1}A[\alpha|\alpha^c]$.

Suppose A is partitioned as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

in which A_{11} is k -by- k and nonsingular. Then

$$\begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix} = \begin{bmatrix} A_{11} & 0 \\ 0 & S \end{bmatrix},$$

where S is the Schur-complement of A_{11} in A . Hence it follows that $\det A = \det A_{11} \det S$.

Finally, we note that if $A[\alpha]$ is a nonsingular principal submatrix of a nonsingular matrix A , then $A^{-1}[\alpha^c] = S^{-1}$, where now S is the Schur-complement of $A[\alpha]$ in A .

1.3 Overview

Notions of “positivity” appear in essentially all branches of mathematics, from positive numbers, to positive functions, to positive matrices. In fact, there is now an entire international journal devoted to the theory and applications of positivity in analysis, which is naturally titled – “Positivity.” Linear algebra is no different, here positivity has taken many forms, for example, positive definite matrices, entry-wise positive matrices, positive stability of a matrix, and matrices with certain collections

of positive minors (these ideas are not mutually exclusive). Applications associated with positivity in Linear Algebra vary from economic models to Markov process (in form of stochastic matrices) in statistics, to stability of systems of linear differential equations.

We begin our introduction to positivity in Linear Algebra by briefly discussing positive definite matrices. An n -by- n symmetric ($A^T = A$) matrix A is said to be *positive (semi) definite* if for every $0 \neq x \in \mathbb{R}^n$, $x^T A x > 0$ ($x^T A x \geq 0$). In other words A is positive definite if A is symmetric and has positive quadratic form. The class of positive definite matrices has been well-studied and appears in most textbooks on Linear Algebra and Matrix Theory (see, for example, [HJ1]). Many equivalent conditions are known for a matrix to be positive definite. For example, A is positive definite if and only if A is symmetric and A has positive eigenvalues, or A is positive definite if and only if A is symmetric and has positive principal minors. The next notion of positivity we wish to discuss is *positive stability*, or all eigenvalues having positive real parts. Admittedly, “stability” of systems of linear differential equations usually means that a certain associated matrix has all its eigenvalues in the left-half plane. However, among linear algebraists (and with good reason), stability is usually taken to mean all eigenvalues are in the right-half plane. (We remark here that the connections between these two ideas is obvious.) There are two results which are often associated with stability: the celebrated stability result due to Lyapunov, and the Routh-Hurwitz stability criterion (see also Chapter 5 of this dissertation). We have already seen that positive definite matrices are positive stable, but Lyapunov’s

theorem shows that positive stable matrices are even more intimately related to all positive definite matrices. (Recall that A^* denotes the conjugate transpose of A .)

THEOREM 1.3.1 (*Lyapunov*) *Let $A \in M_n(\mathbb{C})$ be given. Then A is positive stable if and only if there exists a positive definite matrix $G \in M_n(\mathbb{C})$ such that $GA + A^*G$ is positive definite.*

Much work has been done related to and extending Lyapunov's result, see, for example, [HJ2] for a modest discussion along these lines. The next result is known as the Routh-Hurwitz stability criterion (see Chapter 5 for definitions and discussion). Without going into much detail here, this criterion states that: the roots of a given polynomial lie in the left-half plane if and only if a certain matrix (defined in terms of this polynomial) has positive leading minors (principal minors whose index sets are $\{1, 2, \dots, k\}$, for $k = 1, 2, \dots, n$). There is an obvious analog for polynomials to have all roots in the right-half plane.

The next idea of positivity is matrices with positive (or nonnegative) entries, or *entry-wise positive (nonnegative) matrices*. Entry-wise positive matrices have had a long illustrious history in Matrix Theory. Such matrices arise in statistics, often in the form of stochastic matrices, and are usually associated with Markov chains or processes. The actual origin of entry-wise positive matrices in Matrix Theory is unclear, however, the first (and still very important) serious result for entry-wise positive matrices was due to Perron (1907).

THEOREM 1.3.2 (*Perron*) *Suppose A is an n -by- n entry-wise positive matrix.*

Then there exists a simple eigenvalue r , with associated entry-wise positive eigenvector x , such that $r > |\lambda|$, for all eigenvalues λ of A different than r .

This result was significantly extended by Frobenius soon afterwards, and so began the so-called “Perron-Frobenius Theory” of entry-wise nonnegative matrices. Many experts in the field have continued to work in the area of entry-wise positive matrices. One particular topic of interest is *inverse positivity*, that is, an n -by- n invertible matrix has an entry-wise positive inverse. One well-studied subclass of inverse positive matrices is known as the M -matrices. An n -by- n matrix A is said to be an M -matrix if $A = tI - B$, where B is an entry-wise nonnegative matrix and $t > \max\{|\lambda| : \lambda \text{ an eigenvalue of } B\}$. Then M -matrices have nonpositive off-diagonal entries, and are (rather obviously) related to entry-wise nonnegative matrices. Moreover, (see [HJ2]) it is known that any M -matrix is positive stable; has positive principal minors; and is, indeed, inverse positive.

The final topic concerned with positivity (and probably most related to the work in this dissertation) is matrices with certain collections of positive (or nonnegative) minors. We call an n -by- n matrix A a P -matrix (P_0 -matrix) if $\det A[\alpha] > 0$ ($\det A[\alpha] \geq 0$), for every $\alpha \subseteq N$. That is, A is a P -matrix (P_0 -matrix) if every principal minor of A is positive (nonnegative). Observe that positive definite and M -matrices are examples of P -matrices (see [HJ2] for a brief discussion on P -matrices). The matrices of interest here are referred to as totally nonnegative matrices and are contained among the class of P_0 -matrices (in the case of square matrices).

An m -by- n matrix A is said to be *totally nonnegative* (TN) if the determinant of

every square submatrix (i.e., minor) is nonnegative. Similarly, an m -by- n matrix A is *totally positive* (TP) if every minor of A is positive. A well-studied class that lies between the above two classes (in the square case) is the class of oscillatory matrices. An n -by- n matrix A is *oscillatory* if A is totally nonnegative and A^k is totally positive for some $k \geq 1$.

At first it may appear that the notion of total positivity is artificial, however, this class of matrices arises in a variety of applications. For example, totally positive (or nonnegative) matrices can be found in statistics (see [GM, Go, He]), mathematical biology (see [GM]), combinatorics (see [BFZ, B, G3, P2]), dynamics (see [GK2]), approximation theory (see [GM, Pr, CP1]), operator theory (see [So]), and geometry (see [St]). Historically, the theory of totally positive matrices originated from the pioneering work of Gantmacher and Krein ([GK1]) and was elegantly brought together in their monograph [GK2]. Also, under the influence of I. Schoenberg (see, for example, [Sc]), Karlin published an influential treatise, Total Positivity ([K2]), which mostly concerns totally positive kernels but also deals with the discrete version, totally positive matrices. Since then there has been a considerable amount of work done on total positivity, some of which is contained in the exceptional survey paper by T. Ando ([A]). (See also [Ko1, M1, M2] for other general references on total nonnegativity.)

We begin this dissertation with a survey of known results on the theory of totally nonnegative matrices. This survey is complemented with some recent (some of which are new) results on other properties of totally nonnegative matrices. We include results on certain perturbations of totally nonnegative matrices and zero/nonzero

patterns of totally nonnegative matrices. We close chapter 2 with two (new) topics on totally nonnegative matrices. The first is concerned with extending a result for positive semidefinite matrices, known as column (row) inclusion (which offers yet another connection between these two classes). Finally, we consider a new type of “sum” of square matrices, which contains the usual sum and direct sum as special cases. We refer to this sum as a sub-direct sum.

In chapter 3, we seek to extend some classical results of Gantmacher and Krein on the eigenvalues of totally nonnegative matrices. As will be shown, the nonzero eigenvalues of an irreducible totally nonnegative matrix are distinct. We also investigate other issues concerning the eigenvalues of totally nonnegative matrices, by determining all possible Jordan canonical forms of irreducible totally nonnegative matrices. We complete this description through $n = 6$, and include many general related results and interesting open problems.

Recall that the class of totally nonnegative matrices is defined in terms of minors. Along these lines, in chapter 4, we investigate inequalities that hold among products of principal minors for every totally nonnegative matrix. We offer a complete characterization through $n = 5$, and prove that many special types of inequalities hold for all n . For example, we determine necessary and sufficient conditions on the index sets $\alpha_1, \alpha_2, \beta_1, \beta_2 \subseteq N$ so that

$$\det A[\alpha_1] \det A[\alpha_2] \leq \det A[\beta_1] \det A[\beta_2],$$

for every n -by- n totally nonnegative matrix A .

Finally, in chapter 5 we consider the entry-wise (or Hadamard) product of totally nonnegative matrices. We denote this product by \circ . Of particular interest is describing the set of m -by- n matrices A , for which $A \circ B$ is totally nonnegative, whenever B is totally nonnegative. We completely describe this set for $\min(m, n) \leq 3$, and derive many interesting properties about the matrices in this set. Also considered are zero/nonzero patterns of matrices in this set, and a certain determinantal inequality related to Hadamard products of matrices.

Chapter 2

Preliminary Results and Discussion

In this chapter we present a survey of known results both of the classical variety and of the modern sort. This survey will be divided into two basic parts. The first will simply list, in a logical and clear manner, various classical results on the properties of totally nonnegative matrices. Most will be stated without proof, but will be accompanied by numerous references where proofs, etc. may be found. The second component will feature more modern results some of which are novel to this thesis. Accompanying proofs and references will be supplied for the results in this portion. Since many of the results in later chapters rely upon the results contained within this important ground laying chapter, we trust that all readers will benefit from this preparatory discussion.

2.1 Background Results

To begin our study of the class of totally nonnegative matrices we state the following elementary result which contains a list of simple but useful facts, most of which follow directly from the definition of total nonnegativity and various well-known determinantal identities, which are discussed in chapter 1.

PROPOSITION 2.1.1 *[GK1, GK2, A] Suppose A and B are two totally nonnegative (positive) rectangular matrices. Then*

- (i) AB is totally nonnegative (positive), assuming the product exists,*
- (ii) A^T , the transpose of A , is totally nonnegative (positive),*
- (iii) $A[\alpha|\beta]$ is totally nonnegative (positive) for any row index set α and column index set β .*

Note that a special case of (i) above is that pre- and post-multiplication of a totally nonnegative (positive) matrix by a nonnegative (positive) diagonal matrix is totally nonnegative (positive). Another relatively simple but extremely useful result is the following relating the inverse of an invertible totally nonnegative matrix to the class of totally nonnegative matrices (see also [GK1, GK2, A]).

PROPOSITION 2.1.2 *If A is a square invertible totally nonnegative matrix (or totally positive), then $SA^{-1}S$ is totally nonnegative (or totally positive) for $S = \text{diag}(1, -1, 1, -1, \dots, \pm 1)$.*

An immediate consequence of the above result is the next fact about the Schur-complement of totally nonnegative matrices (see also [A]). Recall from chapter 1 that the Schur-complement of an invertible submatrix $A[\gamma]$ in A , denoted by $A/A[\gamma]$, is $A/A[\gamma] = A[\gamma^c] - A[\gamma^c|\gamma](A[\gamma])^{-1}A[\gamma|\gamma^c]$.

PROPOSITION 2.1.3 *If A is a square invertible totally nonnegative matrix (or totally positive), then $A/A[\alpha^c]$, the Schur-complement of $A[\alpha^c]$ in A , is totally nonnegative (or totally positive), for all index sets α with $d(\alpha) = 0$.*

We note here that an auxiliary assumption was made in the above, namely the matrix was assumed to be invertible. However, ‘‘Schur-complements’’ can exist when A is singular, as long as the appropriate principal submatrix is nonsingular. Thus our goal here is to prove a similar result as the proposition above without the assumption that A is nonsingular. Curiously, this next result does not seem to appear in the literature. However, since it does follow rather easily from some classical ideas we consider this result a ‘preliminary’ result and it is probably known to many researchers.

PROPOSITION 2.1.4 *If A is a square totally nonnegative matrix with $A[\alpha^c]$ invertible, then $A/A[\alpha^c]$, the Schur-complement of $A[\alpha^c]$ in A , is totally nonnegative, for all index sets α with $d(\alpha) = 0$.*

Proof. Firstly, assume that $\alpha = \{k + 1, k + 2, \dots, n\}$ and suppose that $A[\alpha^c]$ is invertible. Define the $(n - k)$ -by- $(n - k)$ matrix $C = [c_{ij}]$ by $c_{ij} = \det A[\alpha^c \cup \{i\} | \alpha^c \cup \{j\}]$, where $i, j \in \alpha$. Then Sylvester’s determinantal identity (1.3) can be written as $\det C[\{i_1, \dots, i_s\} | \{j_1, \dots, j_s\}] = (\det A[\alpha^c])^{s-1} \det A[\alpha^c \cup \{i_1, \dots, i_s\} | \alpha^c \cup \{j_1, \dots, j_s\}]$.

From the above identity it follows that $C = (\det A[\alpha^c])(A/A[\alpha^c])$. Moreover, since A is totally nonnegative, by the identity above, C is totally nonnegative and hence $A/A[\alpha^c]$ is totally nonnegative. A similar argument holds for the case when $\alpha = \{1, 2, \dots, k\}$. Thus we have shown that $A/A[\alpha^c]$ is totally nonnegative whenever $\alpha = \{1, 2, \dots, k\}$ or $\alpha = \{k, k+1, \dots, n\}$. The next step is to use the so-called quotient property of Schur-complements. Recall that if $\beta \subseteq \alpha$, then $A/A[\alpha] = [(A/A[\beta])/(A[\alpha]/A[\beta])]$. Suppose α is a fixed index set with $d(\alpha) = 0$ and $|\alpha| < n$. Then we may write α as $\alpha = \{i, i+1, \dots, i+k\}$, where $1 \leq i < i+k \leq n$. Thus $\alpha^c = \{1, 2, \dots, i-1, i+k+1, \dots, n\}$, and define $\beta = \{1, 2, \dots, i-1\}$ if $i > 1$, otherwise $\beta = \{i+k+1, \dots, n\}$ if $i+k < n$. Then by the quotient property described above we have that $A/A[\alpha^c] = [(A/A[\beta])/(A[\alpha^c]/A[\beta])]$. By the previous remarks $A/A[\beta]$ is totally nonnegative, and moreover, since $A[\alpha^c]/A[\beta]$ is a submatrix of $A/A[\beta]$ in the bottom right corner it follows that $(A/A[\beta])/(A[\alpha^c]/A[\beta])$ is totally nonnegative, and hence that $A/A[\alpha^c]$ is totally nonnegative. This completes the proof. ■

We note here that the requirement of $d(\alpha) = 0$ is necessary, as in general Schur-complements of totally nonnegative matrices need not be totally nonnegative.

We now move onto more general results concerning the properties of totally nonnegative matrices. The first of these results is a very useful and appealing criterion (or characterization) of totally positive matrices which was first proved by Fekete [F] in 1913.

THEOREM 2.1.5 (*Fekete's Criterion*) *An m -by- n matrix A is totally positive if and only if $\det A[\alpha|\beta] > 0$, for all $\alpha \subseteq \{1, 2, \dots, m\}$ and $\beta \subseteq \{1, 2, \dots, n\}$, with*

$|\alpha| = |\beta|$ and $d(\alpha) = d(\beta) = 0$.

In other words a matrix is totally positive if and only if the determinant of every square submatrix based on *contiguous* row and column index sets is positive. In the n -by- n case, for example, this reduces the number of minors to check from 4^n to the order of n^4 . Later Cryer [C2], improved upon this result and reduced the complexity to the order of n^3 . More recently, Gasca and Pena [GP4] have even further reduced the number of minors to the order of n^2 . These counts do not take into account the number of operations required to compute the determinant of a given matrix. Unfortunately, Fekete's criterion, in its above form, does not hold in general if "totally positive" is replaced with "totally nonnegative" and " > 0 " is replaced with " ≥ 0 ". Consider the following simple example:

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 0 & 1 \\ 2 & 0 & 1 \end{bmatrix}.$$

It is not difficult to verify that the every minor of A based on contiguous row and column sets is nonnegative, but $\det A[\{1, 3\}] = -3$. Thus A is not totally nonnegative. However, weaker versions (and not as useful) of Fekete's criterion depending on the rank of the matrix do exist for totally nonnegative matrices, see for example [A]. Various generalizations of Fekete's criterion were also obtained in [BFZ] and [FZ] (see also [SS], for other work along these lines). Later we shall describe an efficient method for checking whether a given matrix is totally nonnegative based on triangular factorizations of totally nonnegative matrices (see also [GP4]).

Our next result is a continuity type result which, basically, states that every totally nonnegative matrix can be approximated arbitrarily closely by totally positive matrices. This result may be found in [A, GK2, W], and for a similar result in the case of infinite matrices see [BoP]. See also [C2] for a proof of this fact using certain triangular factorizations that exist for totally nonnegative matrices.

THEOREM 2.1.6 *The closure of the totally positive matrices (in the usual topology on $M_{m,n}$) is the totally nonnegative matrices.*

The next result is concerned with the so-called *variation-diminishing property of totally positive matrices*. This result was proved by Schoenberg [Sc], and is used in many applications. In particular, this result was used to extend some eigenvalue interlacing results for totally positive matrices (see [Pi1]). We begin with some background including both notation and definitions. For a given vector $c = (c_1, c_2, \dots, c_n)^T \in \mathbb{R}^n$ we define two quantities associated with the number of sign changes of the vector c . These are:

$V^-(c)$ — the number of sign changes in the sequence c_1, c_2, \dots, c_n with the zero elements discarded; and

$V^+(c)$ — the maximum number of sign changes in the sequence c_1, c_2, \dots, c_n , where the zero elements are arbitrarily assigned the values $+1$ and -1 .

For example, $V^-((1, 0, 1, -1, 0, 1)) = 2$, and $V^+((1, 0, 1, -1, 0, 1)) = 4$. Schoenberg's variation-diminishing result can now be stated as follows (see also [A]).

THEOREM 2.1.7 *Let A be an m -by- n real matrix with $m \geq n$. Then A is totally positive if and only if $V^+(Ax) \leq V^-(x)$, for all nonzero $x \in \mathbb{R}^n$.*

An important topic concerning totally nonnegative matrices, which will be discussed in much more detail in the next chapter, is the location and distribution of the eigenvalues of such matrices. Similarly, knowledge concerning the associated eigenvectors is also vital to the study of matrix theory. See [K1] for various results concerning the eigenvalues and eigenvectors of oscillatory matrices. Consider the following result which can be found in [GK1, GK2] and also [A].

THEOREM 2.1.8 *Let A be an n -by- n oscillatory matrix. Then the eigenvalues of A are positive, real and distinct. Moreover, the eigenvector x_k corresponding to the k^{th} largest eigenvalue has exactly $k - 1$ variations in sign (i.e., $V^-(x_k) = k - 1$) for $k = 1, 2, \dots, n$.*

Let $a_1 \geq a_2 \geq \dots \geq a_n$ and $b_1 \geq b_2 \geq \dots \geq b_{n-1}$ arranged in decreasing order. Then we say the sequence $\{b_i\}$ *interlaces* the sequence $\{a_i\}$ if $a_n \leq b_{n-1} \leq a_{n-1} \leq \dots \leq b_1 \leq a_1$. Further, if all of the above inequalities can be taken to be strict, we say the sequence $\{b_i\}$ *strictly interlaces* the sequence $\{a_i\}$

This result was originally proved by Gantmacher and Krein [GK1], and can also be found in [G1].

THEOREM 2.1.9 *If A is an n -by- n oscillatory matrix, then the eigenvalues of A strictly interlace the eigenvalues of the two principal submatrices of order $n - 1$ obtained from A by deleting the first row and column or the last row and column.*

In [GK2] it is also shown that these are (in general) the only two submatrices of order $n - 1$ of A for which an interlacing result is guaranteed. The above result was extended slightly by Friedland [Fr], and generalized even further by Pinkus [Pil]. Later, Garloff [G1] used Theorem 2.1.9 to prove a majorization result between the diagonal entries and the eigenvalues of a totally nonnegative matrix. We recall the definition of majorization here. Let $a_1 \geq a_2 \geq \dots \geq a_n$ and $b_1 \geq b_2 \geq \dots \geq b_n$ be two sequences of numbers arranged in decreasing order. We say the sequence $\{a_i\}$ *majorizes* the sequence $\{b_i\}$ if

$$\sum_{i=1}^k b_i \leq \sum_{i=1}^k a_i, \quad k = 1, 2, \dots, n-1, \quad \text{and} \quad \sum_{i=1}^n b_i = \sum_{i=1}^n a_i.$$

THEOREM 2.1.10 *Let $n \geq 2$ and $A = [a_{ij}]$ be an oscillatory matrix. Then the main diagonal entries of A are majorized by the eigenvalues of A .*

Observe that this result parallels the classical result for positive definite matrices (see [HJ1]).

2.2 Factorizations of Totally Nonnegative Matrices

We devote an entire section to factorizations (mostly triangular) of totally nonnegative matrices. This topic is very important in matrix theory and has received much attention in the specific case of totally nonnegative matrices. The list of references for this topic is considerably lengthy, see for example, [A, BFZ, CP1, CP2, C1, C2, FZ, GP1, GP2, GP3, GP4, L, P1, RH, W] for further discussion on factorizations of totally nonnegative matrices

Recall that an n -by- n matrix A is said to have an LU factorization if A can be written as $A = LU$, where L is an n -by- n lower triangular matrix and U is an n -by- n upper triangular matrix. The following remarkable result is, in some sense, one of the most important and useful results in the study of totally nonnegative matrices. This result first appeared in [C1] (although the reduction process proved in [W] is a necessary first step to a LU -factorization result) for nonsingular totally nonnegative matrices and later was proved in [C2] for general totally nonnegative matrices, see also [A] for another proof of this fact.

THEOREM 2.2.1 *Let A be an n -by- n matrix. Then A is totally nonnegative if and only if A has an LU factorization such that both L and U are n -by- n totally nonnegative matrices.*

See also chapter 4 of this thesis where we use factorizations of totally nonnegative matrices in studying certain determinantal inequalities. (In fact LU factorizations of

TN matrices comes up in each subsequent chapter in this thesis.)

There has been a considerable amount of work done on factorizations of totally nonnegative matrices particularly into upper and lower elementary bidiagonal matrices. By definition, an *elementary bidiagonal matrix* is an n -by- n matrix whose main diagonal entries are all equal to one, and there is at most one nonzero off-diagonal entry and this entry must occur on the super- or subdiagonal. To this end, we denote by $E_k(\mu) = [c_{ij}]$ ($2 \leq k \leq n$), the lower elementary bidiagonal matrix whose elements are given by

$$c_{ij} = \begin{cases} 1, & \text{if } i = j, \\ \mu, & \text{if } i = k, j = k - 1, \\ 0, & \text{otherwise,} \end{cases}$$

that is,

$$E_k(\mu) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \mu & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}.$$

Some very useful properties of these elementary bidiagonal matrices are $E_k^{-1}(\mu) = E_k(-\mu)$; $E_k(\mu)E_k(\nu) = E_k(\mu+\nu)$; and $E_k(\mu)E_l(\nu) = E_l(\nu)E_k(\mu)$ except when $|k-l| = 1$, and $\mu\nu \neq 0$. Observe that if $\mu \geq 0$, then $E_k(\mu)$ is totally nonnegative. These facts are interesting exercises and their proofs are omitted. The next result may be found

in many places, but see [GP4] for the version stated here.

THEOREM 2.2.2 *Let A be an n -by- n nonsingular totally nonnegative matrix. Then A can be written as*

$$A = (E_2(l_k))(E_3(l_{k-1})E_2(l_{k-2})) \cdots (E_n(l_{n-1}) \cdots E_3(l_2)E_2(l_1))D \\ (E_2^T(u_1)E_3^T(u_2) \cdots E_n^T(u_{n-1})) \cdots (E_2^T(u_{k-2})E_3^T(u_{k-1}))(E_2^T(u_k)),$$

where $k = \binom{n}{2}$; $l_i, u_j \geq 0$ for all $i, j \in \{1, 2, \dots, k\}$; and D is a positive diagonal matrix.

Note that in the above factorization each elementary bidiagonal matrix is in fact totally nonnegative. If a matrix admits such a factorization we say that A has a *bidiagonal factorization*.

As far as we can figure Theorem 2.2.2 was first stated in [L], where it was used for an application to Lie-group theory. Loewner [L] attributed this result to Whitney [W], although Whitney did not make such a claim in her paper. More accurately, Whitney [W] proved a lemma that is required to prove such a result. More recently, such authors as Gasca, Pena and others considered certain bidiagonal factorizations for which a specified order of row operations is required to ensure a special and appealing form for the factorization. They refer to this particular factorization as a Neville factorization or elimination (see [CP2, GP1, GP2, GP3, GP4]). Cryer [C2] (see also [C1]) seems to be the first to prove that singular totally nonnegative matrices also admit a bidiagonal factorization. See [BFZ] and the sequel [FZ] for an excellent treatment of the combinatorial and algebraic aspects of bidiagonal factorizations.

Some of the graphical representations provided in [BFZ] are employed throughout this thesis, and have been very useful for proving many results (see chapter 4. for example).

We discuss this graphical representation in more detail now. The connection that was observed in [BFZ] and later in [FZ] is a very important and extremely useful one between the products of elementary bidiagonal matrices and combinatorics. To begin, consider the following example. Let

$$L = E_3(l) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & l & 1 \end{bmatrix}.$$

be a 3-by-3 lower elementary bidiagonal matrix. Associated with such bidiagonal matrices is the following weighted diagram (or digraph) see Figure 2.1. This diagram

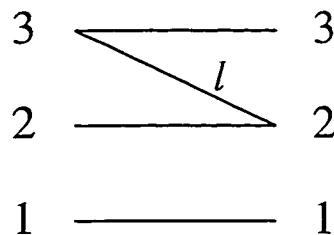


Figure 2.1: Elementary Diagram.

is to be interpreted by reading left to right. All horizontal edges correspond to the main diagonal entries of L and have a weight of 1, whereas the slanted edge corresponds to the (3,2) entry of L (this is the reason it connects the horizontal edges between vertices 3 and 2), and it has weight l . We consider collections of vertex disjoint (or independent) directed paths from a subset of indices on the left (corresponding to the row indices of a given minor) to a (possibly different) subset

of indices on the right (corresponding to the column indices of a given minor). Each such collection constitutes a term in a given minor, and we sum all such terms to compute the minor. For instance, suppose we are given the diagram in Figure 2.1, and we wish to compute the $(3,2)$ entry of the corresponding matrix (which of course is a special minor of the matrix). To do this we consider all the paths initiating at 3 and terminating at 2. Because of the simplicity of this particular diagram there is only one such path. To calculate the corresponding term in the minor (or in this case an entry) simply multiply all the weights of the edges included along this path. In this case the $(3,2)$ entry is equal to $1 \cdot l \cdot 1 = l$. Also, by convention, if there does not exist a desired collection of paths we assign the value 0 to this expression. Such a convention is well-defined since in this case the corresponding minor will be zero. Since all directions are left to right it follows that every other off-diagonal entry (using this diagram) is 0. Finally, observe that each main diagonal entry of the corresponding matrix is equal to 1. Thus we obtain the matrix L . To compute a nontrivial minor of L using this diagram we consider collections of vertex disjoint paths. For example, to compute $\det A[\{1, 3\}|\{1, 2\}]$, we consider pairs (since the minor is of size 2) of vertex disjoint paths initiating at the set $\{1, 3\}$ and terminating at the set $\{1, 2\}$. In this case there exists only one such pair of independent paths with associated term l .

In the case of an upper elementary bidiagonal matrix, say L^T , we have the following associated diagram (see Figure 2.2). A diagonal matrix $D = \text{diag}(d_1, d_2, d_3)$ may be represented by the diagram depicted in Figure 2.3. The real beauty of this device is in the adaptation of it to a product of elementary bidiagonal matrices. We illustrate

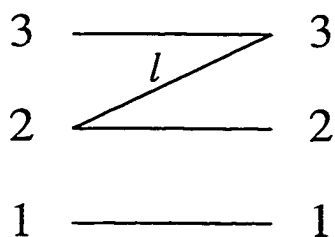


Figure 2.2: Second Elementary Diagram.

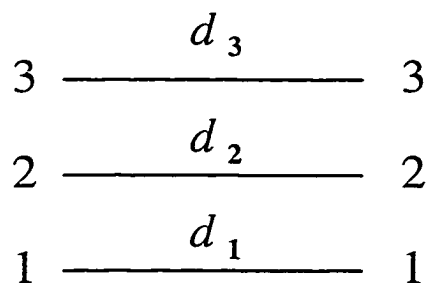


Figure 2.3: Diagonal Elementary Diagram.

this notion with the following example. Suppose A is a 3-by-3 matrix which is factored as $A = E_2(l_3)E_3(l_2)E_2(l_1)DE_2^T(u_1)E_3^T(u_2)E_2^T(u_3)$, where $D = \text{diag}(d_1, d_2, d_3)$. The associated diagram for A is obtained simply by concatenating the diagrams corresponding to each matrix in the factorization according to the order in which they appear. In this case the diagram corresponding to A is given by Figure 2.4. Minors

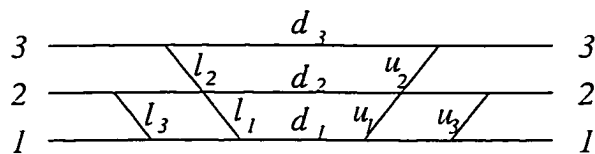


Figure 2.4: General 3-by-3 Diagram.

are computed from the above figure in the same manner as before. For example,

to compute $\det A[\{2, 3\}|\{1, 2\}]$, consider all pairs of independent paths initiating at $\{2, 3\}$ and terminating at $\{1, 2\}$. In order to have non-intersecting paths, the path beginning at 3 must terminate at 2, and the path beginning at 2 must terminate at 1. In this case there exists only one such pair, and hence $\det A[\{2, 3\}|\{1, 2\}] = (l_2 d_2)(l_3 d_1)$. What about the principal minor $\det A[\{2, 3\}]$? Observe that one possible pair of independent paths from $\{2, 3\}$ to $\{2, 3\}$ is the pair that proceeds straight across the diagram, and yields the term $d_2 d_3$. For all other possible pairs of paths we consider two cases:

1. The path beginning at 3 on the left goes straight across the diagram to 3 on the right.
2. The path beginning at 3 on the left uses the edge with weight l_2 .

Case (1): In this case this path yields a weight of d_3 . There are four possibilities for the path beginning and ending at 2 (ignoring the path that goes straight across). These paths yield the weights $d_1 l_3 u_1$, $d_1 l_3 u_3$, $d_1 l_1 u_1$, and $d_1 l_1 u_3$. So in this case the expression for the sum of all of these paths is given by $d_3((l_1 + l_3)(u_1 + u_3))d_1$.

Case (2): For this case the path from 3 to 3 has the weighting $l_2 d_2 u_2$. Notice that in this case there is only one possible path from 2 to 2 which does not intersect the path from 3 to 3, and it has the weighting $l_3 d_1 u_3$, which gives rise to the expression $(l_2 d_2 u_2)(l_3 d_1 u_3)$.

Thus the sum of all of these expression is equal to the minor $\det A[\{2, 3\}]$, that is $\det A[\{2, 3\}] = d_2 d_3 + d_3((l_1 + l_3)(u_1 + u_3))d_1 + (l_2 d_2 u_2)(l_3 d_1 u_3)$.

In the general case we know that any n -by- n nonsingular totally nonnegative matrix admits a bidiagonal factorization (see Theorem 2.2.2). For the factorization given in Theorem 2.2.2 the corresponding diagram is given in Figure 2.5.

Other factorizations have also been considered for totally nonnegative matrices see, for example, [C2, GP3, P1, RH]. The following factorization seems to be useful for various spectral problems for totally nonnegative matrices.

THEOREM 2.2.3 *If A is an n -by- n totally nonnegative matrix, then there exists a totally nonnegative matrix S and a tridiagonal totally nonnegative matrix T such that*

(i) $TS = SA$, and

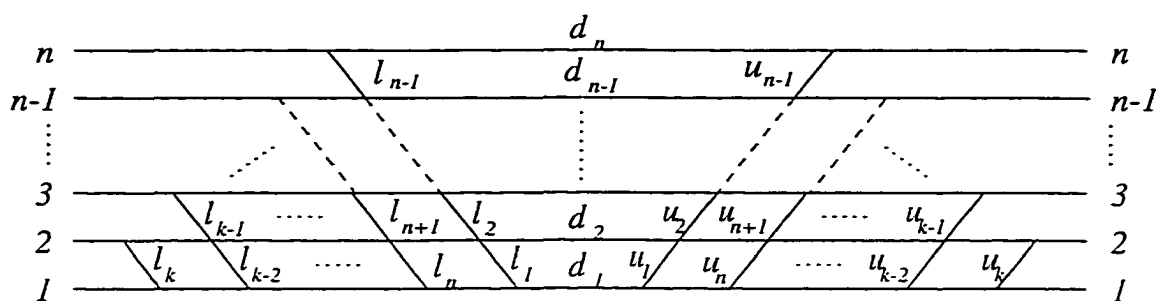
(ii) *the matrices A and T have the same eigenvalues.*

Moreover, if A is nonsingular, then S is nonsingular.

This result was first proved in the nonsingular case by Rainey and Habetler [RH] and the general case was later proved by Cryer [C2]. In chapter 3 we will make use of a slightly modified version of this factorization to aid in characterizing properties about the positive eigenvalues of an irreducible totally nonnegative matrix.

2.3 Classical Determinantal Inequalities

Relationships among principal minors, in particular determining inequalities that exists among products of principal minors, is very important and a well-studied topic for

Figure 2.5: General n -by- n diagram.

many special subclasses of matrices. In the case of positive semidefinite matrices and M -matrices many (classical) inequalities are known to hold and are associated with names like: Hadamard, Fischer and Koteljanskii. Not surprisingly, such inequalities are known to hold for the class of totally nonnegative matrices (although the proofs are vastly different), and we state them here in increasing order of generality.

Let $A = [a_{ij}]$ be an n -by- n totally nonnegative matrix. Then the following inequalities hold:

- Hadamard: ([GK2, A, Ko2])

$$\det A \leq \prod_{i=1}^n a_{ii},$$

- Fischer: ([GK2, A, Ko2]) Let $S \subseteq N = \{1, 2, \dots, n\}$,

$$\det A \leq \det A[S] \cdot \det A[N \setminus S],$$

- Koteljanskii: ([Ko2]) Let $S, T \subseteq N$,

$$\det A[S \cup T] \cdot \det A[S \cap T] \leq \det A[S] \cdot \det A[T].$$

In chapter 4 we investigate further inequalities among principal minors of totally nonnegative matrices.

2.4 Altering Totally Nonnegative Matrices

Often it is useful to classify and understand various operations or alterations which leave a given class of matrices invariant. For example, if A is an n -by- n positive semidefinite matrix, then C^*AC is a positive semidefinite matrix for all n -by- m matrices C . That is, the class of positive semidefinite matrices is closed under congruence (see also [HJ1]). In this section we outline various operations which leave the class of totally nonnegative matrices invariant. We have already discussed at least one such operation, namely matrix multiplication (see Proposition 2.1.1). The first topic we consider here is *insertion*.

PROPOSITION 2.4.1 *Let $A = [a_1, a_2, \dots, a_n]$ be an m -by- n totally nonnegative matrix whose i^{th} column is a_i ($i = 1, 2, \dots, n$). Then for fixed but arbitrary k ($0 \leq k \leq n$), the m -by- $(n+1)$ matrix $\hat{A} = [a_1, \dots, a_k, a_k, a_{k+1}, \dots, a_n]$ is a totally nonnegative matrix.*

Proof. The proof follows by observing that any minor of \hat{A} excluding column k or column $k+1$ is a minor of A , which is nonnegative; otherwise the minor includes both columns k and $k+1$ and hence is zero. ■

A similar statement concerning repeating rows follows from the above proposition and transposition. Furthermore, we can describe properties enjoyed by the set of all column vectors that may be inserted into a totally nonnegative matrix and remain a totally nonnegative matrix.

PROPOSITION 2.4.2 *Let $A = [a_1, a_2, \dots, a_n]$ be an m -by- n totally nonnegative matrix whose i^{th} column is a_i ($i = 1, 2, \dots, n$). Suppose C denotes the set of all column vectors b for which the m -by- $(n + 1)$ matrix $\hat{A} = [a_1, \dots, a_k, b, a_{k+1}, \dots, a_n]$ is a totally nonnegative matrix (here k is fixed but arbitrary). Then C is a nonempty convex cone.*

Proof. By the preceding proposition we have that $C \neq \emptyset$. The remaining claims are elementary exercises using the multi-linearity of the determinant. For example, suppose $b, c \in C$, $t \in (0, 1)$ and consider any square submatrix of $[a_1, \dots, a_k, tb + (1 - t)c, a_{k+1}, \dots, a_n]$, that involves column $k + 1$. Denote this submatrix as follows $[a'_i, \dots, a'_j, tb' + (1 - t)c', a'_l, \dots, a'_r]$. Then $\det[a'_i, \dots, a'_j, tb' + (1 - t)c', a'_l, \dots, a'_r] = t \det[a'_i, \dots, a'_j, b', a'_l, \dots, a'_r] + (1 - t) \det[a'_i, \dots, a'_j, c', a'_l, \dots, a'_r] \geq 0$, since $b, c \in C$. Similar calculations can be used to show that C is a cone. ■

Observe that as a special case inserting a zero column vector preserves the property of being totally nonnegative.

Inserting columns (or rows) and preserving total positivity seems to be a more difficult problem and requires a delicate argument. However, we do have the following result, which, after discussion with experts, seems to be new.

THEOREM 2.4.3 *Let $A = [a_1, a_2, \dots, a_n]$ be an m -by- n totally positive matrix whose i^{th} column is a_i ($i = 1, 2, \dots, n$). Then for fixed but arbitrary k , there exists an n -vector b such that the m -by- $(n + 1)$ matrix $\hat{A} = [a_1, \dots, a_k, b, a_{k+1}, \dots, a_n]$ is a totally positive matrix.*

The proof of this result is complicated and requires a couple of technical lemmas, and can be found in [FJS1]. (Theorem 2.4.3 was also used in conjunction with certain totally positive matrix completion problems, see [FJS2].) Consider the case of nonsingular totally nonnegative matrices (a class that lies between TP and TN). A natural question is: Is it possible to insert a row and a column into an n -by- n nonsingular totally nonnegative matrix and produce an $(n + 1)$ -by- $(n + 1)$ nonsingular totally nonnegative matrix? In general this question is still open, however, in many special cases, such as, direct summing such a matrix with the identity matrix (i.e., the direct sum of A and B is the matrix $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$) the answer is positive. We conjecture that in general the answer to this question is positive, that is, it is always possible to insert a row and column and preserve the property of being an invertible totally nonnegative matrix. Also note that an immediate consequence of Proposition 2.4.2 and Theorem 2.4.3 is that we may border a totally nonnegative or totally positive matrix and remain totally nonnegative or totally positive, respectively.

Also of interest are perturbations which leave a given class invariant. For example, if A is a positive semidefinite matrix, then $A + D$ is positive semidefinite for all nonnegative diagonal matrices D . It is an easy exercise to show this result does not hold in general for the class of totally nonnegative matrices, but a much weaker version does hold.

PROPOSITION 2.4.4 *If A is an m -by- n totally nonnegative matrix, then increasing the $(1,1)$ or the (m,n) entries of A results in a totally nonnegative matrix.*

The proof basically follows from the fact that both the $(1,1)$ and the (m,n) entries enter positively into any square minor that includes them. In general, these are the only two entries that may be arbitrarily increased and leave the class of totally nonnegative matrices invariant. See also a later section (section 2.6) on decreasing certain entries of a totally nonnegative matrix and preserve total nonnegativity.

Let e_i denote the i^{th} standard basis vector (that is, the vector consisting of a 1 in the i^{th} position and zeros otherwise).

COROLLARY 2.4.5 *Let A be an m -by- n totally nonnegative matrix, and let $\delta \geq 0$.*

Then

$$[\delta e_1 | A], [A | \delta e_m], \begin{bmatrix} \delta e_i^T \\ A \end{bmatrix}, \text{ and } \begin{bmatrix} A \\ \delta e_n^T \end{bmatrix},$$

are all totally nonnegative matrices.

The above result follows directly from Propositions 2.4.2 and 2.4.4.

We close this section with a result concerning simultaneous row and column permutations of a totally nonnegative matrix. Recall that such classes as the positive semidefinite, M - and P -matrices are all closed under simultaneous permutation of rows and columns. However, in general, the class of totally nonnegative matrices is not closed under simultaneous permutation of rows and columns. Nevertheless the following result holds (a proof is included for completeness).

PROPOSITION 2.4.6 *Let ρ denote the n -by- n permutation matrix induced by the permutation $i \rightarrow n-i+1$, ($1 \leq i \leq n$), and suppose A is an n -by- n totally nonnegative (positive) matrix. Then $\rho A \rho$ is a totally nonnegative (positive) matrix.*

Proof. Consider an arbitrary minor of $\rho A \rho$. Then by the Cauchy-Binet identity for determinants (see (1.4)),

$$\det \rho A \rho[\alpha|\beta] = \sum_{|\gamma|=|\alpha|} \det \rho A[\alpha|\gamma] \det \rho[\gamma|\beta].$$

By the definition of ρ it follows that the above sum reduces to $\det \rho A[\alpha|\tilde{\beta}] \det \rho[\tilde{\beta}|\beta]$, where $\tilde{\beta}$ is defined by $j \in \tilde{\beta}$ if and only if $n - j + 1 \in \beta$. Another application of Cauchy-Binet reveals that

$$\det \rho A \rho[\alpha|\beta] = \left(\sum_{|\gamma|=|\alpha|} \det \rho[\alpha|\gamma] \det A[\gamma|\tilde{\beta}] \right) \det \rho[\tilde{\beta}|\beta].$$

Similarly, the above sum reduces to $\det \rho[\alpha|\tilde{\alpha}] \det A[\tilde{\alpha}|\tilde{\beta}] \det \rho[\tilde{\beta}|\beta]$, where $\tilde{\alpha}$ is defined similarly as $\tilde{\beta}$. It is not difficult to show that $\det \rho[\gamma|\tilde{\gamma}] = 1$, if $|\gamma| \equiv 0$, or $1 \pmod{4}$, otherwise $\det \rho[\gamma|\tilde{\gamma}] = -1$. Since $|\alpha| = |\beta|$, we have that $\det \rho A \rho[\alpha|\beta] = \det A[\tilde{\alpha}|\tilde{\beta}] \geq 0$ (> 0). Hence $\rho A \rho$ is totally nonnegative (positive). ■

2.5 Zero-Nonzero Patterns of Totally Nonnegative Matrices

In this section we are concerned with zero-nonzero patterns of totally nonnegative matrices. As observed in the previous section inserting a zero row (or column) into a totally nonnegative matrix preserves the property of being totally nonnegative. Thus for the remainder of this section we assume that our matrices do not contain any zero rows or columns. To begin our analysis we make the following definitions.

DEFINITION 2.5.1 An m -by- n *sign pattern* is an m -by- n array of symbols chosen from $\{+, 0, -\}$. A *realization* of a sign pattern, S , is a real m -by- n matrix A such that:

$$a_{ij} > 0 \text{ when } s_{ij} = +; \quad a_{ij} < 0 \text{ when } s_{ij} = -; \quad \text{and } a_{ij} = 0 \text{ when } s_{ij} = 0.$$

We also let $*$ denote a nonzero entry of a matrix whenever a sign is not specified. There are two natural mathematical notions associated with various sign-pattern problems. They are the notions of *require* and *allow*. We say an m -by- n sign pattern S *requires property* P if every realization of S has property P . On the other hand we say a sign pattern S *allows property* P if there exists a realization of S with property P .

An m -by- n matrix is TN_k (TP_k) for $1 \leq k \leq \min(m, n)$, if all minors of size at most k are nonnegative (positive). We call an m -by- n matrix TN_+ if each square

submatrix is either sign-singular (singular by virtue of the its sign-pattern) or has positive determinant.

DEFINITION 2.5.2 An m -by- n matrix is said to be *in double echelon form* if

(i) Each row of A has one of the following forms:

1. $(*, *, \dots, *)$,
2. $(*, \dots, *, 0, \dots, 0)$,
3. $(0, \dots, 0, *, \dots, *)$, or
4. $(0, \dots, 0, *, \dots, *, 0, \dots, 0)$.

(ii) The first and last nonzero entries in row $i + 1$ are not to the left of the first and last nonzero entries in row i , respectively ($i = 1, 2, \dots, n - 1$).

Note that there is an obvious analog of the above definition for m -by- n sign-patterns. Thus a matrix in double echelon form may appear as follows

$$\begin{bmatrix} * & * & 0 & \dots & 0 \\ * & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & * \\ 0 & \dots & 0 & * & * \end{bmatrix}.$$

Recall that increasing the $(1,1)$ or (m,n) entry of an m -by- n totally nonnegative matrix preserves the property of being totally nonnegative. This observation is needed in the proof of the next result.

THEOREM 2.5.3 *Let S be an m -by- n $(0, +)$ -sign pattern in double echelon form.*

Then S allows a TN_+ matrix.

Proof. Since S has no zero rows or columns $s_{11} = +$. Assign the value 1 to all the positive positions in row 1. Consider the left most $+$ in row 2, say $s_{2j} = +$, but $s_{2k} = 0$ for all $k < j$. Let $B = S[\{1, 2\}|\{1, 2, \dots, j\}]$. Then s_{2j} enters positively into any minor of B , hence there exists a choice for the value of s_{2j} so that B is TN_+ . Continue to choose values s_{2l} ($l > j$) large enough so that there exists a 2-by- l matrix with sign pattern $S[\{1, 2\}|\{1, 2, \dots, l\}]$ that is TN_+ until a zero element is encountered. Thus we have a 2-by- n TN_+ matrix C that is a realization of $S[\{1, 2\}|\{1, 2, \dots, n\}]$. Suppose, by induction, there exists a p -by- q matrix B a realization of $S[\{1, 2, \dots, p\}|\{1, 2, \dots, q\}]$, in which the (p, q) has not been specified. but both $B[\{1, 2, \dots, p\}|\{1, \dots, q-1\}]$ and $B[\{1, \dots, p-1\}|\{1, 2, \dots, q\}]$ are TN_+ matrices. Observe that any submatrix which does not involve b_{pq} is contained in either $B[\{1, 2, \dots, p\}|\{1, \dots, q-1\}]$ or $B[\{1, \dots, p-1\}|\{1, 2, \dots, q\}]$, and hence is TN_+ . Thus we are only concerned with submatrices which involve b_{pq} . Then we claim that the $(p, q)^{th}$ entry of B can be chosen so that B is TN_+ . If $s_{pq} = 0$. then since S is in double echelon form either row p or column q of B is zero, and hence B is a TN_+ matrix. Otherwise $s_{pq} = +$, and we may choose a value b_{pq} for s_{pq} large enough so that B is a TN matrix. If B is TN_+ , then we are done. Thus assume there exists a k -by- k submatrix of B (which must involve b_{pq}) that is singular but not sign-singular. Denote this submatrix by C and suppose C is in the following form $C = \left[\begin{array}{c|c} C_{11} & c_{12} \\ \hline c_{21} & b_{pq} \end{array} \right]$. Since C is not sign-singular, it follows that C_{11} is not sign-singular.

To see this suppose that C_{11} is sign-singular. Then there exists a zero block in C_{11} of size at least k (see [Ry]). Since C is in double echelon form it follows that there exists a zero block of size at least $k + 1$, and hence C is sign-singular, which is a contradiction. Since C_{11} is not sign-singular, it must be nonsingular and in this case we may increase the value b_{pq} so that C is nonsingular. The proof now follows by induction. ■

We now come to our main observations for this section.

THEOREM 2.5.4 *Let S be an m -by- n $(0, +)$ -pattern with no zero rows or columns.*

Then the following are equivalent:

1. *S allows a TN_+ matrix;*
2. *S allows a TN matrix;*
3. *S allows a TN_2 matrix;*
4. *S is a double echelon pattern.*

Proof. The implications $(1) \Rightarrow (2) \Rightarrow (3)$ are trivial since the containments $TN_+ \subseteq TN \subseteq TN_2$ are obvious. Suppose S allows a TN_2 matrix denoted by A , and assume that some entry, say $s_{pq} = 0$, otherwise S has no zero entries, and we are done. Hence $a_{pq} = 0$. Since S has no zero rows or columns some entry of A in row p is nonzero. Assume $a_{pt} > 0$. There are two cases to consider: $t > q$; or $t < q$. Suppose $t > q$. Since A is TN_2 it follows that $a_{sq} = 0$ for all $s > p$. Since S has no zero lines, some entry in the q^{th} column must be nonzero. Moreover, this entry

must occur with a row index less than p . Again since A is TN_2 it follows that $a_{ps} = 0$ for all $s < q$. Continuing this argument implies that $a_{ij} = 0$ for all $p \leq i \leq m$ and $1 \leq j \leq q$. For case 2, a similar argument shows that $a_{ij} = 0$ for all $1 \leq i \leq p$ and $q \leq j \leq n$. A routine induction implies that A and hence S is in double echelon form. The implication (4) \Rightarrow (1) is Theorem 2.5.3. ■

Observe that a similar argument used in the proof of (3) \Rightarrow (4) above may also be used to prove the following result.

COROLLARY 2.5.5 *Let A be an m -by- n totally nonnegative matrix with no zero rows or columns. Then A is in double echelon form.*

A result similar to the above corollary, though stated slightly differently, appeared in [R]. We conclude this section with some results on (0,1)-totally nonnegative matrices.

LEMMA 2.5.6 *Let A be an n -by- n (0,1)-lower (upper) triangular matrix with ones on and below (above) the main diagonal. Then A is a totally nonnegative matrix.*

Proof. Consider the (0,1)-matrix, $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. It is easy to verify that A is totally nonnegative. By Proposition 2.4.2 we can border A by a column of zeros on the right and then repeat the bottom row of A without affecting total nonnegativity. Next, by Proposition 2.4.4 we may increase the lower right entry to 1, which does not disturb the property of being TN. Thus, our new matrix, $\tilde{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$, is TN. We may continue this procedure to show that an n -by- n full lower triangular (0,1)-matrix is TN. The upper triangular version follows by transposition. ■

LEMMA 2.5.7 Any $(0,1)$ -matrix in the form

$$\left[\begin{array}{ccc} 0 & \cdots & 0 \\ & \ddots & \vdots \\ & & 0 \\ \text{1's} & & \end{array} \right]$$

is totally nonnegative.

Proof. This follows immediately from Proposition 2.4.1 and Lemma 2.5.6. ■

THEOREM 2.5.8 Let $A = [a_{ij}]$ be an m -by- n $(0,1)$ -matrix with no zero rows or columns. Then A is totally nonnegative if and only if A is in double echelon form and does not contain the submatrix:

$$B = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Proof. First, observe that the matrix, B is not TN because $\det(B) = -1$. The necessity is trivial since B is not TN and the property of total nonnegativity is inherited. To prove sufficiency we will proceed by induction on $m + n$. Note that this statement is not difficult to verify if $n \leq 3$. Suppose $m + n = k$, and that the result holds for all m' -by- n' such matrices with $m' + n' < k$. Consider the m -by- n matrix A . We may assume that A has no repeated consecutive rows or columns, otherwise the submatrix \hat{A} of A obtained by deleting a repeated row or column satisfies the induction hypothesis and hence is TN. Therefore A is TN by Proposition 2.4.1. Hence we may assume that $a_{11} = 1$. If $a_{12} = 0$, then the submatrix $A[\{2, 3, \dots, m\}|\{1, 2, \dots, n\}]$ satisfies the induction hypothesis and hence is TN. In this case A is again TN by

Proposition 2.4.4. Therefore we assume $a_{12} = 1$, and $a_{21} = 1$ (by transposition). Observe, that since A is in double echelon form and $a_{12} = a_{21} = 1$ it follows that $a_{22} = 1$. If $a_{33} = 0$, then either $a_{13} = a_{23} = 0$ or $a_{31} = a_{32} = 0$. In each case, since A is in double echelon form, either row 1 = row 2 (when $a_{13} = a_{23} = 0$) or column 1 = column 2 (otherwise), and hence we may delete a row (column) and apply the induction hypothesis. Therefore, we may assume $a_{33} = 1$, and similarly, we assume that $a_{23} = a_{32} = 1$, because otherwise, either row 1 = row 2 or column 1 = column 2. There are three cases to consider: $a_{13} = 0$; $a_{13} = 1$ and $a_{31} = 0$; and $a_{13} = a_{31} = 1$.

Case 1: $a_{13} = 0$.

$$\begin{bmatrix} 1 & 1 & 0 & \cdots & \\ 1 & 1 & 1 & \cdots & \\ 1 & 1 & 1 & & \\ \vdots & \vdots & & * & \\ & & & & \ddots \end{bmatrix}.$$

In which case we have that $a_{31} = 1$, because if $a_{31} = 0$, then the submatrix B would appear in the upper left corner of A . Now, if $a_{44} = 0$, then either row 2 = row 3 (if $a_{34} = a_{24} = 0$), or column 1 = column 2 (if $a_{42} = a_{43} = 0$). Thus, we may assume $a_{44} = 1$. Similarly, $a_{42} = a_{43} = 1$, otherwise column 1 = column 2. Additionally, $a_{34} = 1$, or else row 2 = row 3. Together, this implies that $a_{41} = 1$.

$$\begin{bmatrix} 1 & 1 & 0 & * & \cdots \\ 1 & 1 & 1 & * & \cdots \\ 1 & 1 & 1 & 1 & \\ 1 & 1 & 1 & 1 & \\ \vdots & \vdots & & & \ddots \end{bmatrix}.$$

If $a_{55} = 0$, then either column 1 = column 2 (if $a_{51} = a_{52} = a_{53} = a_{54} = 0$) or else row 3 = row 4 (if $a_{35} = a_{45} = 0$), and thus we may assume $a_{55} = 1$. Similarly, $a_{52} = a_{53} = a_{54} = 1$, because otherwise column 1 = column 2. Therefore, $a_{51} = 1$, because if $a_{51} = 0$, then B would appear as a submatrix of A . We may also assume that $a_{45} = 1$, for otherwise, row 3 = row 4. Thus, our matrix has the form:

$$\begin{bmatrix} 1 & 1 & 0 & * & * & \dots \\ 1 & 1 & 1 & * & * & \dots \\ 1 & 1 & 1 & 1 & * & \\ 1 & 1 & 1 & 1 & 1 & \\ 1 & 1 & 1 & 1 & 1 & \\ \vdots & \vdots & & & & \ddots \end{bmatrix}.$$

Continuing this method of analysis implies that every entry of A on and below the main diagonal must be equal to 1. Thus, A is TN by Lemma 2.5.7.

Case 2: $a_{13} = 1$, $a_{31} = 0$.

Observe that this case follows from Case 1 and transposition. The conclusion in this case, however, will be that every entry of A on and *above* the main diagonal must be equal to 1.

Case 3: $a_{13} = 1$ and $a_{31} = 1$.

Applying similar reasoning as in the previous two cases we can assume that $a_{44} = a_{24} = a_{34} = a_{43} = a_{42} = 1$. Then the argument proceeds in the same manner as before with three cases to consider: (1): $a_{14} = 0$; (2): $a_{14} = 1$ and $a_{41} = 0$, or (3): $a_{14} = a_{41} = 1$. Each of these cases is handled by similar arguments as the cases (1), (2) and (3) above, and the conclusions are similar in

each case. Continuing in this manner will prove that any $(0,1)$ -matrix in double echelon form that does not contain B as a submatrix is totally nonnegative. ■

2.6 Retractibility of Totally Nonnegative Matrices

We have already seen that if we increase the $(1,1)$ or the (m,n) entry of an m -by- n totally nonnegative (positive) matrix, then the resulting matrix is totally nonnegative (positive). In this section we investigate further which entries of a totally nonnegative (positive) matrix may be perturbed (that is, increased or decreased) so that the result is a totally nonnegative (positive) matrix. Such issues have already been addressed for other positivity classes of matrices, for example, if A is an n -by- n positive semidefinite, M -, P -, or inverse M -matrix, then $A + D$ (D a nonnegative diagonal matrix) is a positive semidefinite, M -, P -, or inverse M -matrix, respectively (see [HJ1, HJ2]). Recall that E_{ij} denotes the n -by- n $(i,j)^{th}$ *standard basis matrix*, that is the matrix whose $(i,j)^{th}$ entry is 1, and all remaining entries are zero. Suppose A is an n -by- n matrix. Then $\det(A - tE_{11}) = \det A - t\det A(\{1\})$. Therefore, if $\det A(\{1\}) \neq 0$, then $\det(A - tE_{11}) = 0$, when $t = \frac{\det A}{\det A(\{1\})}$. We are now in a position to make the following definitions.

DEFINITION 2.6.1 Let \mathcal{C} denote a given subclass of the n -by- n P_0 -matrices, and suppose A is an n -by- n matrix. Then we define:

(i) $A^R = \{A - tE_{11} : t \in [0, \frac{\det A}{\det A(\{1\})}]\}$ — “the set of retractions of A ”,

(ii) $\mathcal{C}^{(R)} = \{A \in \mathcal{C} : A^R \subseteq \mathcal{C}\}$ — “the retractable subset of \mathcal{C} ”,

(iii) $\mathcal{C}_R = \cup_{A \in \mathcal{C}} A^R$ — “the set of all retractions of matrices in \mathcal{C} ”.

If, in (i), $\det A(\{1\}) = 0$, then the interval for t is defined to be the single point zero. The notions of retractibility (of a matrix) and retractible sets are very important for studying certain determinantal inequalities, such as Oppenheim’s inequality for entry-wise (or Hadamard) products of positive semidefinite matrices (see [HJ1]), see, for example, Chapter 5, where entry-wise products of totally nonnegative matrices is discussed. It is well-known (see [HJ1] or [HJ2]), and is not difficult to prove, that if $\mathcal{C} = PSD$, the set of all positive semidefinite matrices, then $PSD^{(R)} = PSD_R = PSD$. Also if $\mathcal{C} = M$, the set of all M -matrices, then $M^{(R)} = M_R = M$. A somewhat more subtle result is that if $\mathcal{C} = TN$, the set of all totally nonnegative matrices, then $TN^{(R)} = TN_R = TN$. This fact will follow immediately from the next lemma.

LEMMA 2.6.2 *Let A be an n -by- n totally nonnegative matrix with $\det A(\{1\}) \neq 0$. Then $A - xE_{11}$ is totally nonnegative for all $x \in [0, \frac{\det A}{\det A(\{1\})}]$.*

Proof. Firstly, observe that for every value $x \in [0, \frac{\det A}{\det A(\{1\})}]$, $\det(A - xE_{11}) \geq 0$. Recall that A admits a UL -factorization (follows from the LU -factorization result and reversal) into totally nonnegative matrices. Partition A as follows,

$$A = \begin{bmatrix} a_{11} & a_{12}^T \\ a_{21} & A_{22} \end{bmatrix},$$

where a_{11} is 1-by-1 and $A_{22} = A(\{1\})$. Partition L and U conformally with A . Then

$$A = \begin{bmatrix} a_{11} & a_{12}^T \\ a_{21} & A_{22} \end{bmatrix} = UL = \begin{bmatrix} u_{11} & u_{12}^T \\ 0 & U_{22} \end{bmatrix} \begin{bmatrix} l_{11} & 0 \\ l_{21} & L_{22} \end{bmatrix} = \begin{bmatrix} u_{11}l_{11} + u_{12}^T l_{21} & u_{12}^T L_{22} \\ U_{22}l_{21} & U_{22}L_{22} \end{bmatrix}.$$

Consider the matrix $A - xE_{11}$, with $x \in [0, \frac{\det A}{\det A(\{1\})}]$. Then

$$A - xE_{11} = \begin{bmatrix} u_{11}l_{11} + u_{12}^T l_{21} - x & u_{12}^T L_{22} \\ U_{22}l_{21} & U_{22}L_{22} \end{bmatrix} = \begin{bmatrix} u_{11} - \frac{x}{l_{11}} & u_{12}^T \\ 0 & U_{22} \end{bmatrix} \begin{bmatrix} l_{11} & 0 \\ l_{21} & L_{22} \end{bmatrix} = U' L,$$

if $l_{11} \neq 0$. Note that if $l_{11} = 0$, then L , and hence A , is singular. In this case $x = 0$ is the only allowed value for x . But in this case the desired result is trivial. Thus we assume that $l_{11} > 0$. To show that $A - xE_{11}$ is totally nonnegative it is enough to verify that $u_{11} - x/l_{11} \geq 0$. Since if this was the case U' is totally nonnegative by Corollary 2.4.5 and as L is totally nonnegative by assumption, we have that their product, $A - xE_{11}$ is totally nonnegative. Since $l_{11} > 0$ and $\det A(\{1\}) > 0$ it follows that L and U_{22} are nonsingular. Hence $0 \leq \det(A - xE_{11}) = (u_{11} - x/l_{11})\det U_{22}\det L$, from which it follows that $u_{11} - x/l_{11} \geq 0$. ■

Note that a similar result holds for decreasing the (n, n) entry by considering the matrix $\rho A \rho$ as in Proposition 2.4.6.

COROLLARY 2.6.3 *Let TN denote the class of all n -by- n totally nonnegative matrices. Then $TN^{(R)} = TN_R = TN$.*

For the remainder of this section we restrict ourselves to the set of totally positive matrices. The first result for this class is a slight strengthening of Lemma 2.6.2. Recall that an n -by- n triangular matrix A is said to a *triangular TP matrix*, if all minors of A are positive, except for those that are zero by virtue of the zero pattern of A . (Recall the definitions of TN_k and TP_k from page 36.)

THEOREM 2.6.4 *Let A be an n -by- n totally positive matrix. Then $A - tE_{11}$ is a TP_{n-1} matrix, for all $t \in [0, \frac{\det A}{\det A(\{1\})}]$.*

Proof. Following the proof of Lemma 2.6.2 we can write

$$A - tE_{11} = \begin{bmatrix} u_{11} - \frac{t}{l_{11}} & u_{12}^T \\ 0 & U_{22} \end{bmatrix} \begin{bmatrix} l_{11} & 0 \\ l_{21} & L_{22} \end{bmatrix} = U' L,$$

where $U' = U - (\frac{t}{l_{11}})E_{11}$, and both U, L are triangular TP matrices (see [C2]). If $u_{11} - t/l_{11} > 0$, then U' and L are triangular TP matrices and hence $A - tE_{11}$ is TP. So consider the case $u_{11} - t/l_{11} = 0$, or equivalently, $\det(A - tE_{11}) = 0$, or $t = \frac{\det A}{\det A(\{1\})}$. Let $B = A - tE_{11}$. Observe that $B[\{1, 2, \dots, n\}|\{2, 3, \dots, n\}]$ and $B[\{2, 3, \dots, n\}|\{1, 2, \dots, n\}]$ are TP matrices since A is TP. Thus the only contiguous minors left to verify are the leading contiguous minors of B . Consider the submatrix $B[\{1, 2, \dots, k\}]$ for $1 \leq k < n$. Then $\det B[\{1, 2, \dots, k\}] = \det A[\{1, 2, \dots, k\}] - t \det A[\{2, \dots, k\}]$. This minor is positive if and only if

$$\det A[\{1, 2, \dots, k\}] > t \det A[\{2, \dots, k\}],$$

which is equivalent to $\det A[\{1, 2, \dots, k\}] \det A(\{1\}) > \det A \det A[\{2, \dots, k\}]$, an example of a Koteljanskii inequality. The only issue left to settle is whether or not equality holds for the above Koteljanskii inequality. We claim here that for a TP matrix every Koteljanskii inequality is strict. Suppose to the contrary, that is, assume there exist two index sets α and β such that $\det A[\alpha \cup \beta] \det A[\alpha \cap \beta] = \det A[\alpha] \det A[\beta]$. For simplicity, we may assume that $\alpha \cup \beta = N$, otherwise replace A by $A[\alpha \cup \beta]$ in the following. By Jacobi's identity (1.1) we have $\det A^{-1}[(\alpha \cup \beta)^c] \det A^{-1}[(\alpha \cap \beta)^c] =$

$\det A^{-1}[\alpha^c] \det A^{-1}[\beta^c]$. Let $C = SA^{-1}S$, for $S = \text{diag}(1, -1, \dots, \pm 1)$. Then C is TP and the above equation implies $\det C = \det C[\alpha^c] \det C[\beta^c]$. By a result in [Ca], C is reducible, which is nonsense since C is TP. Thus $A - tE_{11}$ is TP_{n-1} , by Fekete's criterion. This completes the proof. ■

Observe that from the proof above it follows that $A - tE_{11}$ is TP for all $t \in [0, \frac{\det A}{\det A(\{1\})})$, and is TP_{n-1} when $t = \frac{\det A}{\det A(\{1\})}$. An obvious next question is what other entries can be increased/decreased to the point of singularity so that the matrix is TP_{n-1} . As it turns out retracting (or decreasing) the (2,2) entry of a TP matrix results in a TP_{n-1} matrix.

THEOREM 2.6.5 *Let A be an n -by- n totally positive matrix. Then $A - tE_{22}$ is TP_{n-1} for all $t \in [0, \frac{\det A}{\det A(\{2\})}]$.*

Proof. Using the fact that all Koteljanskii inequalities are strict it follows that all of the leading principal minors of $A - tE_{22}$ are positive. Consider the submatrix $B = (A - tE_{22})[\{1, 2, \dots, n\}|\{2, 3, \dots, n\}]$. To show that B is TP we need only consider the contiguous minors of B that involve the first and second row and first column, all other minors are positive by assumption. Let C denote such a submatrix of B . To compute $\det C$, expand the determinant along the second row of C . Then $\det C = (-1)^{1+2}(-t)\det C(\{2\}|\{1\}) + \det A[\alpha|\beta]$, where $\det A[\alpha|\beta]$ is some minor of A . Thus $\det C$ is a positive linear combination of minors of A , and hence is positive. Therefore B is TP. Similar arguments show that $(A - tE_{22})[\{2, \dots, n\}|\{1, 2, \dots, n\}]$ is TP. This completes the proof. ■

A similar fact holds for the retractibility of the $(n-1, n-1)$ entry of a TP matrix.

The next result follows directly from Theorems 2.6.4 and 2.6.5.

COROLLARY 2.6.6 *Let $n \leq 4$. If A is a n -by- n totally positive matrix and $1 \leq i \leq n$, then $A - tE_{ii}$ is TP_{n-1} for all $t \in [0, \frac{\det A}{\det A(\{i\})}]$.*

According to the next example we cannot decrease any other interior main diagonal entry (in general) of a TP matrix and stay TP_{n-1} .

EXAMPLE 2.6.7 Consider the following matrix.

$$A = \begin{bmatrix} 100 & 10 & 7/5 & 2 & 1 \\ 22 & 5 & 2 & 3 & 2 \\ 3 & 1 & 1.01 & 2 & 3 \\ 1 & 1 & 2 & 5 & 12 \\ 1/2 & 2 & 5 & 15 & 50 \end{bmatrix}.$$

Then A is a totally positive matrix with $\frac{\det A}{\det A(\{3\})} \approx .03$. However, $\frac{\det A[\{1,2,3\}|\{3,4,5\}]}{\det A[\{1,2\}|\{4,5\}]} = .01$. Thus for $t \in (.01, .03]$, $\det(A - tE_{33})[\{1,2,3\}|\{3,4,5\}] < 0$, and hence $A - tE_{33}$ is not TP_4 .

Up to this point we have only considered retracting on a single diagonal entry, and obvious next step is to consider increasing or decreasing off-diagonal entries in a TP matrix. We begin our study of perturbing off-diagonal entries by considering the (1,2) entry.

THEOREM 2.6.8 *Let A be an n -by- n totally positive matrix. Then $A + tE_{12}$ is TP_{n-1} for all $t \in [0, \frac{\det A}{\det A(\{1\}|\{2\})}]$.*

Proof. Since the (1,2) entry of A enters negatively into $\det A$ we increase a_{12} to $a_{12} + \frac{\det A}{\det A(\{1\}|\{2\})}$ so that $\det(A + tE_{12}) = 0$. Observe that the submatrix $(A + tE_{12})[\{2, \dots, n\}|\{1, 2, \dots, n\}]$ is equal to $A[\{2, \dots, n\}|\{1, 2, \dots, n\}]$ and hence is TP. Moreover, $(A + tE_{12})[\{1, 2, \dots, n\}|\{2, \dots, n\}]$ is TP since we have increased the “(1,1)” entry of a TP matrix. The only remaining minors to verify are the leading principal minors of $A + tE_{12}$. Observe that for $t \in [0, \frac{\det A}{\det A(\{1\}|\{2\})}]$,

$$\begin{aligned} 0 &\leq \det(A + tE_{12})\det A[\{2, 3, \dots, n-1\}] \\ &= \det(A + tE_{12})[\{1, 2, \dots, n-1\}]\det A[\{2, 3, \dots, n\}] \\ &\quad - \det(A + tE_{12})[\{1, 2, \dots, n-1\}|\{2, \dots, n\}]\det A[\{2, \dots, n\}|\{1, 2, \dots, n-1\}], \end{aligned}$$

follows by Sylvester’s identity (see chapter 1). Hence $\det(A + tE_{12})[\{1, 2, \dots, n-1\}] > 0$. Replacing $A + tE_{12}$ by $(A + tE_{12})[\{1, 2, \dots, n-1\}]$ in the above identity yields $\det(A + tE_{12})[\{1, 2, \dots, n-2\}] > 0$ and so on. This completes the proof. ■

Using transposition and reversal the conclusion of Theorem 2.6.8 holds when the (1,2) entry of a totally positive matrix A is replaced by the (2,1), $(n-1, n)$ and $(n, n-1)$ entries of A . Unfortunately, this is all that can be said positively concerning increasing or decreasing off-diagonal entries. Consider the following example.

EXAMPLE 2.6.9 Let

$$A = \begin{bmatrix} 13 & 33 & 31 & 10 \\ 132 & 383 & 371 & 120 \\ 13 & 38 & 37 & 12 \\ 1 & 3 & 3 & 1 \end{bmatrix}.$$

Then $\frac{\det A}{\det A(\{1\}|\{3\})} = 1$ and $\frac{\det A(\{1,2\}|\{3,4\})}{a_{24}} = 1/12$. Thus for $t \in (1/12, 1]$ $\det(A - tE_{13})(\{1,2\}|\{3,4\}) < 0$. Thus $A - tE_{13}$ is not TP_3 .

Similar examples can also be constructed in the case of the (2,3) and (1,4) entries, and consequently in all remaining off-diagonal entries of an n -by- n totally positive matrix.

2.7 Row and Column Inclusion

In this section we generalize in a certain way a fact known for positive semidefinite (PSD) matrices to a much wider setting that includes most totally nonnegative matrices and most (possibly singular) M-matrices. In this way we provide a further direction of unification between these classes.

It is known that if $A = [a_{ij}]$ is positive semidefinite, then any column vector of the form

$$\begin{bmatrix} a_{i_1 j} \\ a_{i_2 j} \\ \vdots \\ a_{i_k j} \end{bmatrix}$$

must lie in the column space of the principal submatrix $A(\{i_1, i_2, \dots, i_k\})$. (This classical fact may be seen in a variety of ways. For example, consider an Hermitian matrix of the form

$$\begin{bmatrix} B & c \\ c^* & d \end{bmatrix}.$$

If the column vector c is *not* in the range of B , null space/range orthogonality implies there is a vector x in the null space of B such that $x^*c < 0$. There is then an $\varepsilon > 0$ such that

$$\begin{bmatrix} x^* & \varepsilon \end{bmatrix} \begin{bmatrix} B & c \\ c^* & d \end{bmatrix} \begin{bmatrix} x \\ \varepsilon \end{bmatrix} < 0.$$

Since A^T is also positive semidefinite, there is an analogous statement about rows. Moreover, for positive semidefinite matrices (though *both* inclusions are valid), it is equivalent to say that for each j either the row lies in the row space *or* the column lies in the column space of the indicated principal submatrix. It is precisely this statement that may be substantially generalized. This fact has been known to several researchers for some time, see [FJ2] for more history and motivation.

For a given m -by- n matrix A , we let $Row(A)$ ($Col(A)$) denote the row (column) space of A , and let $\text{rank}(A)$ denote the rank of A .

Let \mathcal{F}_0 (\mathcal{F}) denote the set of n -by- n P_0 - (P -) matrices which also satisfy Fischer's inequality. The class \mathcal{F}_0 contains many familiar classes of matrices for example: positive semidefinite, M-matrices, totally nonnegative matrices (and their permutation similarities) and triangular matrices with nonnegative main diagonal.

Using the notation developed above, we can reformulate the column inclusion result for positive semidefinite matrices as follows: If A is an n -by- n positive semidefinite matrix, then for any $\alpha \subseteq N$ and for each j , $A[\alpha|\{j\}] \in Col(A[\alpha])$.

The next result can be found in [FJ2, Thm. 6.1]

THEOREM 2.7.1 *If $A \in \mathcal{F}_0$ and $\alpha \subseteq N$, is such that $\text{rank}(A[\alpha]) \geq |\alpha| - 1$, then for each j either $A[\{j\}|\alpha] \in \text{Row}(A[\alpha])$ or $A[\alpha|\{j\}] \in \text{Col}(A[\alpha])$.*

Some condition imposed upon $\text{rank}(A[\alpha])$ is necessary, as may be seen from the following example.

EXAMPLE 2.7.2 Let

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Then $A \in \mathcal{F}_0$. Let $\alpha = \{1, 3\}$. Then $A[\alpha] = 0$. However, $A[\{2\}|\alpha] \neq 0$ and $A[\alpha|\{2\}] \neq 0$.

Note the above matrix is a reducible TN matrix. To generalize further, we use the following concept.

DEFINITION 2.7.3 *An n -by- n matrix A satisfies the *principal rank property* (PRP), if every principal submatrix A' of A has in turn an invertible *principal* submatrix that is of the order $\text{rank}(A')$.*

The following is simply a recasting of the previous definition. A satisfies the PRP if for every α , there exists $\beta \subseteq \alpha$, with $|\beta| = \text{rank}(A[\alpha])$ so that $A[\beta]$ is invertible.

Observe that the principal rank property is inherited by principal submatrices. The next lemma (perhaps of independent interest) gives a sufficient condition for a matrix to satisfy the PRP.

LEMMA 2.7.4 *Let A be an n -by- n matrix. Suppose the algebraic multiplicity of 0 equals the geometric multiplicity of 0, for every principal submatrix of A . Then A satisfies the PRP.*

Proof. Let A' be any k -by- k principal submatrix of A . If A' is nonsingular, then there is nothing to do. Thus suppose $\text{rank}(A') = r$ ($r < k$). Since the algebraic multiplicity of 0 equals the geometric multiplicity of 0 for A' , it follows that the characteristic polynomial of A' is equal to

$$\det(xI - A') = x^{k-r}(x^r - s_1x^{r-1} + \cdots + (-1)^r s_r),$$

in which $s_r \neq 0$. Since s_r is equal to the sum of all the r -by- r principal minors of A' ,

it follows that there exists at least one nonsingular r -by- r principal submatrix of A' .

This completes the proof. ■

Note that symmetric matrices are examples of matrices which satisfy the above lemma, and hence satisfy the PRP. The converse to the above general lemma is easily seen not to hold in general. The simplest example demonstrating this fact is

$$A = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}.$$

However, for P_0 -matrices the condition given in the above lemma is clearly also necessary, since for a P_0 -matrix, the coefficient s_r will be nonzero if and only if there is a nonsingular r -by- r principal submatrix of A .

THEOREM 2.7.5 ([FJ2, Thm. 6.5]) *If $A \in \mathcal{F}_0$, satisfies the PRP, and $\alpha \subseteq N$, then for each j either $A[\{j\}|\alpha] \in \text{Row}(A[\alpha])$ or $A[\alpha|\{j\}] \in \text{Col}(A[\alpha])$.*

Theorem 2.7.5 has two hypotheses: \mathcal{F}_0 and PRP. Neither of these can be omitted in general. For example,

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

satisfies PRP but not the row or column inclusion conclusion of Theorem 2.7.5, and

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

is in \mathcal{F}_0 , does not satisfy PRP, and does not satisfy the conclusion of Theorem 2.7.5 for $\alpha = \{1, 3\}$.

Since any positive semidefinite matrix satisfies the PRP (by the previous lemma) and is in \mathcal{F}_0 (see [HJ1]) we have the following.

COROLLARY 2.7.6 *If A is an n -by- n positive semidefinite matrix, then for any $\alpha \subseteq N$ and for each j , $A[\alpha|\{j\}] \in \text{Col}(A[\alpha])$.*

We now specialize to the class of totally nonnegative matrices (TN), a subclass of \mathcal{F}_0 . We may extend our row or column inclusion results somewhat to *non*-principal submatrices for this class. Before we state our next result we need the following notation. Let $\alpha, \beta \subseteq N$ and let $i_1 < i_2 < \dots < i_k$ be the elements of α arranged in increasing order, and let $j_1 < j_2 < \dots < j_k$ be the elements of β arranged in increasing order. Define $i_0 = 1, j_0 = 1, i_{k+1} = n$ and $j_{k+1} = n$. For

$t = 0, 1, \dots, k$, let $A[i_1, \dots, i_t, (i_t, i_{t+1}), i_{t+1}, \dots, i_k | \beta]$ denote the submatrix of A obtained from $A[\alpha | \beta]$ by inserting any row s , for which $i_t < s < i_{t+1}$. Similarly, let $A[\alpha | j_1, \dots, j_t, (j_t, j_{t+1}), j_{t+1}, \dots, j_k]$ denote the submatrix obtained from $A[\alpha | \beta]$ by inserting any column q , for which $j_t < q < j_{t+1}$.

THEOREM 2.7.7 *Let A be an n -by- n TN matrix and let $\alpha, \beta \subseteq N$, with $|\alpha| = |\beta|$. Suppose that $\text{rank}(A[\alpha | \beta]) = p$, and either $p \geq |\alpha| - 1$ or at least one of $d(\alpha) = 0$ or $d(\beta) = 0$. Then either*

$$\text{rank}(A[i_1, \dots, i_t, (i_t, i_{t+1}), i_{t+1}, \dots, i_k | \beta]) = p$$

or

$$\text{rank}(A[\alpha | j_1, \dots, j_t, (j_t, j_{t+1}), j_{t+1}, \dots, j_k]) = p,$$

for $t = 0, 1, \dots, k$.

Proof. Let $\alpha, \beta \subseteq N$, with $|\alpha| = |\beta|$. Firstly, suppose $d(\alpha) = 0$. The case when $d(\beta) = 0$ will then follow by transposition. Let $s, q \in N$, with $s < i_1$ and $q < j_1$ (if such an s, q exist). The case in which $s > i_k$ and $q > j_k$ is handled similarly. Assume for the purpose of contradiction that $\text{rank}(A[\alpha \cup \{s\} | \beta]) > p$ and $\text{rank}(A[\alpha | \beta \cup \{q\}]) > p$. Consider the submatrix $A[\alpha \cup \{s\} | \beta \cup \{q\}]$, which is totally nonnegative and has rank equal to $p + 2$. Thus there exists a nonsingular submatrix A' of $A[\alpha \cup \{s\} | \beta \cup \{q\}]$ of order $p + 2$. Since the $\text{rank}(A[\alpha | \beta]) = p$, it follows that $A' = A[\gamma \cup \{s\} | \delta \cup \{q\}]$, in which $\gamma \subseteq \alpha$, $|\gamma| = p + 1$ and $\delta \subseteq \beta$, $|\delta| = p + 1$. Further, $A[\gamma | \delta]$ is principal in A' , since $s < i_1$ and $q < j_1$. Observe that since $\text{rank}(A[\alpha | \beta]) = p$,

$A[\gamma|\delta]$ is singular. Consequently,

$$0 < \det A' \leq \det A[\gamma|\delta] \cdot a_{sq} = 0.$$

This is a contradiction and hence the conclusion holds.

Finally, suppose $\text{rank}(A[\alpha|\beta]) \geq |\alpha| - 1$, and let $|\alpha| = k$. Define $i_0 = j_0 = 1$ and $i_{k+1} = j_{k+1} = n$. Let t be any fixed integer, $0 \leq t \leq k$. Assume there exists a row indexed by s , $i_t < s < i_{t+1}$, and a column indexed by q , $j_t < q < j_{t+1}$, in which $\text{rank}(A[i_1, \dots, i_t, s, i_{t+1}, \dots, i_k|\beta]) > p$ and $\text{rank}(A[\alpha|j_1, \dots, j_t, q, j_{t+1}, \dots, j_k]) > p$. In fact both must equal $p + 1$. Then $\text{rank} A[\alpha \cup \{s\}|\beta \cup \{q\}] = p + 2$. Again consider the nonsingular submatrix A' of $A[\alpha \cup \{s\}|\beta \cup \{q\}]$, which (as before) must be of the form $A[\gamma \cup \{s\}|\delta \cup \{q\}]$. Applying similar arguments as above we will arrive at a contradiction. This completes the proof. ■

We wish to remark here that if for some t ($0 \leq t \leq k$) $i_t = i_{t+1} - 1$ and $j_t < j_{t+1} - 1$, then there may exist a q , with $j_t < q < j_{t+1}$ and

$$\text{rank}(A[\alpha|\beta \cup \{q\}]) > \text{rank}(A[\alpha|\beta]).$$

Thus, the conclusion of the above theorem holds only if for a fixed t

$(0 \leq t \leq k)$ $i_t < i_{t+1} - 1$ and $j_t < j_{t+1} - 1$.

We complete this discussion with an example, which illustrates three possibilities. The first two give insight into the necessity of the hypotheses in Theorem 2.7.7. The final one illustrates the previous remark.

EXAMPLE 2.7.8 Let

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

It is not difficult to determine that A is TN.

(i) Let $\alpha = \{2, 4\}$ and $\beta = \{2, 4\}$. Then $\text{rank}(A[\alpha|\beta]) = 1$. However, if we let $s = 1$ and $q = 5$, then $\text{rank}(A[\alpha \cup \{s\}|\beta]) = 2$ and $\text{rank}(A[\alpha|\beta \cup \{q\}]) = 2$. Thus it is necessary to include a row and a column from the same “gap” in α and β .

(ii) Similar arguments can be applied in the case $\alpha = \{1, 3\}$ and $\beta = \{4, 5\}$, and with $s = q = 2$.

(iii) Let $\alpha = \{2, 3\}$ and $\beta = \{1, 3\}$. Suppose $q = 2$. Then

$$\text{rank}(A[\alpha|\beta \cup \{q\}]) > \text{rank}(A[\alpha|\beta]).$$

2.8 Sub-direct Sums

Let $0 \leq k \leq m, n$ and suppose that

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \in M_m(\mathbb{C}) \text{ and } B = \begin{bmatrix} B_{22} & B_{23} \\ B_{32} & B_{33} \end{bmatrix} \in M_n(\mathbb{C}),$$

in which $A_{22}, B_{22} \in M_k(\mathbb{C})$. Then we call

$$C = \begin{bmatrix} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} + B_{22} & B_{23} \\ 0 & B_{32} & B_{33} \end{bmatrix}$$

the k -subdirect sum of A and B , which we denote by $A \oplus_k B$. When the value of k is irrelevant or clear, we may just refer to a *subdirect sum*, and, when $k = 0$, a 0-subdirect sum is a familiar direct sum, and \oplus_0 will be abbreviated to \oplus . Of course, the k -subdirect sum of two matrices is generally not commutative. See [FJ1] for more information and background (including references) on sub-direct sums of other positivity classes of matrices.

For a given class of matrices Π we address four natural questions: (I) If A and B are in Π must a 1-subdirect sum C be in Π ; (II) If

$$C = \begin{bmatrix} C_{11} & C_{12} & 0 \\ C_{21} & C_{22} & C_{23} \\ 0 & C_{32} & C_{33} \end{bmatrix}$$

is in Π , may C be written as $C = A \oplus_1 B$, such that A and B are both in Π , when C_{22} is 1-by-1; and (III) and (IV) the corresponding questions with 1 replaced by $k > 1$.

We begin with a simple determinantal identity, that is useful in consideration of 1-subdirect sums. We provide a proof here for completeness.

LEMMA 2.8.1 *Let*

$$A = \begin{bmatrix} A_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in M_m(\mathbb{C}) \text{ and } B = \begin{bmatrix} b_{22} & b_{23} \\ b_{32} & B_{33} \end{bmatrix} \in M_n(\mathbb{C}),$$

in which a_{22} and b_{22} are both 1-by-1. Then

$$\det(A \oplus_1 B) = \det A_{11} \det B + \det A \det B_{33}. \quad (2.1)$$

Proof. It is routine to verify that if A_{11} and B_{33} are both singular, then $A \oplus_1 B$ is necessarily singular; hence (2.1) holds in this special case. Suppose, then, without loss of generality, that A_{11} is nonsingular. Using Schur complements (see chapter 1) it follows that

$$\det \begin{bmatrix} A_{11} & a_{12} & 0 \\ a_{21} & a_{22} + b_{22} & b_{23} \\ 0 & b_{32} & B_{33} \end{bmatrix} = \det A_{11} \det \begin{bmatrix} a_{22} + b_{22} - a_{21} A_{11}^{-1} a_{12} & b_{23} \\ b_{32} & B_{33} \end{bmatrix}.$$

Expanding the latter determinant by the first row gives

$$\begin{aligned} \det A_{11} \det \begin{bmatrix} b_{22} & b_{23} \\ b_{32} & B_{33} \end{bmatrix} + \det A_{11} (a_{22} - a_{21} A_{11}^{-1} a_{12}) \det B_{33} \\ = \det A_{11} \det B + \det A \det B_{33}. \blacksquare \end{aligned}$$

In [FJ1] we considered questions (I)-(IV) for various positivity classes: positive (semi) definite; M -matrices; symmetric M -matrices; P -matrices; P_0 -matrices; doubly nonnegative matrices (entry-wise nonnegative positive semidefinite matrices); and completely positive matrices (matrices of the form BB^T with B entry-wise nonnegative). One of the results proved in [FJ1] is the next result which addresses specifically questions (I) and (II) for each of the above positivity classes.

THEOREM 2.8.2 ([FJ1]) *Let Π denote any one of the above positivity classes.*

Suppose

$$C = \begin{bmatrix} C_{11} & c_{12} & 0 \\ c_{21} & c_{22} & c_{23} \\ 0 & c_{32} & C_{33} \end{bmatrix}$$

is an n -by- n matrix with C_{11} and C_{33} square, and with c_{22} 1-by-1. Then C is in Π if and only if C may be written as $C = A \oplus_1 B$, in which both A and B are in Π .

For more information on questions (III) and (IV) for these positivity classes see [FJ1]. We now turn our attention to the class of totally nonnegative matrices (see also [FJ1]). The next result answers questions (I) and (II) for the class TN.

THEOREM 2.8.3 *Let*

$$C = \begin{bmatrix} C_{11} & c_{12} & 0 \\ c_{21} & c_{22} & c_{23} \\ 0 & c_{32} & C_{33} \end{bmatrix} \in M_n(\mathbb{R})$$

be an n -by- n matrix with C_{11} and C_{33} square, and with c_{22} 1-by-1. Then C is TN if and only if C may be written as $C = A \oplus_1 B$, in which both A and B are TN.

Proof. First, suppose both A and B are TN, and let

$$A = \begin{bmatrix} A_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in M_m(\mathbb{R}) \text{ and } B = \begin{bmatrix} b_{22} & b_{23} \\ b_{32} & B_{33} \end{bmatrix} \in M_r(\mathbb{R}),$$

where a_{22} and b_{22} are both 1-by-1. Firstly, we note that any principal submatrix of $A \oplus_1 B$ either involves the overlapping entry and hence is nonnegative by (2.1); or does not involve the overlapping entry and is nonnegative since it is a direct sum of TN matrices. Let $C = A \oplus_1 B$, and let $C[\alpha|\beta]$ be any square submatrix of C . Let $\alpha_1 = \alpha \cap \{1, 2, \dots, m-1\}$, $\alpha_2 = \alpha \cap \{m+1, m+2, \dots, n\}$, $\beta_1 = \beta \cap \{1, 2, \dots, m-1\}$ and $\beta_2 = \beta \cap \{m+1, m+2, \dots, n\}$. Further, we can assume $\alpha_1, \alpha_2, \beta_1$ and β_2 are all

nonempty, otherwise it is straightforward to verify that $\det C[\alpha|\beta] \geq 0$. Suppose $m \notin \alpha \cap \beta$. Then either $m \notin \alpha$ and $m \notin \beta$ or, without loss of generality, $m \in \alpha$ and $m \notin \beta$. First assume $m \notin \alpha$ and $m \notin \beta$. Then

$$C[\alpha|\beta] = \begin{bmatrix} A_{11}[\alpha_1|\beta_1] & 0 \\ 0 & B_{33}[\alpha_2|\beta_2] \end{bmatrix}.$$

If $|\alpha_1| = |\beta_1|$, then $C[\alpha|\beta]$ is a direct sum of TN matrices, and hence is TN. Otherwise $|\alpha_1| \neq |\beta_1|$, and without loss of generality, assume $|\alpha_1| > |\beta_1|$ (the case $|\alpha_1| < |\beta_1|$ follows by symmetry). Therefore the size of the larger zero block is $|\alpha_1| - |\beta_1|$. Furthermore,

$$|\alpha_1| + |\beta_2| \geq |\beta_1| + |\beta_2| + 1 = |\beta| + 1 = |\alpha| + 1,$$

hence $\det C[\alpha|\beta] = 0$ (see [Ry, p. 55]). Now assume $m \in \alpha$ and $m \notin \beta$. Then

$$C[\alpha|\beta] = \begin{bmatrix} A_{11}[\alpha_1|\beta_1] & 0 \\ a_{21}[\beta_1] & b_{23}[\beta_2] \\ 0 & B_{33}[\alpha_2|\beta_2] \end{bmatrix}.$$

If $|\alpha_1| = |\beta_1|$, then $|\alpha_2| + 1 = |\beta_2|$. Hence $C[\alpha|\beta]$ is block triangular and it follows that $\det C[\alpha|\beta] \geq 0$. Otherwise, $|\alpha_1| \neq |\beta_1|$. If $|\alpha_1| > |\beta_1|$, then $\det C[\alpha|\beta] = 0$ (see [Ry, p. 55]), and if $|\alpha_1| < |\beta_1|$, then either $C[\alpha|\beta]$ is block triangular or $C[\alpha|\beta]$ is singular. Thus suppose $m \in \alpha \cap \beta$. Again there are two cases to consider. Suppose $|\alpha_1| = |\beta_1|$. Then $|\alpha_2| = |\beta_2|$, and $\det C[\alpha|\beta] \geq 0$ follows from (2.1). Otherwise suppose $|\alpha_1| \neq |\beta_1|$, and without loss of generality, assume $|\alpha_1| > |\beta_1|$ (the case $|\alpha_1| < |\beta_1|$

follows by symmetry). The order of the larger zero block is $|\alpha_1|$ -by- $|\beta_2|$, and

$$|\alpha_1| + |\beta_2| \geq |\beta_1| + |\beta_2| + 1 = |\beta| = |\alpha|.$$

If $|\alpha_1| + |\beta_2| > |\beta|$, then, as before, $\det C[\alpha|\beta] = 0$, see [Ry, p. 55]. So assume $|\alpha_1| + |\beta_2| = |\beta|$, in which case, it follows that $C[\alpha|\beta]$ is a block triangular matrix with the diagonal blocks being TN, hence $\det C[\alpha|\beta] \geq 0$.

Conversely, suppose C is TN. Since

$$\begin{bmatrix} C_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \text{ and } \begin{bmatrix} c_{22} & c_{23} \\ c_{32} & C_{33} \end{bmatrix},$$

are TN, choose $\hat{a}_{22} \geq 0$ and $\hat{b}_{22} \geq 0$ as small as possible so that

$$\hat{A} = \begin{bmatrix} C_{11} & c_{12} \\ c_{21} & \hat{a}_{22} \end{bmatrix} \text{ and } \hat{B} = \begin{bmatrix} \hat{b}_{22} & c_{23} \\ c_{32} & C_{33} \end{bmatrix},$$

are both TN. Assume $\hat{A} \in M_m(\mathbb{R})$ and $\hat{B} \in M_r(\mathbb{R})$. In what follows we are only concerned with square submatrices of \hat{A} and \hat{B} . Let

$\Gamma = \{ \hat{A}[\alpha|\beta] : m \in \alpha, \beta \subseteq \{1, 2, \dots, m\} \text{ and } \hat{A}[\alpha|\beta] \text{ is singular} \}$, and let

$\Lambda = \{ \hat{B}[\gamma|\delta] : 1 \in \gamma, \delta \subseteq \{1, 2, \dots, r\} \text{ and } \hat{B}[\gamma|\delta] \text{ is singular} \}$. Observe that Γ (similarly Λ) is nonempty. Since if Γ was empty, then for every $m \in \alpha, \beta \subseteq \{1, 2, \dots, m\}$, $\det \hat{A}[\alpha|\beta] > 0$. Then by continuity we may decrease \hat{a}_{22} , while \hat{A} remains TN, which contradicts the minimality of \hat{a}_{22} . Therefore Γ is nonempty and similarly so is Λ .

Suppose for the moment that $\hat{a}_{22} + \hat{b}_{22} \leq c_{22}$. Then increase \hat{a}_{22} to a_{22} and increase

\hat{b}_{22} to b_{22} so that $a_{22} + b_{22} = c_{22}$, and let

$$A = \begin{bmatrix} C_{11} & c_{12} \\ c_{21} & a_{22} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} b_{22} & c_{23} \\ c_{32} & C_{33} \end{bmatrix}.$$

By the observation preceding Theorem 2.8.3, A and B are TN, and $C = A \oplus_1 B$. Thus, if we can show $a_{22} + \hat{b}_{22} \leq c_{22}$, then the proof is complete. So suppose $a_{22} + \hat{b}_{22} > c_{22}$. Then one of two possibilities can occur. Either $\hat{A}[\alpha - \{m\}|\beta - \{m\}]$ is singular, for every $\hat{A}[\alpha|\beta] \in \Gamma$, or there exists $\hat{A}[\alpha_0|\beta_0] \in \Gamma$ such that $\hat{A}[\alpha_0 - \{m\}|\beta_0 - \{m\}]$ is nonsingular. Suppose that $\hat{A}[\alpha - \{m\}|\beta - \{m\}]$ is singular, for every $\hat{A}[\alpha|\beta] \in \Gamma$. In this case, each such $\det \hat{A}[\alpha|\beta]$ does not depend on a_{22} , and hence a_{22} may be decreased without affecting $\det \hat{A}[\alpha|\beta]$. Also, as previously noted, if $m \in \alpha'$, $\beta' \subseteq \{1, 2, \dots, m\}$ and $\det \hat{A}[\alpha'|\beta'] > 0$, then a_{22} may be decreased in this case. However, this contradicts the minimality of a_{22} . Thus there exists $\hat{A}[\alpha_0|\beta_0] \in \Gamma$ such that $\hat{A}[\alpha_0 - \{m\}|\beta_0 - \{m\}]$ is nonsingular. Similar arguments also show that there exists $\hat{B}[\gamma_0|\delta_0] \in \Lambda$ such that $\hat{B}[\gamma_0 - \{1\}|\delta_0 - \{1\}]$ is nonsingular. Furthermore, if a_{22} is decreased, then $\det \hat{A}[\alpha_0|\beta_0] < 0$. Since $a_{22} + \hat{b}_{22} > c_{22}$, decrease a_{22} to a'_{22} such that $a'_{22} + \hat{b}_{22} = c_{22}$. Then

$$C' = \begin{bmatrix} \hat{A}[\alpha_0 - \{m\}|\beta_0 - \{m\}] & c_{12}[\alpha_0 - \{m\}] \\ c_{21}[\beta_0 - \{m\}] & a'_{22} \end{bmatrix} \oplus_1 \hat{B}[\gamma_0|\delta_0],$$

is a submatrix of C . However, $0 \leq \det C' < 0$, by (2.1). This is a contradiction. hence $a_{22} + \hat{b}_{22} \leq c_{22}$, which completes the proof. ■

Recall that totally nonnegative matrices are closed under usual matrix multiplication. Roy Mathias offered the following (in a private communication) simple argument to show that the 1-subdirect sum of totally nonnegative matrices is totally

nonnegative. Suppose A and B are m -by- m and n -by- n totally nonnegative matrices, respectively. Then let $X = A \oplus B$, and define the following $(m+n)$ -by- $(m+n-1)$ totally nonnegative matrix

$$S = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 & \cdots \\ \vdots & \ddots & 1 & 0 & \cdots \\ 0 & \cdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix},$$

where the identity matrix in the upper left corner is m -by- m (the same size as A).

Then a routine computation yields that $S^T X S = A \oplus_1 B$, and since both S and X are totally nonnegative it follows that $A \oplus_1 B$ is totally nonnegative.

It is easy to verify that the class TN is not closed under addition, hence it follows that the answer to question (III) is negative. However, even more can go wrong.

EXAMPLE 2.8.4

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then A and B are both totally nonnegative matrices. However,

$$A \oplus_2 B = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 \\ 0 & 2 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

is not a totally nonnegative matrix since $\det(A \oplus_2 B) = -2$. Note that the sum in the overlapping positions is a totally nonnegative matrix.

To address question (IV), we will make strong use of the fact that TN matrices are LU factorable into TN matrices L and U , but we need rather more than what is known. When the TN matrix C has our form

$$C = \begin{bmatrix} C_{11} & C_{12} & 0 \\ C_{21} & C_{22} & C_{23} \\ 0 & C_{32} & C_{33} \end{bmatrix},$$

we need that, in addition to being lower and upper triangular, respectively, the (correspondingly partitioned) L has its 3,1 block and U its 1,3 block 0. In case C_{11} is invertible, a simple partitioned calculation reveals that this must occur in any LU factorization of C (since L_{11} and U_{11} will be invertible). However, when C_{11} is singular (which implies by Fischer's inequality that C is singular), there can occur TN matrices C of our form that have TN LU factorizations with positive entries in the L_{31} or U_{13} blocks. Fortunately, though, LU factorizations will be highly non unique in this case, and there will always exist ones of the desired form. Thus, an auxiliary assumption that C_{11} is invertible would avoid the need for Lemma 2.8.5 in the proof of Theorem 2.8.6, but this somewhat specialized lemma (perhaps of independent interest, though it requires an elaborate proof) shows that such an auxiliary assumption is unnecessary.

For simplicity of notation (see also [C2]), we let

$$G_2 = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix},$$

and let $F_r(c)$, $c \geq 0$, denote the following r -by- r matrix,

$$F_r(c) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0_{r-2} & 0 \\ c & 0 & 1 \end{bmatrix}.$$

Observe, that both of the above matrices are TN. We also note that the property TN is not in general preserved under row and column operations. However, it follows

(see, for example, [W]), that if $A = [a_{ij}]$ is TN and say $a_{1j}, a_{1,j+1} > 0$, and $a_{1k} = 0$, for $k > j + 1$, then the matrix A^* obtained from A by subtracting $a_{1,j+1}/a_{1j}$ times column j from column $j + 1$ is TN. That is $A = A^*U$, where U is the upper triangular TN matrix given by $U = I_{j-1} \oplus F_2^T(a_{1,j+1}/a_{1j}) \oplus I_{n-j-1}$. This notion was generalized in [C2]. Suppose $A = [a_{ij}]$ is TN, $a_{1j} > 0$, and for some $t > j$, $a_{1t} > 0$, but $a_{1k} = 0$ for all $j < k < t$ and $k > t$. Then, since A is TN it follows that $a_{ik} = 0$ for $1 \leq i \leq n$ and $j < k < t$. Let A^* be obtained from A by subtracting a_{1t}/a_{1j} times column j from column t . Thus $A = A^*U$, where U is the upper triangular TN matrix given by $U = I_{j-1} \oplus F_{t-j+1}^T(a_{1t}/a_{1j}) \oplus I_{n-t}$. It is shown in [C2] that the matrix A^* is TN.

LEMMA 2.8.5 *Let*

$$C = \begin{bmatrix} C_{11} & C_{12} & 0 \\ C_{21} & C_{22} & C_{23} \\ 0 & C_{32} & C_{33} \end{bmatrix},$$

in which C_{22} and C_{33} are square and C_{11} is m -by- m ($m \geq 1$). Suppose C is TN.

Then

$$C = LU = \begin{bmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ 0 & L_{32} & L_{33} \end{bmatrix} \cdot \begin{bmatrix} U_{11} & U_{12} & 0 \\ 0 & U_{22} & U_{23} \\ 0 & 0 & U_{33} \end{bmatrix},$$

in which L and U (partitioned conformally with C) are both TN.

Proof. First, we note that the algorithm given in [C2] applied to C , always produces a U , with $U_{13} = 0$. The proof of the lemma will use induction on m , the size of C_{11} . Suppose $m = 1$, and consider two cases; $C_{11} > 0$ (C_{11} is 1-by-1), or $C_{11} = 0$. If $C_{11} > 0$, then C_{11} is nonsingular and we are done by previous observations. Thus suppose $C_{11} = 0$. Since C is TN, either row or column 1 must be zero. Assume,

without loss of generality, that $C_{12} = 0$; otherwise consider C^T . If $C_{21} = 0$ as well, then by the comments on the previous page, C has a desired LU -factorization.

Therefore assume $C_{21} \neq 0$. By the algorithm given in [C2] C may be written as,

$$C = \begin{bmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \cdot \begin{bmatrix} U_{11} & U_{12} & 0 \\ 0 & U_{22} & U_{23} \\ 0 & 0 & U_{33} \end{bmatrix}.$$

Since $C_{21} \neq 0$, it follows that $U_{11} \neq 0$. However, $L_{31}U_{11} = 0$, hence $L_{31} = 0$. Now assume the result is true for all such C , with C_{11} l -by- l , $1 \leq l \leq m - 1$. Suppose $C = [c_{ij}]$ is as above, with C_{11} m -by- m . Again we consider two cases; $c_{11} = 0$ and $c_{11} > 0$. Suppose $c_{11} = 0$, and assume, without loss of generality, column 1 of C is zero. Let j be the smallest integer (larger than 1) for which $c_{1j} > 0$. If no such j exists, then the result follows easily by induction. Since C is TN it follows that $c_{it} = 0$, for all i, t such that $t < j$. Applying column operations like those discussed before the lemma, C may be written as

$$C = L'U' = \left[\begin{array}{ccc|c} 0 & 0 & c_{1j} & 0 \\ 0 & 0 & c_{2j} & \\ \vdots & \vdots & \vdots & \\ 0 & 0 & c_{nj} & \end{array} \right] \cdot U',$$

(C' is $(n - 1)$ -by- $(n - j)$) where L' and U' are TN. Since the first $j - 1$ columns of C are all 0, observe that L' may be written as

$$L' = \left[\begin{array}{ccc|c} c_{1j} & 0 & 0 & 0 \\ 0 & 0 & c_{2j} & \\ \vdots & \vdots & \vdots & \\ 0 & 0 & c_{nj} & \end{array} \right] \cdot U^{(1,2)}U^{(2,3)} \dots U^{(j-1,j)},$$

in which $U^{(r,r+1)} = I_{r-1} \oplus G_2 \oplus I_{n-r-1}$. Note, when $r = 1$ we ignore the first summand.

It is easy to verify that $U^{(r,r+1)}$ is TN. Let

$$L'' = \left[\begin{array}{ccc|c} c_{1j} & 0 & 0 & 0 \\ 0 & 0 & c_{2j} & \\ \vdots & \vdots & \vdots & C' \\ 0 & 0 & c_{nj} & \end{array} \right]$$

and let $U'' = U^{(1,2)}U^{(2,3)} \dots U^{(j-1,j)}U'$. Then U'' is TN since U'' is a product of TN matrices and L'' is TN since L' is TN, and $C = L''U''$. Now let

$$C'' = \left[\begin{array}{c|c} 0 & c_{2j} \\ \vdots & \vdots \\ 0 & c_{nj} \end{array} \middle| C' \right]$$

(C'' is $(n-1)$ -by- $(n-1)$). By induction, $C'' = \bar{L}\bar{U}$, in which \bar{L} and \bar{U} are TN, and with $\bar{L}_{31} = 0$ and $\bar{U}_{13} = 0$. Then

$$C = \begin{bmatrix} c_{1j} & 0 \\ 0 & \bar{L} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & \bar{U} \end{bmatrix} \cdot U'',$$

is a desired LU factorization. Finally, suppose $c_{11} > 0$, and assume column 1 of C is equal to $[c_{11}, c_{21}, \dots, c_{l1}, 0, \dots, 0, c_{j1}, 0, \dots, 0]^T$, where $c_{l1}, c_{j1} > 0$ and $1 \leq l \leq j-1$. Since C is TN and $c_{i1} = 0$, for $l < i < j$, $c_{it} = 0$, for $l < i < j$ and $1 \leq t \leq n$. Observe that if $L^{(l,j)} = I_{l-1} \oplus F_{j-l+1}(\frac{c_{j1}}{c_{l1}}) \oplus I_{n-j}$, then $L^{(l,j)}$ is TN. Furthermore, C may be written as $C = L^{(l,j)}C'$, in which the first column of C' is equal to $[c_{11}, c_{21}, \dots, c_{l1}, 0, \dots, 0]^T$. Moreover, it follows from a result in [C2], that C' is TN. Applying similar row operations we obtain $C = L'\tilde{C}$, in which L' and \tilde{C} are both TN and with the first column of \tilde{C} given by $[c_{11}, 0, \dots, 0]^T$. Note that L' is equal to a product of lower triangular TN matrices each of the form $L^{(\alpha,\beta)}$. Applying the first step of the algorithm given in [C2] we may write \tilde{C} as

$$\tilde{C} = \tilde{L}U' = \begin{bmatrix} c_{11} & 0 \\ 0 & C'' \end{bmatrix} \cdot U',$$

in which \tilde{L} and U' are both TN. By induction we may write $C'' = L''U''$, in which L'' and U'' are TN, and with $L''_{31} = 0$ and $U''_{13} = 0$. Thus we may write C as

$$C = L' \begin{bmatrix} c_{11} & 0 \\ 0 & L'' \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & U'' \end{bmatrix} U',$$

which is a desired decomposition. This completes the proof. ■

We are now in a position to prove an affirmative answer to question (IV) for the class TN.

THEOREM 2.8.6 *Let*

$$C = \begin{bmatrix} C_{11} & C_{12} & 0 \\ C_{21} & C_{22} & C_{23} \\ 0 & C_{32} & C_{33} \end{bmatrix},$$

in which C_{11} and C_{33} are square, and with C_{22} k -by- k . If C is TN, then C can be written as $C = A \oplus_k B$ so that A and B are TN.

Proof. By Lemma 2.8.5, $C = LU$, in which both L and U are TN, and $L_{31} = 0$ and $U_{13} = 0$. Then it is easy to check that

$$C = LU = \begin{bmatrix} L_{11}U_{11} & L_{11}U_{12} & 0 \\ L_{21}U_{11} & L_{22}U_{22} + L_{21}U_{12} & L_{22}U_{23} \\ 0 & L_{32}U_{22} & L_{33}U_{33} + L_{32}U_{23} \end{bmatrix}.$$

Hence C can be written as

$$C = \begin{bmatrix} L_{11}U_{11} & L_{11}U_{12} & 0 \\ L_{21}U_{11} & L_{21}U_{12} & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & L_{22}U_{22} & L_{22}U_{23} \\ 0 & L_{32}U_{22} & L_{33}U_{33} + L_{32}U_{23} \end{bmatrix}.$$

Notice that if

$$A = \begin{bmatrix} L_{11}U_{11} & L_{11}U_{12} \\ L_{21}U_{11} & L_{21}U_{12} \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \cdot \begin{bmatrix} U_{11} & U_{12} \\ 0 & 0 \end{bmatrix},$$

and

$$B = \begin{bmatrix} L_{22}U_{22} & L_{22}U_{23} \\ L_{32}U_{22} & L_{33}U_{33} + L_{32}U_{23} \end{bmatrix} = \begin{bmatrix} L_{22} & 0 \\ L_{32} & L_{33} \end{bmatrix} \cdot \begin{bmatrix} U_{22} & U_{23} \\ 0 & U_{33} \end{bmatrix},$$

then $C = A \oplus_k B$. Each of the four matrices on the right is easily seen to be TN because L and U are, and it follows from the multiplicative closure of the class TN that both A and B are TN, which completes the proof of the theorem. ■

Chapter 3

Spectral Properties and Jordan Structures

A very important topic in matrix theory is the location and distribution of the eigenvalues of a given matrix or a fixed class of matrices. Some classical results include the discs of Gershgorin and the ovals of Cassini for general complex matrices. Other examples are the inclusion/exclusion wedges for P-matrices, and the celebrated Perron-Frobenius theory for entry-wise nonnegative matrices (see [HJ1, HJ2]).

Approximately sixty years ago Gantmacher and Krein [GK2], who were originally interested in oscillation dynamics, undertook a careful study into the theory of totally nonnegative matrices. Of the many topics they considered, one was the properties of the eigenvalues of totally nonnegative matrices. In particular, they proved that the eigenvalues of a totally positive matrix are real, positive and distinct. We include a proof of this fact for completeness of composition, and also because this particular

proof offers an interesting application of the aforementioned Perron-Frobenius theory. By continuity (see Theorem 2.1.6), for example, it follows that the eigenvalues of a totally nonnegative matrix are real and nonnegative.

Before we prove this result we need some definitions. Given an m -by- n matrix A we define the k^{th} compound of A , which we denote by $C_k(A)$, to be the $\binom{m}{k}$ -by- $\binom{n}{k}$ matrix whose $(i, j)^{\text{th}}$ entry is $\det A[\alpha_i | \beta_j]$, where α_i and β_j are k -subsets, order lexicographically, of $\{1, 2, \dots, m\}$ and $\{1, 2, \dots, n\}$, respectively. It follows from a classical theorem of Kronecker (see also [HJ1]) that if $m = n$, then the eigenvalues of $C_k(A)$ are all of the k -fold products of the eigenvalues of A .

We are now in a position to prove the classical result of Gantmacher and Krein [GK2] on the eigenvalues of a totally positive matrix.

THEOREM 3.0.7 *Let A be an n -by- n totally positive matrix. Then the eigenvalues of A are real, positive and distinct.*

Proof. Suppose $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues of A , arranged such that $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$. By Perron's Theorem (see [HJ1]) it follows that λ_1 is real positive and strictly larger in modulus than any other eigenvalue of A . Suppose, to the contrary, that λ_j ($j \geq 2$) is the first eigenvalue of A that is not a real number. It is clear that the product $\lambda_1 \lambda_2 \cdots \lambda_j$ is an eigenvalue of $C_j(A)$ of maximum modulus, and hence since $C_j(A)$ is an entry-wise positive matrix it follows from Perron's Theorem that $\lambda_1 \lambda_2 \cdots \lambda_j$ is a real positive number. Applying similar arguments to $C_{j-1}(A)$ implies that $\lambda_1 \lambda_2 \cdots \lambda_{j-1}$ is a real positive number. Consequently, λ_j is a real positive number. Therefore, every eigenvalue of A is real and also positive.

Applying Perron's Theorem to $C_j(A)$ yields, $\lambda_1 \lambda_2 \cdots \lambda_j > \lambda_1 \lambda_2 \cdots \lambda_{j-1} \lambda_{j+1}$, which implies $\lambda_j > \lambda_{j+1}$. Thus, the eigenvalues of A are distinct. ■

A simple consequence of the previous result (also observed in [GK2] and in [GK1]) is that the eigenvalues of an oscillatory matrix are real, positive and distinct. Another less obvious consequence, which follows from a continuity argument (using Theorem 2.1.6), is that the eigenvalues of a totally nonnegative matrix are real and nonnegative. See also [R] for more background information on classes of matrices with real distinct spectrum. See [Fr] and [Pil] for weak interlacing properties of the eigenvalues of totally positive matrices, and see [Ev] for some remarks on the distribution of eigenvalues of oscillatory matrices. In [BJ1] it is shown that any collection of n positive distinct numbers can be realized as the spectrum of an n -by- n totally positive matrix. (For other results along these lines we refer the reader to [G2].)

For the most part this is the extent of what is known about the properties of the eigenvalues of a totally nonnegative matrix. It is our purpose here to explore further the properties of the eigenvalues of a general totally nonnegative matrix.

We say that an n -by- n ($n \geq 2$) matrix A is *reducible* if there exists a permutation matrix P (a *permutation* matrix P is an n -by- n (0,1) matrix that satisfies $PP^T = P^T P = I$), so that

$$PAP^T = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix},$$

where the matrix 0 is an $(n-r)$ -by- r zero matrix ($n-1 \geq r \geq 1$). Otherwise we say A is an *irreducible* matrix. We begin our discussion by assuming that A is irreducible.

This assumption is not without loss of generality assumption, but is made to avoid certain trivialities. Observe that any collection of n nonnegative real numbers may be realized as the eigenvalues of a reducible totally nonnegative matrix, simply by using entry-wise (and hence totally) nonnegative diagonal matrices. The next result was originally observed in [GK2], and can also be found in [R], although stated slightly differently. The proof we present here utilizes the fact that certain totally nonnegative matrices are in double echelon form. This fact will be used later in this chapter.

LEMMA 3.0.8 *Suppose $A = [a_{ij}]$ is a totally nonnegative matrix. Then A is irreducible if and only if $a_{ij} > 0$ for all i, j such that $|i - j| \leq 1$.*

Proof. The sufficiency of the condition that $a_{ij} > 0$ for all i, j with $|i - j| \leq 1$ is trivial. On the other hand suppose A is irreducible. Thus A has no zero lines, and hence must be in double echelon form. Suppose $a_{ii} = 0$ for some i . Then $a_{st} = 0$ for $1 \leq s \leq i$ and $i \leq t \leq n$, or $a_{st} = 0$ for $i \leq s \leq n$ and $1 \leq t \leq i$, from which it follows that A is reducible, and hence we have a contradiction. Therefore we may assume $a_{ii} > 0$ for all i . Similarly, if $a_{i,i+1} = 0$ for some $i = 1, 2, \dots, n - 1$, then $a_{st} = 0$ for all $1 \leq s \leq i$ and $i \leq t \leq n$. Again A is reducible, another contradiction. This completes the proof. ■

Note that if A is a TP_2 matrix, then in fact, it follows that A is irreducible if and only if $a_{ij} > 0$ for all i, j such that $|i - j| \leq 1$ (see Theorem 2.5.4). We call an n -by- n matrix $A = [a_{ij}]$ *tridiagonal* if $a_{ij} = 0$, for all i, j with $|i - j| \geq 2$. Hence tridiagonal matrices and irreducible totally nonnegative matrices are somewhat

related (see lemma above). The following result is well-known and many elementary proofs of this result have been presented, however, a more difficult problem is locating a proof of this fact in the literature.

LEMMA 3.0.9 *Let T be an n -by- n irreducible entry-wise nonnegative tridiagonal matrix. Then the eigenvalues of T are real and distinct.*

It is well-known (see [A, GK2]) that a nonsingular irreducible totally nonnegative matrix is oscillatory, and hence has positive distinct spectrum. Henceforth we consider (for the next section) only irreducible totally nonnegative matrices that are singular.

3.1 Main Result

Before we come to our key lemma for this section we include a brief discussion concerning elementary row operations applied to totally nonnegative matrices.

As in Section 1.8 (see also [C2]), we let $F_r(c)$, $c \geq 0$, denote the following r -by- r matrix,

$$F_r(c) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0_{r-2} & 0 \\ c & 0 & 1 \end{bmatrix}.$$

Observe, that $F_r(c)$ is a totally nonnegative matrix. Unfortunately, the property of being totally nonnegative is not in general preserved under row and column operations. However, in [W] it is proved that if $A = [a_{ij}]$ is totally nonnegative and $a_{1j} > 0$ and $a_{1k} = 0$ for $k > j + 1$, then the matrix A^* obtained from A by subtracting $a_{1,j+1}/a_{1j}$ times column j from column $j + 1$ is totally nonnegative. In this

case $A = A^*U$, where U is the upper triangular totally nonnegative matrix given by $U = I_{j-1} \oplus F_2^T(a_{1,j+1}/a_{1j}) \oplus I_{n-j-1}$. This result was generalized in [C2]. Suppose $A = [a_{ij}]$ is totally nonnegative, $a_{1j} > 0$, and for some $t > j$, $a_{1t} > 0$, but $a_{1k} = 0$ for all $j < k < t$ and $k > t$. Then, since A is totally nonnegative the restrictions concerning the zero-nonzero patterns imply that $a_{ik} = 0$ for $1 \leq i \leq n$ and $j < k < t$. Let A^* be obtained from A by subtracting a_{1t}/a_{1j} times column j from column t . Thus $A = A^*U$, where U is the upper triangular totally nonnegative matrix given by $U = I_{j-1} \oplus F_{t-j+1}^T(a_{1t}/a_{1j}) \oplus I_{n-t}$. In [C2] it is proved that the matrix A^* is totally nonnegative.

We begin our analysis with a central lemma, referred to as the “basic lemma”, from which the main result may be obtained by a sequential application of this lemma. In [RH] a similar “reduction” type basic lemma was proved for lower Hessenberg (a matrix $B = [b_{ij}]$ is said to be a *lower Hessenberg matrix* if $b_{ij} = 0$, for all i, j with $i + 1 < j$) totally nonnegative matrices, and later extended to general totally nonnegative matrices. This lemma in [RH] was then used to prove Theorem 2.2.3 for nonsingular totally nonnegative matrices. Since the auxiliary assumption of irreducibility is necessary for our analysis (and was not in [RH]), we are required to prove a different reduction type lemma (see Lemma 3.1.2). Firstly, consider the following definition.

DEFINITION 3.1.1 Let A and B be two square matrices, not necessarily of the same size. Then we say that A and B have the *same nonzero Jordan structure* if the nonzero eigenvalues of A and B can be put into 1-1 correspondence so that each

corresponding pair has the same number and sizes of Jordan blocks. Further, if A and B have the same size, we say that A and B have the same *qualitative Jordan structure*, if they have the same nonzero Jordan structure, and if the number and sizes of the Jordan blocks corresponding to zero coincide.

For example, suppose A has eigenvalues 2 (with one 2-by-2 Jordan block) and 3 (with one 3-by-3 Jordan block). Then any 5-by-5 matrix with two nonzero eigenvalues, one of which has 1 2-by-2 Jordan block and the other with a 3-by-3 Jordan block has the same nonzero and qualitative Jordan structure as A . Recall that if A is an n -by- m matrix and B is a m -by- n matrix then AB and BA have the same nonzero Jordan structure, and two matrices that are similar have the same qualitative Jordan structure.

We are now in a position to state and prove the “basic lemma”.

LEMMA 3.1.2 (Basic Lemma) *Suppose $A = [a_{ij}]$ is an n -by- n irreducible totally nonnegative matrix in which $a_{1,j+1} > 0$, for some j with $2 \leq j + 1 \leq n$ and $a_{1t} = 0$ for all $t > j + 1$ (if $j + 1 = n$, then ignore t). Then there exists an irreducible totally nonnegative matrix A' such that:*

(i) $(A')_{1,j+1} = 0$; and

(ii) A' is either n -by- n and similar to A , or is $(n - 1)$ -by- $(n - 1)$ and has the same nonzero Jordan structure as A .

Proof. Define i_1 to be the smallest row index such that $a_{i_1,j+2} > 0$. (Note that since A is irreducible $i_1 \leq j + 1$.) If $j + 1 = 2$, then proceed to row i_1 . Since

A is irreducible, $a_{st} > 0$ for all s, t with $|s - t| \leq 1$ (by Lemma 3.0.8), hence it follows that $i_1 = 2$, in the case when $j + 1 = 2$. Otherwise $j + 1 > 2$. Use column $j \geq 2$ to eliminate $a_{1,j+1}$ (note that since A is in double echelon form $a_{1j} > 0$), via the elementary upper triangular bidiagonal nonsingular matrix $S = I - \alpha e_j e_{j+1}^T$. Consider the matrix $A' = S^{-1}AS$. It is shown in [W] that AS is totally nonnegative, and since S^{-1} is totally nonnegative, we have that A' is totally nonnegative. Clearly the $(1, j + 1)^{st}$ entry of A' is zero. Observe that A' will be in double echelon form, unless A' contains a zero column, which necessarily must be the $(j + 1)^{st}$ column of A' . Assume for the moment that the $(j + 1)^{st}$ column of A' is nonzero. Then we must show that A' is irreducible. There are two cases to consider: (1): $i_1 < j + 1$; or (2): $i_1 = j + 1$. Suppose that the first nonzero entry in column $j + 1$ of A' is in row t , where $2 \leq t \leq i_1$. If $i_1 < j + 1$, then we have $t \leq i_1 < j + 1$ and in this case let $A' = [c_{kl}]$. Thus $c_{kl} > 0$ whenever $|k - l| \leq 1$, so in particular, $c_{j,j+1} > 0$, $c_{j+1,j+1} > 0$, and $c_{j+2,j+1} > 0$. Otherwise, suppose $i_1 = j + 1$. Now if $t < i_1$, then the same reasoning as above applies. Therefore assume that $t = i_1 = j + 1$. After eliminating $a_{1,j+1}$, the worst possible case for column $j + 1$ of AS is that it has the form $[0, 0, \dots, 0, *, \dots]^T$, where the first nonzero occurs in row $j + 1$. However, upon premultiplying AS by S^{-1} we add a positive multiple of row $j + 1$ to row j . By the definition of t , it follows that the $(j + 1, j + 1)^{st}$ entry of AS is nonzero, and hence $c_{j,j+1} > 0$. Consider the entry $c_{j+2,j+1}$. We claim that this entry is positive. To prove this suppose this was not the case, i.e., assume $c_{j+2,j+1} = 0$. But in this case $a_{j+2,j}$ must be positive since it was used to eliminate this entry. However, since A' is in

double echelon (because it does not have any nonzero columns) it is not possible that $c_{j+2,j+1} = 0$. Therefore A' is irreducible since $c_{kl} > 0$ for all k, l with $|k - l| \leq 1$. We remark here that adding row $j + 1$ to row j might (inadvertantly) alter the zero nonzero pattern above the superdiagonal and to the right of column $j + 1$. But this does not affect the proof, since the resulting matrix will still be in double echelon form, as needed. Finally, suppose the $(j + 1)^{st}$ column of A' is zero. Then (as in [RH] and [C2]) consider the matrix obtained by deleting the $(j + 1)^{st}$ row and column of A' , and denote it by A'' . Since A'' can be obtained from A' by a projection, the nonzero Jordan structure of A'' is the same as A' , which in turn is the same as A . Moreover, since A'' is a submatrix of A' , A'' is totally nonnegative. The only point remaining to prove is that A'' is irreducible. Assume that the j^{th} , $(j + 1)^{st}$ and $(j + 2)^{nd}$ columns of A' are $[*, *, \dots, *, ?, \dots]^T$ (here the ? occurs in the row $j + 2$), $[0, 0, \dots, 0]^T$, and $[\dots, ?, *, *, \dots]^T$ (here the ? occurs in row j), respectively. To show A'' is irreducible it is enough to show that $c_{j+2,j} > 0$ and $c_{j,j+2} > 0$, where $A' = [c_{kl}]$. Since A' is obtained from AS by adding a positive multiple of row $j + 1$ to row j , it follows that $c_{j,j+2} > 0$, as $a_{j+1,j+2} > 0$, by the irreducibility of A . For the purpose of a contradiction, assume that $c_{j+2,j} (= a_{j+2,j}) = 0$. Since A is in double echelon form it follows that $a_{kj} (= c_{kj}) = 0$, for all $k \geq j + 2$. However, column j was used to annihilate column $j + 1$, and hence $a_{k,j+1} = 0$, for all $k \geq j + 2$. But this contradicts the irreducibility of A . Therefore $c_{j+2,j} > 0$ and $c_{j,j+2} > 0$, and thus A'' is irreducible. Thus completes the proof of the basic lemma. ■

We are now in a position to state our main results concerning the eigenvalues of

an irreducible totally nonnegative matrix.

THEOREM 3.1.3 *Let A be an irreducible totally nonnegative matrix. Then there exists a irreducible tridiagonal totally nonnegative matrix T (not necessarily of the same size) with the same nonzero Jordan structure as A . Moreover, T is obtained from A by a sequence of similarity transformations and projections.*

Proof. Successive application of the basic lemma (Lemma 3.1.2) results in a k -by- k ($k \leq n$) irreducible lower Hessenberg totally nonnegative matrix L , which has the same nonzero Jordan structure as A . Consider the matrix $H = L^T$. Apply successive application of the basic lemma to H , so as to reduce H to an irreducible totally nonnegative tridiagonal matrix T . Observe that the zero-nonzero pattern of H below the subdiagonal is preserved by successive application of the basic lemma. Furthermore, if a zero column is not encountered along the way, then it is clear that the resulting matrix is an irreducible totally nonnegative tridiagonal matrix. Thus, suppose that at some stage in the reduction process a zero column is produced, say the j^{th} column is the first zero column obtained in this process. In this case column $j - 1$ was used to annihilate column j . Since H is upper Hessenberg (the transpose of a lower Hessenberg matrix), $h_{t,j-1} = 0$ for $t \geq j + 1$. Thus in order for column $j - 1$ to annihilate column j , $h_{t,j} = 0$ for $t \geq j + 1$. However, this contradicts the irreducibility of H for the case $j < n$. Therefore only the final column of H can possibly be annihilated. Hence it follows that after applying the basic lemma repeatedly to H , the resulting matrix T is an irreducible totally nonnegative tridiagonal with the same nonzero Jordan structure as H . Since any matrix is similar to its transpose (see

[HJ1]) it follows that T has the same nonzero Jordan structure as A . The fact that T is obtained from A by a sequence of similarity transformations and projections is evident from applying the basic lemma, and using the fact that any matrix is similar to its transpose. This completes the proof. ■

THEOREM 3.1.4 *Let A be an n -by- n irreducible totally nonnegative matrix. Then the positive eigenvalues of A are distinct.*

Proof. By the previous theorem there exists an irreducible totally nonnegative tridiagonal matrix T , with the same nonzero Jordan structure as A . By Lemma 3.0.9 the positive eigenvalues of T are distinct, hence the positive eigenvalues of A are distinct. ■

COROLLARY 3.1.5 *The eigenvalues of a totally positive matrix are real positive and distinct.*

We note here that the size of the tridiagonal matrix obtained in Theorem 3.1.3 is either the same as the number of nonzero eigenvalues of A , or is this number plus one. In the next section we will see that this quantity (namely, the number of nonzero eigenvalues of A) will play a central role in our analysis of the qualitative Jordan structures of totally nonnegative matrices.

3.2 Jordan Structures of Totally Nonnegative Matrices

In an attempt to characterize all possible (qualitative) Jordan structures of totally nonnegative matrices, we take the first natural step and consider all possible ranks and principal ranks of totally nonnegative matrices. For more information of the Jordan canonical form of a matrix, and its implications see [HJ1]. (Recall the definitions of nonzero and qualitative Jordan structure from the previous section.) Recall that the *rank* of a given m -by- n matrix A , denoted by $\text{rank}(A)$, is the size of the largest invertible square submatrix of A . Naturally, the *principal rank* of an n -by- n matrix A , denoted by $\text{prank}(A)$, is the size of the largest invertible principal submatrix of A (see also chapter 2, section 2.7 for similar notions). Note that the inequality, $1 \leq \text{prank}(A) \leq \text{rank}(A) \leq \min(m, n)$ follows directly from the definitions above. One topic of interest is characterizing all the triples $(n, \text{rank}(A), \text{prank}(A))$, where n is the size of the totally nonnegative matrix A .

Firstly, we note that a k -by- k Jordan block corresponding to a nonnegative eigenvalue is a totally nonnegative matrix (see also chapter 5, section 5.2). Hence in the reducible case all possible Jordan structures (or all possible Jordan canonical forms) exist for the class of reducible totally nonnegative matrices. Henceforth, we consider only totally nonnegative matrices that are irreducible. Recall that in this case (see previous section), the positive eigenvalues of every such totally nonnegative matrix are distinct. Hence the Jordan blocks corresponding to the positive eigenvalues are

necessarily 1-by-1. Therefore the only issue left to consider is the sizes and the number of Jordan blocks corresponding to the eigenvalue zero.

It follows from Lemma 2.7.4 that if A is a P_0 -matrix (and in particular totally nonnegative), then $\text{prank}(A)$ is equal to the number of nonzero (positive) eigenvalues of A . Hence if A is an n -by- n totally nonnegative matrix, then $n - \text{prank}(A)$ is equal to the sum of the sizes of the Jordan blocks corresponding to the eigenvalue zero of A , and $n - \text{rank}(A)$ is equal to the number of Jordan blocks corresponding to zero. Another well-known notion is how to determine the size of the largest Jordan block corresponding to the eigenvalue zero (in fact it is known for any eigenvalue). In the case of irreducible totally nonnegative matrices this reduces to the following: if k is the smallest positive integer such that $\text{rank}(A^k) = \text{prank}(A)$, then k is equal to the size of the largest Jordan block corresponding to the eigenvalue zero. Observe that k , as defined above, satisfies $k \leq \text{rank}(A) - \text{prank}(A) + 1$.

We begin by considering the following examples which demonstrate that there do exist nontrivial (larger than one) Jordan blocks corresponding to zero for irreducible totally nonnegative matrices.

EXAMPLE 3.2.1 Let $A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix}$. Then A is a 3-by-3 totally positive matrix.

Consider the 4-by-4 irreducible totally nonnegative matrix,

$$B = \begin{bmatrix} 3 & 3 & 2 & 1 \\ 2 & 2 & 3 & 2 \\ 1 & 1 & 2 & 3 \\ 1 & 1 & 2 & 3 \end{bmatrix}.$$

Then $\text{rank}(B) = 3$ and $\text{prank}(B) = 2$, from which it follows that B has one 2-by-2 Jordan block corresponding to zero, and two positive distinct eigenvalues.

We note here that in general this “asymmetric” bordering of a totally nonnegative matrix preserves the rank, but may change the principal rank. Observe, that if we border the matrix B above in a similar manner, then the resulting totally nonnegative matrix has the same rank and principal rank as B . We will come back to this asymmetric bordering later in this section.

EXAMPLE 3.2.2 Consider the n -by- n lower Hessenberg $(0,1)$ -matrix

$$H = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & \cdots & 1 & 1 \end{bmatrix}.$$

Then, by Lemma 2.5.7, H is totally nonnegative matrix, and it is clear that H is irreducible (see Lemma 3.0.8). Moreover, it is not difficult to verify that the rank(A) = $n - 1$ (observe, that H is singular, and that $H[\{1, 2, \dots, n - 1\}|\{2, 3, \dots, n\}]$ is a nonsingular lower triangular matrix). A more difficult exercise is to prove that $\text{prank}(A) = \lceil \frac{n}{2} \rceil$. To prove this observe that if n is odd (even), then $H[\{1, 3, 5, \dots, n\}]$ ($H[\{2, 4, 6, \dots, n\}]$) is a nonsingular lower triangular matrix. Hence $\text{prank}(H) \geq \lceil \frac{n}{2} \rceil$. To show $\text{prank}(H) \leq \lceil \frac{n}{2} \rceil$, suppose there exists an index set α such that $|\alpha| > \lceil \frac{n}{2} \rceil$, and $\det H[\alpha] > 0$. Then α must contain at least one consecutive pair of indices, say i and $i + 1$ are in α , where $1 < i < n - 1$. Since H is lower Hessenberg and $\det H[\alpha] > 0$, it follows that index $i + 2 \in \alpha$. Applying the same reasoning to the pair $i + 1$ and $i + 2$, we may conclude $i + 3 \in \alpha$. However, continuing in this manner will show that $H[\alpha]$ is singular, since either both indices $n - 1$ and n will be in α , or the maximum of α is less than n , in which case there will exist a pair of indices $k, k + 1$ in α and

$k + 2$ not in α .

The fact that any totally nonnegative matrix has a bidiagonal factorization can also be employed to tackle certain problems concerning Jordan structures. In particular, the quantities rank and principal rank of a given matrix and irreducibility can all be interpreted via the diagram associated with a particular bidiagonal factorization of a totally nonnegative matrix. For example, from a given diagram we can verify whether or not the associated totally nonnegative matrix is irreducible by determining if there exists a path in this diagram from any index i to each of $i - 1$, i and $i + 1$ (ignoring $i - 1$, when $i = 1$ and $i + 1$, when $i = n$). If such a path exists, then for each i , $a_{i,i-1}$, a_{ii} and $a_{i,i+1}$ are all positive, which implies by Lemma 3.0.8, that the associated totally nonnegative matrix is irreducible. Since rank and principal rank are defined in terms of nonsingular submatrices, it follows that rank and principal rank can be interpreted as the largest collection of vertex disjoint paths beginning on the left and terminating on the right, in the diagram, and the largest collection of vertex disjoint paths which begin and terminate in the same index set, respectively.

We begin our analysis by considering the triple $(n, \text{rank}, \text{prank})$ among the class of irreducible totally nonnegative matrices. Firstly, observe that the triples $(n, 1, 1)$ and (n, n, n) certainly exist, for all $n \geq 1$, by considering the matrix J of all ones, and any n -by- n totally positive matrix, respectively. Thus for $n \leq 2$, we have completely characterized all possible triples. However, for $n = 3$, the triples $(3, 2, 2)$ and $(3, 2, 1)$ have not been shown to be realizable. Consider, firstly, the triple $(3, 2, 2)$, and suppose that

$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is a 2-by-2 totally positive matrix. Then the matrix $B = \begin{bmatrix} a & b & b \\ c & d & d \\ c & d & d \end{bmatrix}$ is a 3-by-3 irreducible totally nonnegative matrix with $\text{rank}(B) = \text{prank}(B) = 2$.

Hence the triple $(3, 2, 2)$ is realizable. What about the triple $(3, 2, 1)$? Suppose there exists an irreducible totally nonnegative matrix $A = [a_{ij}]$ with $\text{rank}(A) = 2$ and $\text{prank}(A) = 1$. Then $a_{ij} > 0$, for all i, j with $|i - j| \leq 1$. Observe that multiplying A by a positive diagonal matrix does not affect rank or principal rank. Hence we may assume that $a_{11} = a_{22} = a_{33} = 1$ and $a_{12} = a_{21}$, $a_{23} = a_{32}$. Thus $A = \begin{bmatrix} 1 & a_{12} & a_{13} \\ a_{12} & 1 & a_{23} \\ a_{31} & a_{23} & 1 \end{bmatrix}$. Since $\text{prank}(A) = 1$ it follows that $a_{12} = a_{23} = 1$ and $a_{13}a_{31} = 1$. It is not difficult to determine that if $a_{13}a_{31} = 1$, then $a_{13} = a_{31} = 1$, as A is totally nonnegative. However, in this case $A = J$, which is a contradiction since $\text{rank}(A) = 2$. Thus the triple $(3, 2, 1)$ is not realizable. This leads us to our first result on these triples, with fixed principal rank. Recall that $l_i = 0$ (or $u_j = 0$) means that the corresponding edge does not appear in the diagram.

PROPOSITION 3.2.3 *The triple $(n, k, 1)$ is realizable by an n -by- n irreducible totally nonnegative matrix if and only if $k = 1$.*

Proof. We have already seen that the triple $(n, 1, 1)$ is realizable for all n . Now assume the triple $(n, k, 1)$ is realizable by an n -by- n irreducible totally nonnegative matrix A . Then A has a bidiagonal factorization, which can be represented by an associated diagram. Since A is irreducible there exists a path from index n to n . Let P denote a shortest such path from n to n . Then we claim that P must intersect the bottom row of this diagram. If not, then since there always exists a path from index 1

to 1 by going along this bottom row, it follows that the principal rank of A would be at least 2, which is a contradiction (see diagram on the left in Figure 3.1). Otherwise, P intersects the bottom row (see diagram on the right in Figure 3.1), then since P is a shortest path it follows that any maximal collection of vertex disjoint paths must intersect P . Hence the rank of A is at most one. Since $\{P\}$ is one maximal collection the rank of A is one. ■

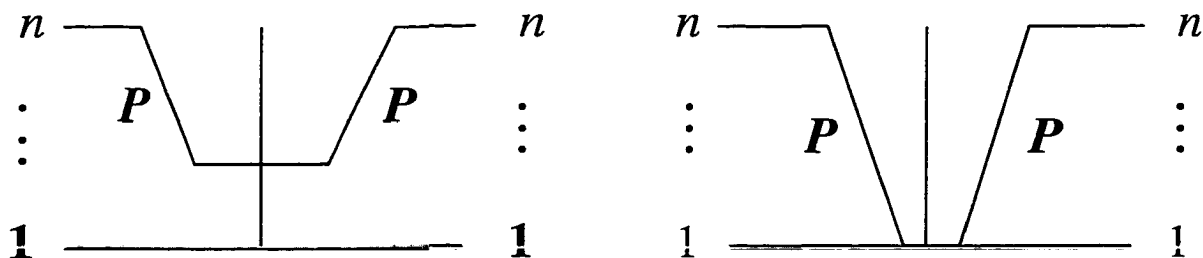


Figure 3.1: Fixed Principal Rank One.

We note here that the above proposition could also have been proved in a similar manner by proving that the triple $(3, 2, 1)$ is not realizable.

Observe that if $\text{rank}(A) = \text{prank}(A)$, then A has $n - \text{prank}(A)$, 1-by-1 Jordan blocks corresponding to zero, and if $\text{rank}(A) = \text{prank}(A) + 1$, then A has exactly one 2-by-2 Jordan block and $n - \text{prank}(A) - 1$, 1-by-1 Jordan blocks corresponding to zero.

We now move onto the case when the principal rank is two.

PROPOSITION 3.2.4 *Suppose the triple $(n, k, 2)$ is realizable by an n -by- n irreducible totally nonnegative matrix. Then $2 \leq k \leq \lceil \frac{n+1}{2} \rceil$. Moreover, each such k is realizable.*

Proof. First observe that $k \geq 2$ is obvious. As before we may assume that the given matrix has a bidiagonal factorization and an associated diagram. Choose a shortest path P from n to n . Since $\text{prank}(A) = 2$, this path cannot intersect that bottom row. Suppose the path P uses the edge with weight d_i . Since $\text{prank}(A) = 2$, it follows that any path beginning at $i - 1$ which is disjoint from P must intersect the bottom row. Hence any maximal collection of vertex disjoint paths from the indices $\{1, 2, \dots, i - 1\}$ that are disjoint from P contains at most one path. In this case it follows that $\text{rank}(A) = k \leq \max_{2 \leq i \leq n} \{\min(i, n - i + 2)\}$. Hence $k \leq \lceil \frac{n+1}{2} \rceil$. To show that every such triple $(n, k, 2)$ with $2 \leq k \leq \lceil \frac{n+1}{2} \rceil$ can be realized, consider the diagram in Figure 3.2 where $i = 2, 3, \dots, n$. ■

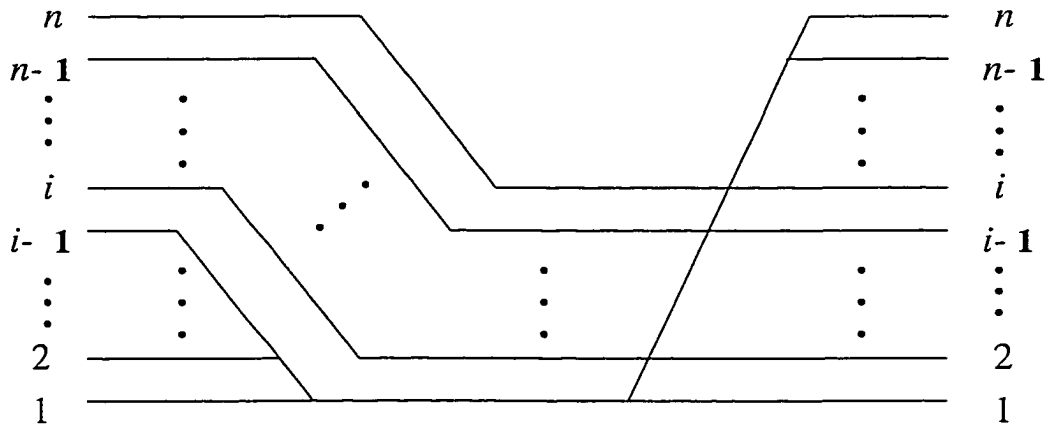


Figure 3.2: Fixed Principal Rank Two.

We can also apply similar techniques to prove the next result. Recall that the $(0,1)$ lower Hessenberg matrix discussed previously has rank equal to $n - 1$ and principal rank equal to $\lceil \frac{n}{2} \rceil$. The next result shows that $\lceil \frac{n}{2} \rceil$ is the smallest possible value for principal rank in the case when rank is $n - 1$.

PROPOSITION 3.2.5 *Suppose the triple $(n, n - 1, k)$ ($n \geq 2$) is realizable by an n -by- n irreducible totally nonnegative matrix. Then $\lceil \frac{n}{2} \rceil \leq k \leq n - 1$. Moreover, for each such k , the triple is realizable.*

Proof. First observe that the inequality $k \leq n - 1$ is obvious. The proof is by induction on n . The claim has already been verified for $n \leq 3$. Let A be an n -by- n irreducible totally nonnegative matrix with rank equal to $n - 1$. As before we assume A has a bidiagonal factorization and an associated diagram, and suppose P denotes the shortest path from n to n . Then there are three cases to consider: (1): P uses the edge with weight d_i with $i \leq n - 2$; (2): P uses the edge with weight d_{n-1} ; or (3): P uses the edge with weight d_n .

Case 1: If P uses the edge with weight d_i with $i \leq n - 2$, then it is not difficult to check that, since P was a shortest such path, the rank of A is at most $n - 2$, a contradiction.

Case 2: If P uses the edge with weight d_{n-1} , then since A has rank $n - 1$ the shortest path from each index j to itself ($1 \leq j \leq n - 2$) can drop at most one level. Moreover, the shortest path from index $n - 2$ to itself cannot intersect P (otherwise $\text{rank}(A) < n - 1$). Consider the diagram induced by the vertices $\{1, 2, \dots, n - 2\}$. Observe that for this diagram the associated totally nonnegative matrix A' (which is not a submatrix of A), satisfies $n - 3 \leq \text{rank}(A') \leq n - 2$. If $\text{rank}(A') = n - 2$, then since the vertex disjoint paths that constitute $\text{rank}(A')$ do not intersect P we have $\text{prank}(A) = n - 1$. If $\text{rank}(A') = n - 3$, then by induction $\lceil \frac{n-2}{2} \rceil \leq \text{prank}(A') \leq n - 3$, and every such value is achievable. Hence $\lceil \frac{n-2}{2} \rceil + 1 \leq \text{prank}(A) \leq n - 3 + 1$ or

$\lceil \frac{n}{2} \rceil \leq \text{prank}(A) \leq n - 2$, and every such value is realizable.

Case 3: Suppose P uses the edge with weight d_n (i.e., P goes straight across the diagram). Consider the diagram induced by the indices $\{1, 2, \dots, n - 1\}$. By induction, the associated totally nonnegative matrix A' is $(n-1)$ -by- $(n-1)$ with $\text{rank}(A') = n - 2$, hence $\lceil \frac{n-1}{2} \rceil \leq \text{prank}(A') \leq n - 2$, and every such value is achievable. Thus it follows that $\lceil \frac{n-1}{2} \rceil + 1 \leq \text{prank}(A) \leq n - 1$, and every such value is achievable. This completes the proof. ■

We are now in a position to classify all possible triples for $n = 4$.

EXAMPLE 3.2.6 Suppose the triple $(4, k, p)$ is realizable by a 4-by-4 irreducible totally nonnegative matrix. The triples $(4, 4, 4)$ and $(4, 1, 1)$ are certainly realizable (by Proposition 3.2.3, the triple $(4, 1, 1)$ is the only possible triple with principal rank equal to one). If the rank (k) is fixed to be 3, then, by Proposition 3.2.5, the only realizable triples are $(4, 3, 3)$ and $(4, 3, 2)$. Similarly, if the rank is fixed at 2, then by Proposition 3.2.4 the triple $(4, 2, 2)$ is the only realizable triple. Hence all possible triples have been determined.

We now move onto a more general result on the Jordan structure of totally nonnegative matrices. Recall that in general that if A is an n -by- n matrix and D is a positive diagonal matrix, then the Jordan structure of A and AD can be vastly different. (For example, it is known that if A is a P-matrix, then there exists a positive diagonal matrix D so that the eigenvalues of AD are positive and distinct, even though A may not even be diagonalizable.) However, in the case of irreducible

totally nonnegative matrices it turns out that the Jordan structure of A and AD (D a positive diagonal matrix) coincide. We begin with the following lemma.

LEMMA 3.2.7 *Let A be an n -by- n totally nonnegative matrix and suppose D an n -by- n positive diagonal matrix. Then*

$$\text{rank}((AD)^k) = \text{rank}(A^k) \quad \text{and} \quad \text{prank}((AD)^k) = \text{prank}(A^k),$$

where $k \geq 1$.

Proof. Let $C_j(A)$ denote the j^{th} compound of A . Since D is a positive diagonal matrix it follows that $C_j(D) = D^j$, is a positive diagonal matrix for all j . Hence $C_j(AD) = C_j(A)C_j(D) = C_j(A)D^j$, where the first equality follows from the Cauchy-Binet identity for determinants (see (1.4)). Since D^j is a positive diagonal matrix the zero/nonzero patterns of $C_j(AD)$ and $C_j(A)$ are the same. Moreover, since $C_j(A)$ and $C_j(AD)$ are entry-wise nonnegative matrices and $C_j(A^k) = (C_j(A))^k$ it follows that the zero/nonzero patterns of each $C_j(A^k)$ is completely determined by $C_j(A)$. Since the zero/non-zero patterns of $C_j(AD)$ and $C_j(A)$ are the same, it follows that the zero/nonzero patterns of $C_j(A^k)$ and $C_j((AD)^k)$ are the same. Observe that the rank and the principal rank of a given matrix is given by the largest j , such that j^{th} compound is nonzero, and the largest j , such that the j^{th} compound has a nonzero diagonal, respectively. Hence it follows that $\text{rank}((AD)^k) =$

$\text{rank}(A^k)$ and $\text{prank}((AD)^k) = \text{prank}(A^k)$, where $k \geq 1$. This completes the proof.

■

We are now in a position to prove that the Jordan structure of A and AD are the same, whenever A is totally nonnegative and irreducible.

THEOREM 3.2.8 *Suppose A is an n -by- n irreducible totally nonnegative matrix and D is positive diagonal matrix. Then A and AD have the same qualitative Jordan structure.*

Proof. Since A is irreducible (and hence AD is) and since $\text{prank}(AD) = \text{prank}(A)$, we have that the number of distinct positive eigenvalues of A and AD are equal. Moreover, since the number and sizes of the Jordan blocks corresponding to zero are completely determined by $\text{rank}(A^k)$ for $k \geq 1$, it follows that A and AD have the same qualitative Jordan structure, since $\text{rank}((AD)^k) = \text{rank}(A^k)$, for $k \geq 1$ (by Lemma 3.2.7). ■

The assumption of irreducibility in the above result is necessary as seen by the following example. Let $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$. Then A is TN and is itself a Jordan block, and hence is not diagonalizable. However, if $D = \text{diag}(1, 2, 3)$, then $AD = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 3 \\ 0 & 0 & 3 \end{bmatrix}$, which has distinct eigenvalues and hence is diagonalizable. Thus A and AD do not have the same qualitative Jordan structure.

We now present of couple of interesting consequences to the above theorem.

COROLLARY 3.2.9 *Suppose A is an n -by- n irreducible totally nonnegative matrix partitioned as follows, $A = \begin{bmatrix} A_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, where A_{11} is $(n-1)$ -by- $(n-1)$ and a_{22} is*

a scalar. Define the $(n + 1)$ -by- $(n + 1)$ matrix B as follows

$$B = \begin{bmatrix} A_{11} & a_{12} & a_{12} \\ a_{21} & a_{22} & a_{22} \\ a_{12} & a_{22} & a_{22} \end{bmatrix}.$$

Then:

- (i) B is totally nonnegative and irreducible,
- (ii) $\text{rank}(B) = \text{rank}(A)$ and $\text{prank}(B) = \text{prank}(A)$,
- (iii) The Jordan structure of B is the same as A , except B has one more 1-by-1 Jordan block associated with the eigenvalue zero.

Proof. Let $S = E_n(-1)$, the n -by- n elementary bidiagonal matrix with a -1 in the $(n, n - 1)$ entry. Then an easy calculation reveals that

$$SBS^{-1} = \begin{bmatrix} A_{11} & 2a_{12} & a_{12} \\ a_{21} & 2a_{22} & a_{22} \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} AD & \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} \\ 0 & 0 \end{bmatrix},$$

where $D = I \oplus [2]$. Observe that $\text{rank}(B^k) = \text{rank}(SB^kS^{-1}) = \text{rank}((SBS^{-1})^k)$. Since $(SBS^{-1})^k = \begin{bmatrix} (AD)^k & (AD)^{k-1} \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} \\ 0 & 0 \end{bmatrix}$, and $\begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix}$ is in the span of AD it follows that

$$\text{rank} \left(\begin{bmatrix} (AD)^k & (AD)^{k-1} \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} \\ 0 & 0 \end{bmatrix} \right) = \text{rank}((AD)^k).$$

By Theorem 3.2.8, we have $\text{rank}(B^k) = \text{rank}((AD)^k) = \text{rank}(A^k)$. The fact that B is totally nonnegative is trivial, and since $a_{22} > 0$ (because A is irreducible), B is irreducible. Also by the symmetry of the bordering, it follows that $\text{rank}(B) = \text{rank}(A)$ and $\text{prank}(B) = \text{prank}(A)$. Finally, (iii) now follows easily. ■

COROLLARY 3.2.10 *If the triple (n, k, p) is realizable by an irreducible totally nonnegative matrix, then the triple $(n + 1, k, p)$ is also realizable by an irreducible totally nonnegative matrix.*

Using the above results we can now classify all possible triples for $n = 5$ and 6.

EXAMPLE 3.2.11 Suppose the triple $(5, k, p)$ is realizable by a 5-by-5 irreducible totally nonnegative matrix. The triples $(5,5,5)$ and $(5,1,1)$ are certainly realizable (by Proposition 3.2.3, the triple $(5,1,1)$ is the only possible triple with a principal rank equal one realization). If the rank (k) is fixed to be 4, then, by Proposition 3.2.5, the only realizable triples are $(5,4,4)$ and $(5,4,3)$. Similarly, if the rank is fixed at 2, then by Proposition 3.2.4 the triple $(5,2,2)$ is the only realizable triple. Finally, suppose the rank is fixed to be 3. Then there are only two possible values for the principal rank: 2 or 3. Recall from Example 3.2.6 that the triples $(4,3,2)$ and $(4,3,3)$ were both realizable. Hence by Corollary 3.2.10, the triples $(5,3,2)$ and $(5,3,3)$ are both realizable.

For the case $n = 6$, the arguments are much the same as above and are omitted here. Following is a list of all the triples that are realizable by 6-by-6 irreducible totally nonnegative matrices: $(6,6,6)$; $(6,5,5)$, $(6,5,4)$, $(6,5,3)$; $(6,4,4)$, $(6,4,3)$, $(6,4,2)$; $(6,3,3)$, $(6,3,2)$; $(6,2,2)$; $(6,1,1)$. Hence all possible triples for $n = 5$ and 6 have been determined.

We now turn our attention to proving some general results on the triples:

$(n, \text{rank}, \text{prank})$.

PROPOSITION 3.2.12 *For $n \geq 1$ and $r \leq n$, the triple (n, r, r) is realizable by an irreducible n -by- n totally nonnegative matrix.*

Proof. Let A be an r -by- r totally positive matrix. Then the triple (r, r, r) is realizable by an irreducible totally nonnegative matrix (namely, A). Hence, by Corollary 3.2.10, the triple (n, r, r) is realizable, since $r \leq n$. ■

Recall that any matrix with the triple (n, r, r) , has $n - r$, 1-by-1 Jordan blocks corresponding to zero.

The asymmetric bordering notion used in Example 3.2.1 may also be used to prove the existence of a general class of triples.

PROPOSITION 3.2.13 *For $r \geq 3$ and $r < n$, the triple $(n, r, r - 1)$ is realizable by an irreducible n -by- n totally nonnegative matrix.*

Proof. We first prove the following claim: The triple $(r + 1, r, r - 1)$ is realizable for $r \geq 3$. To prove this claim let A be an r -by- r totally positive matrix partitioned as follows,

$$A = \begin{bmatrix} a_{11} & A_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

where A_{12} is $(n - 1)$ -by- $(n - 1)$. Define the $(r + 1)$ -by- $(r + 1)$ matrix A' as

$$A' = \begin{bmatrix} a_{11} & a_{11} & A_{12} \\ a_{21} & a_{21} & a_{22} \\ a_{21} & a_{21} & a_{22} \end{bmatrix}.$$

Then A' is an irreducible totally nonnegative matrix, and it is clear that $\text{rank}(A') = \text{rank}(A) = r$. What about the $\text{prank}(A')$? First observe that

$$\det A'[\{2, 3, \dots, r\}] = \det A[\{2, 3, \dots, r\}|\{1, 2, \dots, r - 1\}] > 0,$$

since A is totally positive. Thus $r \geq \text{prank}(A') \geq r - 1$. Suppose $\text{prank}(A') = r$, and let $\alpha \subseteq \{1, 2, \dots, r + 1\}$, with $|\alpha| = r$ and $\det A[\alpha] > 0$. There are two cases to consider: $1 \in \alpha$; or $1 \notin \alpha$. Suppose $1 \in \alpha$. Then $2 \notin \alpha$ since $\det A[\alpha] > 0$, so $\alpha = \{1, 3, 4, \dots, r + 1\}$. But since rows r and $r + 1$ of A' are the same, it follows that $\det A[\alpha] = 0$, which is a contradiction. Otherwise, suppose $1 \notin \alpha$. Then $\alpha = \{2, 3, \dots, r + 1\}$, and again $\det A[\alpha] = 0$, which is also a contradiction. Hence $\text{prank}(A') = r - 1$. Thus the triple $(r + 1, r, r - 1)$ is realizable. Then, by Corollary 3.2.10, the triple $(n, r, r - 1)$ is realizable for all $n > r$. This completes the proof. ■

We note that the requirement that $r \geq 3$ is necessary since if the principal rank is equal to one, then the rank is necessarily equal to one. Similarly, $r < n$ is also necessary. Recall that any matrix with the triple $(n, r, r - 1)$, has one 2-by-2 and $n - r - 1$ 1-by-1 Jordan blocks corresponding to zero. We now consider a more general result whose proof follows slightly the proof of the previous result.

THEOREM 3.2.14 *For $k \geq 0$, $r \geq k + 2$, and $n \geq r + k$, the triple $(n, r, r - k)$ is realizable by an irreducible n -by- n totally nonnegative matrix.*

Proof. We first prove that the triple $(r + k, r, r - k)$ is realizable, from which the general result will follow by Corollary 3.2.10. Let A be an r -by- r totally positive matrix. Let $A^{(k)}$ be the $(r + k)$ -by- $(r + k)$ irreducible totally nonnegative matrix obtained from A by k successive applications of the asymmetric bordering scheme used in the proof of the previous result. Then $\text{rank}(A^{(k)}) = r$. Moreover, $\det A^{(k)}[\{k + 1, \dots, r\}] = \det A[\{k + 1, \dots, r\}|\{1, 2, \dots, r - k\}] > 0$, since A is totally

positive. Hence $r \geq \text{prank}(A^{(k)}) \geq r - k$. Finally, suppose $\alpha \subseteq \{1, 2, \dots, r + k\}$ with $\det(A^{(k)}[\alpha]) > 0$. Then by the construction of $A^{(k)}$ it follows that α can contain at most one index from $\{1, 2, \dots, k + 1\}$ and at most one index from $\{r, r + 1, \dots, r + k\}$. In other words $|\alpha| \leq |\{k + 2, \dots, r - 1\}| + 2 = r - k$. Hence $\text{prank}(A^{(k)}) = r - k$. This completes the proof. ■

It is worth mentioning that while this asymmetric bordering scheme has proved useful for determining certain triples, it is not clear how this bordering scheme affects the Jordan structure of an irreducible totally nonnegative matrix.

We are now in a position to prove the next result for the case when the principal rank is fixed to be two.

THEOREM 3.2.15 *Let A be an irreducible totally nonnegative matrix with fixed principal rank equal to two. Then the size of the largest Jordan block corresponding to zero is at most two.*

Proof. As before we assume that A has a bidiagonal factorization, and an associated diagram. Let P be a shortest path from index n on the left to index n on the right. Since $\text{prank}(A) = 2$, it follows that this path P does not intersect the bottom of this diagram. Suppose P drops to level i , that is, does not use any edges of the diagram induced by the set $\{1, 2, \dots, i - 1\}$. Then (as in the case of Proposition 3.2.4) any path from any index $2, 3, \dots, i - 1$, disjoint from P , must intersect the bottom row, otherwise $\text{prank}(A) > 2$. To show that the size of the largest Jordan block is at most two we will show $\text{rank}(A^2) = 2$. To prove this it is enough to show that any path

from any of the indices $\{i, i + 1, \dots, n - 1\}$ to $\{1, 2, \dots, n\}$ must either intersect P , or the bottom row, or terminate among the indices $\{1, 2, \dots, i - 1\}$. Suppose there exists a path Q originating at some index $t \in \{i, i + 1, \dots, n - 1\}$ and terminating at $s \in \{i, i + 1, \dots, n - 1\}$ without intersecting P or the bottom row. Since Q does not intersect P it must drop below level i , as P was a shortest path. Assume $t \geq s$ (the argument for $s \leq t$ is similar). Since A is irreducible there exists a path from s to $s + 1$, but in this case such a path must intersect Q . We claim that any path from s , disjoint from P , must intersect the bottom level. To see this suppose there exists a path R from s , that does not intersect the bottom level. (see also Figure 3.3). Recall

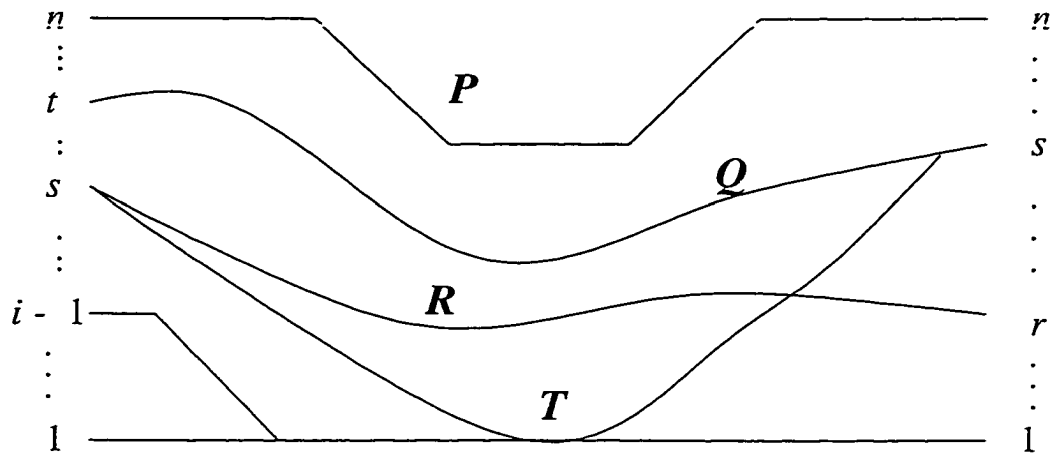


Figure 3.3: Principal Rank Two.

that any path T from s to s , disjoint from P , must intersect the bottom level (since $\text{prank}(A) = 2$), and hence any such path must intersect R . Thus there exists a path from s to s that does not intersect P and is disjoint from the bottom level: Take R until the intersection of R and T , then follow T until it intersects Q (which may be at s), and then proceed to s . This contradicts that fact that $\text{prank}(A) = 2$. Therefore

any path originating from $\{i, i + 1, \dots, n - 1\}$ must satisfy one of the following:

1. intersects P ,
2. intersects the bottom row,
3. terminates in $\{1, 2, \dots, i - 1\}$, or
4. if it terminates at $s \geq i$, then any path beginning at s , that is disjoint from P , must intersect the bottom row.

(We note that these cases are not mutually exclusive.) It now follows that the rank of A^2 is two. Certainly, the $\text{rank}(A^2) \geq 2$, as $\text{prank}(A) = 2$. Suppose there exists at least three vertex disjoint paths constituting the rank of A^2 . Since P was chosen to be a shortest such path at most one of these paths can intersect P . Moreover, since these paths are vertex disjoint at most one can terminate among the vertices $\{1, 2, \dots, i - 1\}$ (which also includes that case of a path intersecting the bottom level). Thus the only possibility left is case 4. But in this case, any path beginning from s , that is disjoint from P , must intersect the bottom level. Hence these paths cannot be disjoint for the diagram representing A^2 (which is obtained simply by concatenating two diagrams associated with A). This completes the proof. ■

COROLLARY 3.2.16 *Let A be an n -by- n irreducible totally nonnegative matrix with $\text{prank}(A) = 2$. Then $\text{rank}(A^2) = \text{prank}(A) = 2$.*

Through $n = 6$ it is not difficult to show that given a complete description of all the triples $(n, \text{rank}, \text{prank})$, and using Theorem 3.2.15, we can completely char-

acterize all possible Jordan structures (or all possible Jordan canonical forms) for every n -by- n irreducible totally nonnegative matrices with $n \leq 6$. For example, when $n = 6$ and (for example) rank fixed at four, the triples $(6,4,4)$, $(6,4,3)$, and $(6,4,2)$ are the only realizable triples in this case. Since $n - \text{rank}(A) = 2$, we know that there must exist two Jordan blocks corresponding to zero. By the remarks following Propositions 3.2.12 and 3.2.13 it follows that for $(6,4,4)$ there are two 1-by-1 Jordan blocks corresponding to zero, and for $(6,4,3)$ there is one 2-by-2 and one 1-by-1 Jordan block corresponding to zero. For the case $(6,4,2)$ there are two possible Jordan structures: (1): one 3-by-3 and one 1-by-1 Jordan block corresponding to zero, or (2): two 2-by-2 Jordan blocks corresponding to zero. By Theorem 3.2.15 it follows that case (1) cannot occur, hence any 6-by-6 irreducible totally nonnegative matrix with rank equal to 4 and principal rank equal to 2 (which do exist) must have two 2-by-2 Jordan blocks corresponding to zero.

In the following list we use JB to mean Jordan block corresponding to zero. Also in this list we do not include the cases when $\text{rank}(A) = 1$ or n . This list represents a complete classification of all possible Jordan structures through $n = 6$.

1. $n = 3$:
 - (a) $\text{rank}(A) = 2$
 - i. $\text{prank}(A) = 2 \Rightarrow$ one 1-by-1 JB,
2. $n = 4$:
 - (a) $\text{rank}(A) = 3$
 - i. $\text{prank}(A) = 3 \Rightarrow$ one 1-by-1 JB,
 - ii. $\text{prank}(A) = 2 \Rightarrow$ one 2-by-2 JB,
 - (b) $\text{rank}(A) = 2$

- i. $\text{prank}(A) = 2 \Rightarrow$ two 1-by-1 JB's,
3. $n = 5$:
- (a) $\text{rank}(A) = 4$
 - i. $\text{prank}(A) = 4 \Rightarrow$ one 1-by-1 JB,
 - ii. $\text{prank}(A) = 3 \Rightarrow$ one 2-by-2 JB,
 - (b) $\text{rank}(A) = 3$
 - i. $\text{prank}(A) = 3 \Rightarrow$ two 1-by-1 JB's,
 - ii. $\text{prank}(A) = 2 \Rightarrow$ one 2-by-2 JB and one 1-by-1 JB,
 - (c) $\text{rank}(A) = 2$
 - i. $\text{prank}(A) = 2 \Rightarrow$ three 1-by-1 JB's,
4. $n = 6$:
- (a) $\text{rank}(A) = 5$
 - i. $\text{prank}(A) = 5 \Rightarrow$ one 1-by-1 JB.
 - ii. $\text{prank}(A) = 4 \Rightarrow$ one 2-by-2 JB,
 - iii. $\text{prank}(A) = 3 \Rightarrow$ one 3-by-3 JB,
 - (b) $\text{rank}(A) = 4$
 - i. $\text{prank}(A) = 4 \Rightarrow$ two 1-by-1 JB's,
 - ii. $\text{prank}(A) = 3 \Rightarrow$ one 2-by-2 JB and one 1-by-1 JB,
 - iii. $\text{prank}(A) = 2 \Rightarrow$ two 2-by-2 JB's,
 - (c) $\text{rank}(A) = 3$
 - i. $\text{prank}(A) = 3 \Rightarrow$ three 1-by-1 JB's,
 - ii. $\text{prank}(A) = 2 \Rightarrow$ one 2-by-2 JB and two 1-by-1 JB's,
 - (d) $\text{rank}(A) = 2$
 - i. $\text{prank}(A) = 2 \Rightarrow$ four 1-by-1 JB's.

For the case $n = 7$, we can use the previous results to classify all possible triples. (We ignore the trivial triples $(7,7,7)$ and $(7,1,1)$ in this discussion.) For instance, by Proposition 3.2.5 the triples $(7,6,6)$, $(7,6,5)$ and $(7,6,4)$ are the only realizable triples when the rank is fixed at 6. All of the remaining realizable triples (which are listed below) follow from the list for $n = 6$ and Corollary 3.2.10, and also by Proposition

3.2.4. The realizable triples for $n = 7$ are: $(7,6,6)$, $(7,6,5)$, $(7,6,4)$; $(7,5,5)$, $(7,5,4)$, $(7,5,3)$; $(7,4,4)$, $(7,4,3)$, $(7,4,2)$; $(7,3,3)$, $(7,3,2)$; $(7,2,2)$. In the case when $n = 7$ using the complete list of triples above and Theorem 3.2.15, it follows that all possible Jordan structures can be characterized (see list to follow) with the exception of one case, namely the triple $(7,5,3)$. For this particular triple there are two possible Jordan structures: (1): one 3-by-3 JB and one 1-by-1 JB, or (2): two 2-by-2 JB's. The Jordan structure in case (1) is possible by considering a matrix which realizes the triple $(6,5,3)$ (which exists), and then using Corollary 3.2.9 to construct a 7-by-7 irreducible totally nonnegative matrix with the desired Jordan structure. For case (2), we do not know of a general technique to rule out or guarantee such a Jordan structure. However, after some effort we came up with the following example. Consider the following matrix,

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Then A is a 7-by-7 irreducible totally nonnegative with $\text{rank}(A) = 5$ and $\text{prank}(A) = 3$, and has Jordan structure,

$$\begin{bmatrix} * & & & & & & \\ & * & & & & & \\ & & * & & & & \\ \hline & & & 0 & 1 & & \\ & & & 0 & 0 & & \\ \hline & & & & & 0 & 1 \\ & & & & & 0 & 0 \end{bmatrix}.$$

Hence the second possible Jordan structure does indeed occur for some 7-by-7 irre-

ducible totally nonnegative matrix. The complete list of all possible Jordan structures for $n = 7$ is given below.

- $n = 7$:

1. $\text{rank}(A) = 6$

- (a) $\text{prank}(A) = 6 \Rightarrow$ one 1-by-1 JB,
- (b) $\text{prank}(A) = 5 \Rightarrow$ one 2-by-2 JB,
- (c) $\text{prank}(A) = 4 \Rightarrow$ one 3-by-3 JB,

2. $\text{rank}(A) = 5$

- (a) $\text{prank}(A) = 5 \Rightarrow$ two 1-by-1 JB's,
- (b) $\text{prank}(A) = 4 \Rightarrow$ one 2-by-2 JB and one 1-by-1 JB.
- (c) $\text{prank}(A) = 3 \Rightarrow$ one 3-by-3 JB and one 1-by-1 JB; or two 2-by-2 JB's.

3. $\text{rank}(A) = 4$

- (a) $\text{prank}(A) = 4 \Rightarrow$ three 1-by-1 JB's,
- (b) $\text{prank}(A) = 3 \Rightarrow$ one 2-by-2 JB and two 1-by-1 JB's.
- (c) $\text{prank}(A) = 2 \Rightarrow$ two 2-by-2 JB's and one 1-by-1 JB,

4. $\text{rank}(A) = 3$

- (a) $\text{prank}(A) = 3 \Rightarrow$ four 1-by-1 JB's,
- (b) $\text{prank}(A) = 2 \Rightarrow$ one 2-by-2 JB and three 1-by-1 JB's.

5. $\text{rank}(A) = 2$

- (a) $\text{prank}(A) = 2 \Rightarrow$ five 1-by-1 JB's.

We conclude this section with a discussion about future work along these lines and a couple of open problems, which we continue to work on.

Firstly, and the most important issue, is classifying all possible Jordan structures for n -by- n irreducible totally nonnegative matrices. By the results presented thus far we have completed this classification through $n = 7$. (In fact, we have now completed this classification through $n = 8$.) We are in the process of working on many new and worthwhile ideas to continue this classification.

A related, but apparently less difficult (although by no means easy), problem is determining which triples $(n, \text{rank}, \text{prank})$ are realizable by the class of n -by- n irreducible totally nonnegative matrices. Again it follows from the analysis in this section that this issue has been settled through $n = 7$. (As with the above problem we have now settled this issue through $n = 8$.) It seems that, at least thus far, a general result for all realizable triples is very possible, and we continue to develop new techniques and ideas to obtain such a general result.

There are two problems that we wish to touch upon here for a couple of reasons. Firstly, answering these questions will definitely shed some light into the previous unresolved issues, and secondly, they were both unexpected, and still (for the most part) remain unexplained. The first problem is concerned with the size of the largest Jordan block corresponding to zero for an n -by- n irreducible totally nonnegative matrix. Through $n = 7$ (and in fact $n = 8$) the size of the largest Jordan block corresponding to zero is at most the principal rank. Moreover, in general this result holds when $\text{prank}(A) \geq \lceil \frac{n}{2} \rceil$, or when $\text{prank}(A) \geq \text{rank}(A) - 1$, or also when $\text{prank}(A) = 1$ or 2 (by Theorem 3.2.15). Recall that the size of the largest Jordan block corresponding to zero is at most $\text{rank}(A) - \text{prank}(A) + 1$. Thus if $\text{prank}(A) \geq \frac{\text{rank}(A)+1}{2}$, then the result holds. At this point this question is still unresolved in general, and we do not know of a good reason for why such a result should hold. As a final note on this problem, we note that this claim on the size of the largest Jordan block is equivalent

to the following equality,

$$\text{rank}(A^{\text{prank}(A)}) = \text{prank}(A).$$

The next and final problem we discuss here is concerned with the existence of the triple, $(n, \text{rank}, \text{prank})$. Consider the triple $(6, 4, 2)$, which was shown to be a realizable triple. (Observe that the Jordan structure associated with such a triple must consist of: 2 positive distinct eigenvalues, and two, 2-by-2 Jordan blocks corresponding to zero.) Then note that the triple $(\delta, \delta - 2 + 1, \delta - 4 + 1) = (6, 5, 3)$ is also a realizable triple. Moreover, this particular rearrangement gives rise to realizable triples for every known realizable triple (compare Propositions 3.2.4 and 3.2.5). If the triple (n, k, p) can be realized, then the triple $(n, n - p + 1, n - k + 1)$ can also be realized by an n -by- n irreducible totally nonnegative matrix. Again we have little to offer about why such a result should be true, but nevertheless, it is an interesting property that these triples seem to enjoy.

Chapter 4

Determinantal Inequalities

In this chapter we investigate possible relationships among the principal minors of totally nonnegative matrices. A natural and important task is to examine all of the inequalities that exist among the principal minors of totally nonnegative matrices.

Determinantal inequalities have been studied extensively for various classes of matrices. Many of the classical inequalities of Hadamard, Fischer, and Koteljanskii (see Chapter 2) and Szasz (see [HJ1]) were first verified for the class of positive semidefinite matrices, and later were shown to also hold for totally nonnegative matrices (see Chapter 2). These inequalities have also been demonstrated for other positivity classes of matrices such as M -matrices and inverse M -matrices (see [HJ2]). More recently, the study of determinantal inequalities has seen a re emergence in modern research among linear algebraists (see [BJ2]).

An important theme developed by Barrett and Johnson [BJ2] was to describe all possible inequalities that exist among products of principal minors of positive

definite matrices via set-theoretic conditions. This natural idea seems to have escaped other experts. Employing this completely combinatorial notion allowed Barrett and Johnson to describe necessary and sufficient conditions for all such inequalities on three or fewer indices, and also verified other inequalities for cases of more indices.

In [FHJ] we utilized the idea in [BJ2] to answer similar questions for the classes of M -matrices and inverse M -matrices. In this paper ([FHJ]) we completely characterized all the inequalities that exist among products of principal minors for both of these classes.

Our intentions here are to employ these previous notions in the case of totally nonnegative matrices. One notable difference between the class of totally nonnegative matrices and the previous classes discussed above is that totally nonnegative matrices are not in general closed under simultaneous permutation of rows and columns, while all of the other classes mentioned above are closed under such an operation.

4.1 Preliminaries

Recall, that for a given n -by- n matrix A we let $A[S]$, $S \subseteq N \equiv \{1, 2, \dots, n\}$ denote the principal submatrix lying in rows and columns indexed by S . For brevity, we may let (S) denote $\det A[S]$.

DEFINITION 4.1.1 Let $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_p\}$ denote a collection of index sets, where $\alpha_i \subseteq N$, $i = 1, 2, \dots, p$. Then we define

$$\alpha(A) = \det A[\alpha_1] \det A[\alpha_2] \cdots \det A[\alpha_p].$$

If, further, $\beta = \{\beta_1, \beta_2, \dots, \beta_q\}$ is another collection of index sets with $\beta_i \subseteq N$, for all i , then we say that

$$\alpha \leq \beta \text{ with respect to TN}$$

if $\alpha(A) \leq \beta(A)$, for every n -by- n TN matrix A .

We shall also consider ratios of products of principal minors. For two given collections α and β of index sets we shall interpret $\frac{\alpha}{\beta}$ as both a numerical ratio $\frac{\alpha(A)}{\beta(A)}$ for a given TN matrix A and also as a formal ratio to be manipulated according to natural rules. When interpreted numerically, such ratios are well-defined because the class of nonsingular TN is preserved under extraction of principal submatrices, so that $\beta(A) \neq 0$ whenever A is a nonsingular TN matrix. Since, by convention, $\det A[\phi] = 1$, we also assume, without loss of generality, that in any ratio $\frac{\alpha}{\beta}$ both collections α and β have the same number of index sets, since if there is a disparity in the total number of index sets between α and β , the one with the fewer sets may be padded out with copies of ϕ . Either α or β may include repeated index sets (which count).

In chapter 2 section 3 we discussed the most famous determinantal inequalities that exist for TN matrices. Each of these inequalities may be written in our form $\alpha \leq \beta$. For example, Hadamard's inequality, $\det A \leq \prod_{i=1}^n a_{ii}$, has $\alpha = \{N, \phi, \dots, \phi\}$ and $\beta = \{\{1\}, \{2\}, \dots, \{n\}\}$, and Koteljanskii's inequality has the collections $\alpha = \{S \cup T, S \cap T\}$ and $\beta = \{S, T\}$. Our main and most general problem of interest is to characterize, via set-theoretic conditions, all pairs of collections of index sets such that

$$\frac{\alpha(A)}{\beta(A)} \leq K,$$

for some constant $K \geq 0$ (which depends on n) and for all n -by- n TN matrices A . If such a constant exists for all TN matrices A we say that the ratio $\frac{\alpha}{\beta}$ is *bounded* with respect to the class of TN matrices.

As mentioned earlier (see [FHJ]) this problem was resolved for the classes of M - and inverse M -matrices, and has received much attention for the class of positive definite matrices. We now present a couple of simple examples to accompany the ideas and definitions discussed above.

EXAMPLE 4.1.2 Let $\alpha = \{\{1, 2\}, \{3\}\}$ and $\beta = \{\{1, 3\}, \{2\}\}$. Suppose $A = [a_{ij}]$ is an arbitrary 3-by-3 TN matrix. Since A is TN, $\det A[\{1, 2\}|\{2, 3\}] = a_{12}a_{23} - a_{13}a_{22} \geq$

0, and $\det A[\{2, 3\}|\{1, 3\}] = a_{21}a_{33} - a_{23}a_{31} \geq 0$. Now consider the product

$$\begin{aligned}
 \det A[\{1, 3\}]\det A[\{2\}] &= (a_{11}a_{33} - a_{13}a_{31})a_{22}, \\
 &= a_{11}a_{33}a_{22} - a_{13}a_{31}a_{22}, \\
 &\geq a_{11}a_{33}a_{22} - a_{12}a_{23}a_{31}, \text{ since } a_{12}a_{23} \geq a_{13}a_{22}, \\
 &\geq a_{11}a_{33}a_{22} - a_{12}a_{21}a_{33}, \text{ since } a_{21}a_{33} \geq a_{23}a_{31}, \\
 &= (a_{11}a_{22} - a_{12}a_{21})a_{33} \\
 &= \det A[\{1, 2\}]\det A[\{3\}].
 \end{aligned}$$

Thus $\alpha(A) \leq \beta(A)$, for all 3-by-3 TN matrices A . Also observe that if the indices 2 and 3 are interchanged in the above collections, then the inequality $\alpha \leq \beta$ no longer holds.

Let α be any given collection of index sets. For $i \in \{1, 2, \dots, n\}$ we define $f_\alpha(i)$ to be the number of index sets in α that contain the element i . In other words $f_\alpha(i)$ counts the multiplicity (or number of occurrences) of the index i in the collection α (see also [BJ2, FHJ]). The next proposition demonstrates a simple necessary (and by no means sufficient) condition for a given ratio of principal minors to be bounded with respect to TN matrices.

PROPOSITION 4.1.3 *Let α and β be two collections of index sets. If $\frac{\alpha}{\beta}$ is bounded with respect to the class of TN matrices, then $f_\alpha(i) = f_\beta(i)$, for every $i = 1, 2, \dots, n$.*

Proof. Suppose there exists an index i for which $f_\alpha(i) > f_\beta(i)$ (if $f_\alpha(i) < f_\beta(i)$ consider the inverse of the matrix used in the argument to follow). For $k \geq 1$ let $D_k = \text{diag}(1, \dots, 1, k, 1, \dots, 1)$, where the number k occurs in the $(i, i)^{\text{th}}$ entry of D_k . Then D_k is an invertible TN matrix for every value k , and $\frac{\alpha(A)}{\beta(A)} = k^{(f_\alpha(i) - f_\beta(i))} = k^t$, where $t \geq 1$. Hence $\frac{\alpha}{\beta}$ is not a bounded ratio. ■

If a given ratio $\frac{\alpha}{\beta}$ satisfies the condition $f_\alpha(i) = f_\beta(i)$, then we say the ratio satisfies ST0 (set-theoretic) (see also [BJ2, FHJ]).

As discussed in chapter 2 the fact that a TN matrix has an elementary bidiagonal factorization proves to be a very useful fact for verifying when a ratio is bounded with respect to the class of TN matrices (see also examples to follow). Moreover, the combinatorial diagrams that are associated with each such factorization aides in this verification process, and in the identification of other possible bounded ratios. We have already seen that an arbitrary n -by- n nonsingular TN matrix can be represented by the following diagram (see Figure 4.1). Here all variables are nonnegative, and

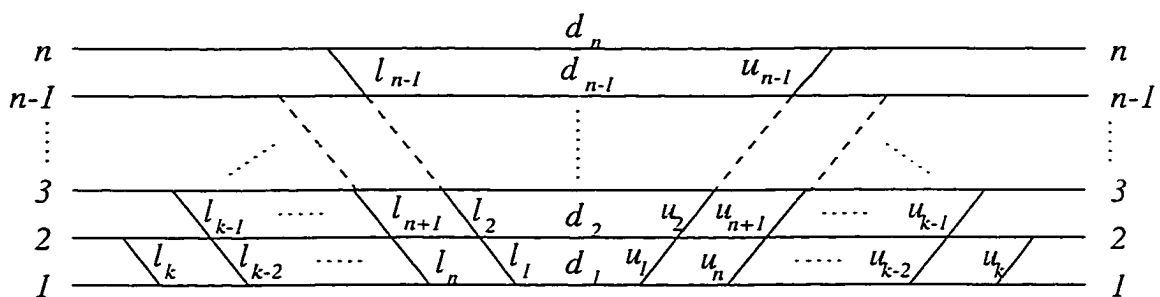


Figure 4.1: General n -by- n diagram.

since A is assumed to be nonsingular it follows that $d_i > 0$, for all $i = 1, 2, \dots, n$.

Suppose we are given a particular ratio $\frac{\alpha}{\beta}$ of index sets, and we wish to verify whether or not it is bounded with respect to the class of TN matrices. By Proposition 4.1.3, $\frac{\alpha}{\beta}$ is bounded only if $\frac{\alpha}{\beta}$ satisfies the condition ST0. From which it follows that

$$\frac{\alpha(DA)}{\beta(DA)} = \frac{\alpha(A)}{\beta(A)},$$

for every positive diagonal matrix D . Thus we may assume, by multiplication of an appropriate diagonal factor, that $d_i = 1$, for all i . Hence it is enough to test a given ratio on all nonsingular TN matrices A , for which $\det A[\{1, 2, \dots, k\}] = 1$, for $k = 1, 2, \dots, n$. Under this normalization, it follows that $\det A[S] = 1 +$ (nonnegative terms), and hence is always bounded away from zero. Similarly, in this case $\beta(A) \geq 1$. To prove a given ratio is unbounded using a bidiagonal factorization under this normalization we must show that there exist choices for $\{l_i\}$ and $\{u_i\}$ such that $\alpha(A)$ increases without bound faster than $\beta(A)$. Observe that both $\alpha(A)$ and $\beta(A)$ are subtraction free polynomial expressions in the nonnegative variables $\{l_i\}$ and $\{u_i\}$. We are interested in the total degree of some sub collection (possibly all) of the l_i 's and u_j 's in $\alpha(A)$ and $\beta(A)$. If there exists a sub collection C for which the degree with respect to this sub collection in $\alpha(A)$ exceeds the corresponding degree

in $\beta(A)$, then we may assign

$$l_i(u_j) = \begin{cases} t, & \text{if } l_i(u_j) \text{ is in } C, \\ 1, & \text{otherwise.} \end{cases}$$

In this case $\alpha(A)$ and $\beta(A)$ are polynomials in the single variable t with $t \geq 0$, and $\deg(\alpha(A)) > \deg(\beta(A))$. Letting $t \rightarrow \infty$ implies that the ratio $\frac{\alpha}{\beta}$ is not a bounded ratio. We illustrate these ideas with the following example.

EXAMPLE 4.1.4 Suppose $\alpha = \{\{1, 3\}, \{2\}\}$ and $\beta = \{\{1, 2\}, \{3\}\}$ (note that this is the permuted version of the collections in the previous example). We claim that the ratio $\frac{\alpha}{\beta}$ is an unbounded ratio. To prove this claim we make use of the ideas discussed above. Let A be an arbitrary nonsingular TN matrix, and assume A is written in the following form

$$A = \begin{bmatrix} 1 & & \\ l_3 & 1 & \\ & & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & & \\ & 1 & \\ & l_2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & & \\ l_1 & 1 & \\ & & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & u_1 & \\ & 1 & \\ & & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & & u_2 \\ & 1 & \\ & & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & u_3 & \\ & 1 & \\ & & 1 \end{bmatrix}.$$

In this case we may also represent A via the following diagram (see Figure 4.2). A

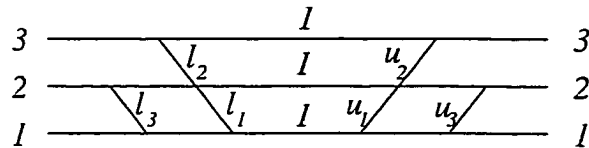


Figure 4.2: 3-by-3 diagram.

simple computation reveals that $(1, 3) = 1 + l_2 u_2$, and $(2) = 1 + (l_1 + l_3)(u_1 + u_3)$, in which case $\alpha(A) = (1 + l_2 u_2)(1 + (l_1 + l_3)(u_1 + u_3))$. Similarly, $\beta(A) = (1)(1 + l_2 u_2 + l_1 l_2 u_1 u_2)$. Note that neither l_3 or u_3 appear in the expression corresponding to $\beta(A)$, while they each appear once in the expression $\alpha(A)$. So if we let $l_3 = t$ (for example) and set all other variables equal to 1, then the matrix A is equal to

$$A = \begin{bmatrix} 1 & 2 & 1 \\ t+1 & 2t+3 & t+2 \\ 1 & 3 & 3 \end{bmatrix},$$

and in this case $\alpha(A) = 4t + 6$, and $\beta(A) = 3$. Consequently, $\frac{\alpha(A)}{\beta(A)} \approx \frac{4t+6}{3}$, which increases without bound as $t \rightarrow \infty$.

It is also possible to use the bidiagonal factorization of a TN matrix to prove that a ratio is bounded. Consider the following example.

EXAMPLE 4.1.5 Let $\alpha = \{\{1, 2\}, \{3\}\}$ and $\beta = \{\{1, 3\}, \{2\}\}$. Note that we have already proven that this ratio is bounded without using these new ideas. Recall from the previous example that $\alpha(A) = (1)(1 + l_2 u_2 + l_1 l_2 u_1 u_2)$, and $\beta(A) = (1 + l_2 u_2)(1 + (l_1 + l_3)(u_1 + u_3))$. Observe that $\beta(A) - \alpha(A) = l_2 u_2 (l_3 u_3 + l_3 u_1 + l_1 u_3)$, which is a subtraction free expression in the nonnegative variables $\{l_i\}$ and $\{u_j\}$ and, in particular, is always nonnegative. Hence $\frac{\alpha(A)}{\beta(A)} \leq 1$, and therefore the given ratio is bounded with respect to the class TN (in fact, $\alpha(A) \leq \beta(A)$).

We call a ratio $\frac{\alpha}{\beta}$ a *basic n -ratio* if the collection α contains all of the subsets of $\{1, 2, \dots, n\}$ with cardinality that is the same parity as n and β consists of all the

remaining subsets of $\{1, 2, \dots, n\}$. For example, the basic 2-ratio is

$$\frac{(1, 2)(\phi)}{(1)(2)}, \text{ (Koteljanskii)}$$

and the basic 3-ratio is

$$\frac{(1, 2, 3)(1)(2)(3)}{(1, 2)(1, 3)(2, 3)(\phi)},$$

and so-on. The class of basic n -ratios, first defined in [BJ2], were essential in [FHJ] as they turned out to be the “generators” of all of the bounded ratios for M - and inverse M -matrices. Their role here in the study of bounded ratios with respect to the class of TN matrices is short lived, and in this case they are not so vital, but we include them as an example class for completeness. We note here that the basic 1- and 2-ratios are bounded with respect to the class TN.

PROPOSITION 4.1.6 *For $n \geq 3$, the basic n -ratio is unbounded with respect to the class TN.*

Proof. Suppose $\frac{\alpha}{\beta}$ denotes the basic n -ratio and let $S = \{1, 2, \dots, n - 2\}$. We assume for simplicity that n is even, say $n = 2k$. The argument when n is odd is very similar. The idea is to count the number of sets in α which do not contain S and are

not contained in S . A routine counting argument shows that this number is equal to,

$$\sum_{j=1}^{k-2} \left[\binom{n}{2j} - \binom{n-2}{2j} \right] + \binom{n}{n-2} - 1, \quad (4.1)$$

and the same count for the collection β is

$$\sum_{j=1}^{k-1} \left[\binom{n}{2j-1} - \binom{n-2}{2j-1} \right] + \binom{n}{n-1} - 2. \quad (4.2)$$

Thus the difference (4.1)-(4.2) can be written as

$$\left[\sum_{j=1}^{n-1} (-1)^j \binom{n}{j} - \sum_{j=1}^{n-3} (-1)^j \binom{n-2}{j} \right] + 1.$$

Recall the well-known elementary combinatorial identity, $\sum_{j=0}^M (-1)^j \binom{M}{j} = 0$, from which it follows the the difference (4.1)-(4.2) is also equal to $(-1 - (-1)^n) - (-1 - (-1)^{n-2}) + 1 = 1$. Thus there are more such sets in α (exactly one more) than in β . Suppose $J \subseteq N$ which does not contain S and is not contained in S . Consider the general diagram given by Figure 4.1. Using the rules for calculating the minor (J) via the diagram (i.e., counting independent paths originating and terminating

in J) it follows that the variable u_{n-2} appears in the expression given by (J) with positive degree. This can be most easily seen by observing that such a set J must contain index $n - 1$ or n , and must be missing at least one of the indices less than $n - 1$. Moreover, u_{n-2} only appears with positive degree in the principal minors (J) for which J is not contained in and does not contain S . Thus if we assign all variables except u_{n-2} the value 1, and let $u_{n-2} = t$ it follows that the degree of t in $\alpha(A)$ is greater than the degree of t in β . Consequently, the basic n -ratio is not bounded with respect to the class TN. ■

Note that in the proof above the key ingredient was the fact the number of sets in α that do not contain S and are not contained in S was greater than the corresponding number in β . This notion gives the following result.

PROPOSITION 4.1.7 *Let $\frac{\alpha}{\beta}$ be a given ratio. If there exists more sets in α that are not contained in $\{1, 2, \dots, k\}$, and do not contain $\{1, 2, \dots, k\}$ (for some $1 \leq k \leq n$), then in the collection β , then the ratio $\frac{\alpha}{\beta}$ is not bounded with respect to the class TN.*

We finish this preliminary discussion with the following definition. Let $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_p\}$ be a given collection of index sets. We let $f_\alpha(J)$ denote the number of index sets in α that contain J , where $J \subseteq N$. In [BJ2] it is shown that if a ratio $\frac{\alpha}{\beta}$ is bounded with respect to the class of positive definite matrices, then $f_\alpha(J) \geq f_\beta(J)$, for every subset $J \subseteq N$. Barrett and Johnson also showed that this condition is not sufficient for a ratio to be bounded with respect to the class of positive definite matrices. In [FHJ] part of one of the main results can be stated as follows: A ratio $\frac{\alpha}{\beta}$ is bounded with respect to the class of M -matrices if and only if it satisfies:

1. $f_\alpha(\{i\}) = f_\beta(\{i\})$, for all $i = 1, 2, \dots, n$, and
2. $f_\alpha(J) \geq f_\beta(J)$, for every subset $J \subseteq N$.

We have already seen that the condition $f_\alpha(J) \geq f_\beta(J)$ is neither necessary (see Example 4.1.2) nor sufficient (the basic 4-ratio for example) for a ratio to be bounded with respect to the class of TN matrices. However, we can prove the following corresponding result. Recall that the dispersion of a given set $S = \{i_1, i_2, \dots, i_k\}$, where $i_j < i_{j+1}$ ($j = 1, 2, \dots, k - 1$) is given by $d(S) = i_k - i_1 - (k - 1)$.

PROPOSITION 4.1.8 *Let $\frac{\alpha}{\beta}$ be a given ratio. If $\frac{\alpha}{\beta}$ is bounded with respect to the class of TN matrices, then $f_\alpha(J) \geq f_\beta(J)$, for all $J \subseteq N$, with $d(J) = 0$, that is for all contiguous subsets J of N .*

Proof. Suppose there exists a set $J \subseteq N$, with $d(J) = 0$ and $f_\alpha(J) < f_\beta(J)$. By Proposition 4.1.3 we may assume that $|J| \geq 2$. We consider two cases.

Case (1): $J = N = \{1, 2, \dots, n\}$.

Consider the diagram in Figure 4.1, which represents an arbitrary nonsingular TN matrix. Observe that the variable u_{n-1} appears with positive degree in a minor (S) if and only if $n \in S$, and $S \neq N$. Hence if we set all weights to 1 except u_{n-1} which we set to t , then the degree of t in $\alpha(A)$ is equal to $f_\alpha(\{n\}) - f_\alpha(N) = f_\beta(\{n\}) - f_\alpha(N)$. Since $f_\alpha(N) < f_\beta(N)$ it follows that the degree of t in $\alpha(A)$ is larger than the degree of t in $\beta(A)$. Consequently, the ratio $\frac{\alpha}{\beta}$ is not bounded with respect to the class of TN matrices.

Case (2): $J \neq N$.

In this case we set all weights of the edges not contained among the indices in J to zero. With this assignment it follows that any minor (S) is equal to $(S \cap J)$. Thus if $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_p\}$, then define $\tilde{\alpha} = \{\alpha_1 \cap J, \alpha_2 \cap J, \dots, \alpha_p \cap J\}$ and define $\tilde{\beta}$ similarly. By the previous remarks (using such a choice of weights) $\frac{\alpha(A)}{\beta(A)} = \frac{\tilde{\alpha}(A)}{\tilde{\beta}(A)}$. But for the new ratio $\frac{\tilde{\alpha}}{\tilde{\beta}}$ we have $f_{\tilde{\alpha}}(J) < f_{\tilde{\beta}}(J)$, where J now represents all possible indices. Thus by the previous case $\frac{\tilde{\alpha}}{\tilde{\beta}}$ is an unbounded ratio, and hence $\frac{\alpha}{\beta}$ is unbounded with respect to the class of TN matrices. This completes the proof. ■

4.2 Operators Preserving Bounded Ratios

In the interest of describing all the inequalities that exist among products of principal minors of totally nonnegative matrices we identify various operations that may be applied to ratios and preserve the property of being bounded with respect to the totally nonnegative matrices. Understanding of how a bounded ratio may be manipulated is vital to our goal of characterizing all such determinantal inequalities.

Firstly we define each operation and then prove that each such operation preserves bounded ratios with respect to the totally nonnegative matrices.

DEFINITION 4.2.1 Let $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_p\}$ and $\beta = \{\beta_1, \beta_2, \dots, \beta_q\}$ be any two collections of index sets from $\{1, 2, \dots, n\}$. Then

(i) **Complement:** $C(\alpha/\beta) = \frac{(\alpha_1^c), (\alpha_2^c), \dots, (\alpha_p^c)}{(\beta_1^c), (\beta_2^c), \dots, (\beta_q^c)}$, where

$$J^c = \{1, 2, \dots, n\} \setminus J, \text{ for } J \subseteq \{1, 2, \dots, n\}.$$

- (ii) **Reversal:** $R(\alpha/\beta)$; every index j in each index set of α and β is replaced (or reversed) by (to) $n - j + 1$. Example: $R\left(\frac{\{1,2\}\{3\}}{\{1,3\}\{2\}}\right) = \frac{\{2,3\}\{1\}}{\{1,3\}\{2\}}$.
- (iii) **Shift:** $S_i(\alpha/\beta)$. Suppose index i does not appear among any of the index sets in either α or β . Then $S_i(\alpha/\beta)$ shifts every index, in each index set, greater than i down by 1. Example: $S_2\left(\frac{\{1,3\}\{4\}}{\{1,4\}\{3\}}\right) = \frac{\{1,2\}\{3\}}{\{1,3\}\{2\}}$.
- (iv) **Insertion:** $A_i(\alpha/\beta)$. Suppose index i does not appear among any of the index sets in either α or β . Then $A_i(\alpha/\beta) = \frac{(\alpha_1 \cup \{i\}), (\alpha_2 \cup \{i\}), \dots, (\alpha_p \cup \{i\})}{(\beta_1 \cup \{i\}), (\beta_2 \cup \{i\}), \dots, (\beta_q \cup \{i\})}$.
- (v) **Deletion:** $D_i(\alpha/\beta)$. Suppose that index i appears in every index set in both α and β . Then $D_i(\alpha/\beta) = \frac{(\alpha_1 \setminus \{i\}), (\alpha_2 \setminus \{i\}), \dots, (\alpha_p \setminus \{i\})}{(\beta_1 \setminus \{i\}), (\beta_2 \setminus \{i\}), \dots, (\beta_q \setminus \{i\})}$.

We begin our analysis with the complement operator. First we recall two well-known facts. If A is an n -by- n invertible matrix, then $\det A^{-1}[J] = \frac{\det A[J^c]}{\det A}$ (Jacobi's identity (1.1)) for every index set J , and if A is an invertible TN matrix, then $SA^{-1}S$ is TN, for $S = \text{diag}(1, -1, \dots, \pm 1)$. Also we note here that $\det SAS[J] = \det A[J]$, for every index set J .

PROPOSITION 4.2.2 *Suppose $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_p\}$ and $\beta = \{\beta_1, \beta_2, \dots, \beta_p\}$ are two collections of index sets. Then α/β is a bounded ratio with respect to the totally nonnegative matrices if and only if $C(\alpha/\beta)$ is bounded with respect to the totally nonnegative matrices.*

Proof. Suppose α/β is bounded. We let α^c denote the collection $\{\alpha_1^c, \alpha_2^c, \dots, \alpha_p^c\}$

(and define β^c similarly). Let A be any invertible totally nonnegative matrix. Then

$$\begin{aligned}
\frac{\alpha^c(A)}{\beta^c(A)} &= \frac{\det A[\alpha_1^c] \cdots \det A[\alpha_p^c]}{\det A[\beta_1^c] \cdots \det A[\beta_p^c]} \\
&= \frac{\det A^{-1}[\alpha_1] \cdots \det A^{-1}[\alpha_p] (\det A)^p}{\det A^{-1}[\beta_1] \cdots \det A^{-1}[\beta_p] (\det A)^p} \\
&= \frac{\det(SA^{-1}S)[\alpha_1] \cdots \det(SA^{-1}S)[\alpha_p]}{\det(SA^{-1}S)[\beta_1] \cdots \det(SA^{-1}S)[\beta_p]} \\
&= \frac{\alpha(SA^{-1}S)}{\beta(SA^{-1}S)},
\end{aligned}$$

which is bounded as $SA^{-1}S$ is totally nonnegative. For the converse, observe that by definition $C^2(\alpha/\beta) = C(C(\alpha/\beta)) = \alpha/\beta$, hence it follows that α/β is a bounded ratio with respect to the totally nonnegative matrices if and only if $C(\alpha/\beta)$ is bounded with respect to the totally nonnegative matrices. ■

Recall that if A is totally nonnegative then $\rho A \rho$ is totally nonnegative (see Proposition 2.4.6), where ρ is the permutation matrix induced by the permutation $i \rightarrow n - i + 1$, for each i . In other words if $A = [a_{ij}]$ is TN, then the matrix $\rho A \rho = [a_{n-i+1, n-j+1}]$ is TN. These facts prove the following result.

PROPOSITION 4.2.3 *Suppose $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_p\}$ and $\beta = \{\beta_1, \beta_2, \dots, \beta_p\}$ are two collections of index sets. Then α/β is a bounded ratio with respect to the totally nonnegative matrices if and only if $R(\alpha/\beta)$ is bounded with respect to the totally nonnegative matrices.*

We also note here that the reversal operator satisfies the relation $R^2(\alpha/\beta) = \alpha/\beta$.

The next operation of interest is shifting. We first consider an example. Suppose we are given the ratio $\frac{(1,3,4)(3)}{(1,3)(3,4)}$. Then $S_2\left(\frac{(1,3,4)(3)}{(1,3)(3,4)}\right) = \frac{(1,2,3)(2)}{(1,2)(2,3)}$. For simplicity, we also define a shift of an index set by the operator s_i . That is, if J is an index set for which the index i does not appear, then we define $s_i(J)$ to be the index set where every index in J greater than i is shifted down by one. For example, $s_3(\{1, 2, 4, 5, 6\}) = \{1, 2, 3, 4, 5\}$. If $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_p\}$ denotes a collection of index sets for which i is not a member of any index set, then $S_i(\alpha) = \{s_i(\alpha_1), \dots, s_i(\alpha_p)\}$.

PROPOSITION 4.2.4 *Suppose $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_p\}$ and $\beta = \{\beta_1, \beta_2, \dots, \beta_p\}$ are two collections of index sets for which index i is not a member of any index set in either collection. Then α/β is a bounded ratio with respect to the totally nonnegative matrices if and only if $S_i(\alpha/\beta)$ is bounded with respect to the totally nonnegative matrices.*

Proof. Suppose α/β is not bounded with respect to the totally nonnegative matrices. Then there exists a sequence of totally nonnegative matrices $\{A_n\}$ such that $\{\frac{\alpha(A_n)}{\beta(A_n)}\}$ is an unbounded set. Define B_n to be $B_n = A_n(\{i\})$. Then B_n is TN, and

$$\alpha(A_n) = \prod_{j=1}^p \det A_n[\alpha_j] = \prod_{j=1}^p \det B_n[s_i(\alpha_j)] = S_i(\alpha)(B_n).$$

The second equality above follows from the fact that the index i does not appear in the collection α . Thus $S_i(\alpha/\beta)$ is not bounded. On the other hand suppose $S_i(\alpha/\beta)$ is not bounded. Then there exists a sequence of totally nonnegative matrices $\{A_n\}$ such that $\{\frac{S_i(\alpha)(A_n)}{S_i(\beta)(A_n)}\}$ is an unbounded set. Embed A_n (by inserting a row and a column) into

a n -by- n TN matrix B_n so that $B_n(\{i\}) = A_n$. Then $\det A_n[s_i(\beta_j)] = \det B_n[\beta_j] > 0$ for each j , and hence it follows that there exists a sequence of TN matrices $\{B_n\}$ for which $\{\frac{\alpha(B_n)}{\beta(B_n)}\}$ is an unbounded set. ■

Before we come to the operations of insertion and deletion we present the following example illustrating that (in general) to preserve bounded ratios it is necessary to insert (and similarly delete) an index into (from) every set.

EXAMPLE 4.2.5 Consider the ratio $\frac{(2,3,4)(3)}{(2,3)(3,4)}$. Then this ratio is bounded by one with respect to the totally nonnegative matrices since it is a Koteljanskii type ratio. However, the ratio $\frac{(2,3,4)(1,3)}{(1,2,3)(3,4)}$, which satisfies (ST0), is not bounded with respect to the totally nonnegative matrices since it fails the condition in Proposition 4.1.8 for the set $J = \{1, 2, 3\}$. Of course, in this case, if we insert the index one into every set of the ratio $\frac{(2,3,4)(3)}{(2,3)(3,4)}$ we obtain another Koteljanskii type ratio, and hence it is bounded by one.

For our analysis we need the following key lemma which may also be of independent interest. We call an n -by- n matrix S a *signature matrix* if S is a diagonal matrix whose main diagonal entries are either 1 or -1. Observe that if S is a signature matrix, then $S = S^T = S^{-1}$. An n -by- n matrix A is signature similar to an n -by- n matrix B if there exists an n -by- n signature matrix S such that $A = SBS$.

LEMMA 4.2.6 *Let A be an invertible n -by- n totally nonnegative matrix. Then $A/A[\alpha]$, the Schur-complement of $A[\alpha]$ in A , is signature similar to a totally nonnegative matrix.*

Proof. Recall the well-known fact that $A/A[\alpha] = (A^{-1}[\alpha^c])^{-1}$ (see [HJ1]). Since A is TN, we have that $A^{-1} = SBS$, for $S = \text{diag}(1, -1, \dots, \pm 1)$ and B totally nonnegative. Then $A/A[\alpha] = (SBS[\alpha^c])^{-1} = (S[\alpha^c]B[\alpha^c]S[\alpha^c])^{-1} = S[\alpha^c](B[\alpha^c])^{-1}S[\alpha^c]$. Since B is TN, $B[\alpha^c]$ is TN, and hence $(B[\alpha^c])^{-1}$ is signature similar to a totally nonnegative matrix C . that is $(B[\alpha^c])^{-1} = S'CS'$, where S' is a signature matrix. Therefore

$$A/A[\alpha] = S[\alpha^c]S'CS'S[\alpha^c] = S''CS'',$$

where $S'' = S[\alpha^c]S'$ is a signature matrix. This completes the proof. ■

Recall that if $d(\alpha^c) = 0$, then $A/A[\alpha]$ is a totally nonnegative matrix (see Propositions 2.1.3 and 2.1.4).

We are now in a position to prove that insertion preserves bounded ratios.

PROPOSITION 4.2.7 *Suppose $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_p\}$ and $\beta = \{\beta_1, \beta_2, \dots, \beta_q\}$ are two collections of index sets such that the index i is not a member of any index set in either α or β . If α/β is a bounded ratio with respect to the totally nonnegative matrices, then $A_i(\alpha/\beta)$ is bounded with respect to the totally nonnegative matrices.*

Proof. Define $\alpha' = \{\alpha_1 \cup \{i\}, \alpha_2 \cup \{i\}, \dots, \alpha_p \cup \{i\}\}$ and $\beta' = \{\beta_1 \cup \{i\}, \beta_2 \cup \{i\}, \dots, \beta_q \cup \{i\}\}$. It is well-known that $\det A[\gamma \cup \{i\}] = a_{ii} \det(A/a_{ii})[s_i(\gamma)]$, for every index set γ that does not contain i , whenever $a_{ii} > 0$. Therefore if A is an arbitrary

nonsingular TN matrix it follows that

$$\begin{aligned}
\alpha'(A) &= \prod_{j=1}^p \det A[\alpha_j \cup \{i\}] \\
&= (a_{ii})^p \prod_{j=1}^p \det(A/a_{ii})[s_i(\alpha_j)] \\
&= (a_{ii})^p S_i(\alpha)(A/a_{ii}) \\
&= (a_{ii})^p S_i(\alpha)(SBS), \text{ by Lemma 4.2.6} \\
&= (a_{ii})^p S_i(\alpha)(B) \\
&\leq K \cdot (a_{ii})^p S_i(\beta)(B), \text{ by Lemma 4.2.4} \\
&= K \cdot (a_{ii})^p S_i(\beta)(SBS) \\
&= K \cdot (a_{ii})^p S_i(\beta)(A/a_{ii}), \text{ by Lemma 4.2.6} \\
&= K \cdot (a_{ii})^p \prod_{j=1}^p \det(A/a_{ii})[s_i(\beta_j)] \\
&= K \cdot \prod_{j=1}^p \det A[\beta_j \cup \{i\}] \\
&= K \cdot \beta'(A),
\end{aligned}$$

where K is a constant. Thus α'/β' is a bounded ratio. ■

Before we consider the deletion operator we first demonstrate a very useful and elementary relationship between the operators A_i and D_i .

LEMMA 4.2.8 *Let $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_p\}$ and $\beta = \{\beta_1, \beta_2, \dots, \beta_p\}$ be two collections of index sets such that the index i is not a member of any index set in either α or β .*

Then

$$C(A_i(\alpha/\beta)) = D_i(C(\alpha/\beta)).$$

Proof. First observe that all of the above operations are well-defined. Since each of these operations is applied set-wise (or set by set) it is enough to consider the case when both α and β contain only a single set. Consider the left-hand side:

$$C(A_i(\alpha/\beta)) = C\left(\frac{\alpha \cup \{i\}}{\beta \cup \{i\}}\right) = \frac{\alpha^c \cap \{i\}^c}{\beta^c \cap \{i\}^c},$$

and the right-hand side is equal to

$$D_i(C(\alpha/\beta)) = D_i\left(\frac{\alpha^c}{\beta^c}\right) = \frac{\alpha^c \cap \{i\}^c}{\beta^c \cap \{i\}^c}.$$

Thus the left-hand side equals the right-hand side, which completes the proof. ■

Some obvious consequences of the above lemma are that $A_i = CD_iC$, and that $D_i = CA_iC$, whenever the operators are defined. A less immediate consequence of the above lemma is the fact that D_i preserves bounded ratios.

PROPOSITION 4.2.9 *Suppose $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_p\}$ and $\beta = \{\beta_1, \beta_2, \dots, \beta_p\}$ are*

two collections of index sets such that the index i is a member of every index set in α and β . If α/β is a bounded ratio with respect to the totally nonnegative matrices, then $D_i(\alpha/\beta)$ is bounded with respect to the totally nonnegative matrices.

Proof. Since α/β is a bounded ratio with respect to the totally nonnegative matrices, it follows from Propositions 4.2.2 and 4.2.7 that $(CA_iC)(\alpha/\beta)$ is a bounded ratio. But by the remarks above $D_i(\alpha/\beta) = (CA_iC)(\alpha/\beta)$, and hence $D_i(\alpha/\beta)$ is a bounded ratio. ■

Propositions 4.2.2, 4.2.3, 4.2.4, 4.2.7, and 4.2.9 show that the following are equivalent:

- (i) α/β is bounded with respect to the TN matrices,
- (ii) $C(\alpha/\beta)$ is bounded with respect to the TN matrices,
- (iii) $R(\alpha/\beta)$ is bounded with respect to the TN matrices,
- (iv) $S_i(\alpha/\beta)$ is bounded with respect to the TN matrices,
- (v) $A_i(\alpha/\beta)$ is bounded with respect to the TN matrices,
- (vi) $D_i(\alpha/\beta)$ is bounded with respect to the TN matrices.

As a final remark in this section we note that if a ratio α/β is bounded by one, then it follows that applying any of the above five operations preserves the inequality between the products of principal minors.

4.3 Special Ratio Types

In this section we demonstrate various types of bounded ratios with respect to the totally nonnegative matrices. Our aim in this section is to use these results as a foundation and hopefully build on them in an attempt to describe all bounded ratios for specific values of n . We begin with the following definition.

DEFINITION 4.3.1 Let $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_p\}$ and $\beta = \{\beta_1, \beta_2, \dots, \beta_p\}$ be two partitions of $\{1, 2, \dots, n\}$. Then the ratio α/β is called an *order-one ratio*.

We have already discussed many examples of bounded (in fact bounded by one) order-one ratios, namely the classical ratios of Hadamard and Fischer. However, there are many other types of order-one ratios that are bounded with respect to the totally nonnegative matrices. Consider the following result.

THEOREM 4.3.2 Let α be a partition of $\{1, 2, \dots, n\}$ consisting of the pair $\{p, q\}$ ($p < q$) and the remaining singletons, and let β be another partition of $\{1, 2, \dots, n\}$ consisting of the pair $\{s, t\}$ ($s < t$) and the remaining singletons. Then the order-one ratio α/β is bounded if and only if $s \leq p < q \leq t$.

Proof. We first verify that the condition $s \leq p < q \leq t$ is necessary. Suppose that this condition fails for the collections α and β . This gives rise to five possible cases: (1): $p < s \leq q \leq t$; (2): $p < s < t < q$; (3): $s \leq p \leq t < q$; (4): $s < t \leq p < q$; (5): $p < q \leq s < t$. Before we investigate each case we note here that the ratio α/β simplifies when it is evaluated at a given matrix, namely $\frac{\alpha(A)}{\beta(A)} = \frac{(p,q)(s,t)}{(s,t)(p)(q)}$. Since

totally nonnegative matrices are closed under extraction of principal submatrices it suffices to assume that $\{p, q, s, t\} = \{1, 2, 3, 4\}$ (in other words we are shifting the indices in this ratio). Consider case (1), namely $p < s \leq q \leq t$. If $s = q$ or $q = t$, then the ratio α/β does not satisfy the condition given in Proposition 4.1.8 for the contiguous subset $\{2, 3\}$. Thus we may assume that $p < s < q < t$. In this case $\frac{\alpha(A)}{\beta(A)} = \frac{(1,3)(2)(4)}{(2,4)(1)(3)}$. Consider the following 4-by-4 totally nonnegative matrix.

$$A = \begin{bmatrix} 1 & 3 & 3 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 1+t & 3+2t & 2+t \\ 0 & t & 1+2t & 2+t \end{bmatrix},$$

with $t \geq 0$. Then it can be easily verified that

$$\frac{\alpha(A)}{\beta(A)} = \frac{(3+2t)(1)(2+t)}{(2)(1)(3+2t)} = \frac{2+t}{2}.$$

Hence this ratio is not bounded with respect to the totally nonnegative matrices. For cases (2) and (3) it is not difficult to check that the ratio α/β does not satisfy the condition given in Proposition 4.1.8, since in both cases s and t will be consecutive integers. The remaining two cases follow from cases (1) and (2) by reversing the indices. This proves the necessity of the condition $s \leq p < q \leq t$. To prove sufficiency we consider two cases: (1): $s = p$ (there is a similar proof for the case when $q = t$); and (2): $s < p < q < t$. For case (1), we have $s = p = 1$, $q = 2$ and $t = 3$. But in this case the ratio $\alpha/\beta = \frac{(1,2)(3)}{(1,3)(2)}$, which has already been shown to be bounded by one in Example 4.1.2. For case (2), the ratio $\alpha/\beta = \frac{(2,3)(1)(4)}{(1,4)(2)(3)}$. Since this ratio satisfies

(ST0) it follows that multiplication via a positive diagonal matrix does change the value of the ratio, hence we may assume that the totally nonnegative matrix A has ones on the main diagonal. Therefore the inequality $\alpha(A) \leq \beta(A)$ is equivalent to $a_{23}a_{32} \geq a_{14}a_{41}$. Observe that $a_{14} \leq a_{13}a_{34}$, and $a_{41} \leq a_{31}a_{43}$. Hence $a_{14}a_{41} \leq a_{13}a_{34}a_{31}a_{43} \leq a_{12}a_{23}a_{21}a_{32}a_{34}a_{43}$, as $a_{13} \leq a_{12}a_{23}$ and $a_{31} \leq a_{32}a_{21}$. Since $a_{12}a_{21}$ and $a_{34}a_{43}$ are both less than one, the result follows. This completes the proof. ■

The above result may be used to verify many other bounded order-one ratios.

COROLLARY 4.3.3 *For $k \geq 2$, $\frac{(1,2,\dots,k)(k+1)}{(1,2,\dots,k-1,k+1)(k)} \leq 1$ with respect to the totally nonnegative matrices.*

Proof. The proof is by induction on k . The case $k = 2$ is verified in Example 4.1.2. Assume the ratio is bounded for all values less than $k + 1$. Let $\alpha(A) = (1, 2, \dots, k)(k + 1)$ and $\beta(A) = (1, 2, \dots, k - 1, k + 1)(k)$. Again we assume that A has all main diagonal entries equal to one. Consider the Schur-complement of A with respect to the the (1,1) entry, A/a_{11} , which is totally nonnegative. Then $\alpha(A) = a_{11} \det(A/a_{11})[\{1, 2, \dots, k - 1\}]a_{k+1,k+1} = \det(A/a_{11})[\{1, 2, \dots, k - 1\}]$, and similarly, $\beta(A) = \det(A/a_{11})[\{1, 2, \dots, k - 2, k\}]$. We make the following claim that $\frac{(A/a_{11})[\{k\}]}{(A/a_{11})[\{k-1\}]} \geq 1$. Observe that this inequality is equivalent to $1 - a_{1k}a_{k1} \geq 1 - a_{k-1,1}a_{1,k-1}$, or $(1, k)(k - 1) \geq (1, k - 1)(k)$, since the main diagonal entries all equal to 1. However, the inequality $(1, k)(k - 1) \geq (1, k - 1)(k)$ follows from Theorem 4.3.2, which proves the claim. Thus

$$\begin{aligned}
\frac{\alpha(A)}{\beta(A)} &= \frac{\det(A/a_{11})[\{1, 2, \dots, k-1\}]}{\det(A/a_{11})[\{1, 2, \dots, k-2, k\}]} \\
&\leq \frac{\det(A/a_{11})[\{1, 2, \dots, k-1\}](A/a_{11})[\{k\}]}{\det(A/a_{11})[\{1, 2, \dots, k-2, k\}](A/a_{11})[\{k-1\}]} \\
&\leq 1,
\end{aligned}$$

by Theorem 4.3.2 since A/a_{11} is totally nonnegative. ■

Applying the reversal operator to the previous inequality yields the inequality

$$\frac{(2, \dots, k+1)(1)}{(1, 3, \dots, k+1)(2)} \leq 1,$$

for $k \geq 2$, with respect to the class of totally nonnegative matrices.

COROLLARY 4.3.4 For $k \geq 3$, $\frac{(1)(2, \dots, k-1)(k)}{(2)(1, 3, \dots, k-2, k)(k-1)} \leq 1$ with respect to the totally nonnegative matrices.

Proof. Applying Theorem 4.3.3 and the corresponding reversed inequality gives

$$(1)(2, \dots, k-1)(k) \leq (1)(2, \dots, k-2, k)(k-1) \leq (2)(1, 3, \dots, k-2, k)(k-1). \blacksquare$$

We now present a general result concerning a special class of order-one ratios that have at most one set of cardinality greater than one.

THEOREM 4.3.5 *The following inequality holds, for any nonsingular totally non-negative matrix*

$$\frac{(i_1, i_2, \dots, i_k) \prod_{j \neq i_t} (j)}{(j_1, j_2, \dots, j_k) \prod_{i \neq j_t} (i)} \leq 1,$$

if and only if $j_1 \leq i_1 < i_k \leq j_k$ and $i_s = j_s$, for $s = 2, 3, \dots, k - 1$.

Proof. The argument to verify that $j_1 \leq i_1 < i_k \leq j_k$ is necessary is similar to the argument presented in the proof of Theorem 4.3.2, and it omitted here. To show that the condition $i_s = j_s$, for $s = 2, 3, \dots, k - 1$ is necessary, suppose s_0 is the smallest $s \in \{2, 3, \dots, k - 1\}$ such that $i_s \neq j_s$. Then there are two cases to consider: (1): $i_{s_0} < j_{s_0}$; or (2): $i_{s_0} > j_{s_0}$. Suppose $i_{s_0} < j_{s_0}$. In what follows we consider the general diagram corresponding to an n -by- n totally nonnegative matrix given in Figure 4.1, and we assign the value of one to the all of the variables l_i . Assign the weight of zero to each variable u_j below the vertex j_{s_0} , and assign the value of t to each variable u_j between the vertices j_{s_0} and $j_{s_0} + 1$, and ones elsewhere. Note that in this case the degree of t in the minor (j_1, j_2, \dots, j_k) is zero. Assume that there are l indices in $\{i_1, i_2, \dots, i_k\}$ that are strictly larger than j_{s_0} . If $l = 0$, then since the ratio satisfies (ST0) all of the singleton minors in $\beta(A)$ have degree zero with respect to t , but the singletons in the numerator consisting of the indices j_{s_0+1}, \dots, j_k all have degree equal to one. Thus it follows (since the list j_{s_0+1}, \dots, j_k consists of at least one index) that the ratio is unbounded. If $l > 0$, then the degree

of t in the minor (i_1, i_2, \dots, i_k) is 1, and since the ratio satisfies (ST0) it follows that the total degree of the singleton minors in the numerator equals the total degree of the singleton minors in the denominator, and hence again the ratio is unbounded. Finally, assume that $i_{s_0} > j_{s_0}$. In this case we assign the following weighting: zero to all the variables u_j above the vertex j_{s_0} and below the vertex j_{s_0-1} ; for the variables u_j between the vertices j_{s_0-1} and j_{s_0} assign the weight of one, except for the variables u_j between j_{s_0-1} and $j_{s_0-1} + 1$, which we assign the weight of t . Then the degree of t in (j_1, j_2, \dots, j_k) is 0. all singletons in the numerator have degree zero, except (j_{s_0}) , which has degree 1. All singletons in the denominator have degree 0, except at most one, in the special case when $s_0 = 2$ and $j_1 < i_1 \leq j_2$. But in this case the singleton (i_1) has degree of t equal to one, and the degree of (i_1, i_2, \dots, i_k) is 1. Thus in either scenario the ratio is unbounded.

To prove sufficiency it suffices to prove the following inequalities hold with respect to the totally nonnegative matrices:

$$(1): \frac{(1,2,\dots,k-1)(k)}{(1,2,\dots,k-2,k)(k-1)} \leq 1, \text{ and}$$

$$(2): \frac{(1)(2,\dots,k-1)(k)}{(2)(1,3,\dots,k-2,k)(k-1)} \leq 1.$$

These inequalities are verified in Corollaries 4.3.3 and 4.3.4. This completes the proof.

■

Before we come to our main observations for this section we first consider the following very interesting and somewhat unexpected ratio. At present the only known proof of the boundedness of this ratio is with the aid of the symbolic computation

package Maple[©].

PROPOSITION 4.3.6 *For any 4-by-4 nonsingular totally nonnegative matrix the following inequality holds,*

$$\frac{(1, 4)(2, 3)}{(1, 3)(2, 4)} \leq 1.$$

Proof. Using the diagram to generate an arbitrary 4-by-4 nonsingular totally nonnegative matrix we can then calculate each of the four 2-by-2 minors above. In fact, $(1, 4) = 1 + u_3u_2l_3l_2 + u_3l_3$, $(2, 3) = 1 + u_6l_4 + u_6l_1 + u_4l_6 + u_4l_4 + u_6l_6 + u_1l_6 + u_1l_4 + u_1l_1 + u_4l_1 + u_6l_6u_5l_5 + u_6l_6u_5l_2 + u_6l_6u_2l_5 + u_6l_6u_2l_2 + u_6l_4u_5l_2 + u_6l_4u_2l_2 + u_4l_6u_2l_5 + u_4l_6u_2l_2 + u_4l_4u_2l_2$, $(1, 3) = 1 + u_5l_5 + u_5l_2 + u_2l_5 + u_2l_2$, and finally $(2, 4) = 1 + u_6l_4 + u_3l_3 + u_6l_1 + u_4l_6 + u_4l_4 + u_6l_6 + u_1l_6 + u_1l_4 + u_1l_1 + u_4l_1 + u_6l_6u_3l_3 + u_6l_4u_3l_3 + u_6l_1u_3l_3 + u_4l_6u_3l_3 + u_4l_4u_3l_3 + u_4l_1u_3l_3 + u_1l_6u_3l_3 + u_1l_4u_3l_3 + u_1l_1u_3l_3 + u_6l_6u_3u_2l_3l_2 + u_6l_4u_3u_2l_3l_2 + u_4l_6u_3u_2l_3l_2 + u_4l_4u_3u_2l_3l_2$. Next we compute the difference between the expressions for the denominator and the numerator, namely, $(1, 3)(2, 4) - (1, 4)(2, 3) = u_5l_5 + u_5l_2 + u_2l_5 + u_2l_2 + u_5u_4l_5l_4 + u_5u_4l_1l_5 + u_5u_4l_1l_2 + u_5u_1l_5l_4 + u_5u_1l_1l_5 + u_5u_1l_1l_2 + u_2u_1l_5l_4 + u_2u_1l_1l_5 + u_2u_1l_1l_2 + u_6l_4u_5l_5 + u_6l_4u_2l_5 + u_6l_1u_5l_5 + u_6l_1u_5l_2 + u_6l_1u_2l_5 + u_6l_1u_2l_2 + u_4l_6u_5l_5 + u_4l_6u_5l_2 + u_4l_4u_5l_2 + u_4l_4u_2l_5 + u_4l_1u_2l_5 + u_4l_1u_2l_2 + u_1l_6u_5l_5 + u_1l_6u_5l_2 + u_1l_6u_2l_5 + u_1l_6u_2l_2 + u_1l_4u_5l_2 + u_1l_4u_2l_2 + u_5l_5u_3l_3 + u_5l_2u_3l_3 + u_2l_5u_3l_3 + u_5l_5u_6l_4u_3l_3 + u_5l_5u_6l_1u_3l_3 + u_5l_5u_4l_6u_3l_3 + u_5l_5u_4l_4u_3l_3 + u_5l_5u_4l_1u_3l_3 + u_5l_5u_1l_6u_3l_3 + u_5l_5u_1l_4u_3l_3 + u_5l_5u_1l_1u_3l_3 + u_5l_5u_6l_4u_3u_2l_3l_2 + u_5l_5u_4l_6u_3u_2l_3l_2 + u_5l_5u_4l_4u_3u_2l_3l_2 + u_5l_2u_6l_1u_3l_3 + u_5l_2u_4l_6u_3l_3 + u_5l_2u_4l_4u_3l_3 + u_5l_2u_4l_1u_3l_3 + u_5l_2u_1l_6u_3l_3 + u_5l_2u_1l_4u_3l_3 + u_5l_2u_1l_1u_3l_3 +$

$u_5l_2^2u_4l_6u_3u_2l_3 + u_5l_2^2u_4l_4u_3u_2l_3 + u_2l_5u_6l_4u_3l_3 + u_2l_5u_6l_1u_3l_3 + u_2l_5u_4l_4u_3l_3 + u_2l_5u_4l_1u_3l_3 +$
 $u_2l_5u_1l_6u_3l_3 + u_2l_5u_1l_4u_3l_3 + u_2l_5u_1l_1u_3l_3 + u_2^2l_5u_6l_4u_3l_3l_2 + u_2^2l_5u_4l_4u_3l_3l_2$. Observe
 that the above expression is a subtraction free expression in nonnegative variables
 and hence is nonnegative. This completes the proof. ■

Arguably, we could perform the calculations presented above by hand and draw
 the same conclusions, but in this case there is no doubt about the calculations above.
 We note here that the above ratio is self-complementary, self-reversing and appears
 to be a very “tight” inequality in general.

We now move on to a different class of ratios which include as a subclass special
 order-one ratios. As we shall see in the next section the following class of ratios
 turns out to be a very important class for classifying all possible bounded ratios with
 respect to the totally nonnegative matrices for specific values of n . Our main result
 for this new class of ratios is the following.

THEOREM 4.3.7 *Let $\alpha_1, \alpha_2, \beta_1$ and β_2 be subsets of $\{1, 2, \dots, n\}$. Then the ratio*

$$\frac{(\alpha_1)(\alpha_2)}{(\beta_1)(\beta_2)}$$
is bounded with respect to the totally nonnegative matrices if and only if it
satisfies (ST0) and

$$\max(|\alpha_1 \cap L|, |\alpha_2 \cap L|) \geq \max(|\beta_1 \cap L|, |\beta_2 \cap L|),$$

for every contiguous index set $L \subseteq \{1, 2, \dots, n\}$ (i.e., $d(L) = 0$).

To begin our analysis and proof of the above result we make a few definitions and develop some notation. We call an index set L an *interval* if L is a contiguous subset of $\{1, 2, \dots, n\}$ (i.e., $d(L) = 0$). Thus any interval L must be a set of the following form $L = \{i, i+1, \dots, i+k\}$, where $1 \leq i \leq i+k \leq n$. Suppose γ is a subset of $\{1, 2, \dots, n\}$ and L is a given interval of $\{1, 2, \dots, n\}$. Then we let $g(\gamma, L) = \max(|\gamma \cap L|, |\gamma^c \cap L|)$. Finally, suppose that $i_1 < i_2 < \dots < i_k$ and $j_1 < j_2 < \dots < j_l$ are indices of $\{1, 2, \dots, n\}$, so that $k-1 \leq l \leq k+1$. Then we say the *sequence* $\{i_t\}$ *interlaces the sequence* $\{j_t\}$ if one of following cases occur:

1. $l = k + 1$ and $j_1 \leq i_1 \leq j_2 \leq i_2 \leq \dots \leq j_k \leq i_k \leq j_l$;
2. $l = k$ and $i_1 \leq j_1 \leq i_2 \leq j_2 \leq \dots \leq j_{k-1} \leq i_k \leq j_l$; or
3. $l = k - 1$ and $i_1 \leq j_1 \leq i_2 \leq j_2 \leq \dots \leq j_l \leq i_k$.

In the event $\gamma = \{i_1, i_2, \dots, i_k\}$ with $i_j < i_{j+1}$ and $\delta = \{j_1, j_2, \dots, j_l\}$ with $j_i < j_{i+1}$ ($k-1 \leq l \leq k+1$), and the sequence $\{i_t\}$ interlaces the sequence $\{j_t\}$, then we say that γ *interlaces* δ . The next proposition follows immediately from the definitions above.

PROPOSITION 4.3.8 *Let γ and δ be two nonempty index sets of $\{1, 2, \dots, n\}$. If δ interlaces δ^c , then $g(\gamma, L) \geq g(\delta, L)$ for every interval L .*

Suppose $\alpha = \{\alpha_1, \alpha_2\}$ and $\beta = \{\beta_1, \beta_2\}$ are two collections of index sets each consisting of two sets. Then we say that the ratio α/β satisfies condition (M) if

$$\max(|\alpha_1 \cap L|, |\alpha_2 \cap L|) \geq \max(|\beta_1 \cap L|, |\beta_2 \cap L|),$$

for every contiguous index set $L \subseteq \{1, 2, \dots, n\}$. The next observation will be very useful later. Suppose α/β satisfies condition (M), then $C(\alpha/\beta)$, $R(\alpha/\beta)$, $S_i(\alpha/\beta)$ (when defined), $A_i(\alpha/\beta)$ (when defined), and $D_i(\alpha/\beta)$ (when defined) all satisfy condition (M). We split the proof of Theorem 4.3.7 into a sequence of lemmata. We begin with following result.

LEMMA 4.3.9 *Let γ and δ be two nonempty index sets of $N = \{1, 2, \dots, n\}$. If the ratio $\frac{(\gamma)(\gamma^c)}{(\delta)(\delta^c)}$ is bounded with respect to the totally nonnegative matrices, then $g(\gamma, N) \geq g(\delta, N)$, in other words $\max(|\gamma|, |\gamma^c|) \geq \max(|\delta|, |\delta^c|)$.*

Proof. Suppose, on the contrary, that $\max(|\gamma|, |\gamma^c|) < \max(|\delta|, |\delta^c|)$. Consider the following assignment of weights to the general diagram given in Figure 4.1. Let $l_j = 1$ for all j , and $u_i = t$ for all i . Then it is not difficult to show that for this special weighting the degree of t in the principal minor (i_1, i_2, \dots, i_k) is equal to $\sum_{j=1}^k (i_j - j)$. Therefore the degree of t in the product $(\gamma)(\gamma^c)$ is equal to $\binom{n+1}{2} - \frac{1}{2}(|\gamma|^2 + |\gamma| + |\gamma^c|^2 + |\gamma^c|)$. A similar expression exists for the degree of t in the product $(\delta)(\delta^c)$. Thus the degree of t in the ratio $\frac{(\gamma)(\gamma^c)}{(\delta)(\delta^c)}$ is given by $\frac{1}{2}(|\delta|^2 + |\delta^c|^2 + n - |\gamma|^2 - |\gamma^c|^2 - n)$ which is

equal to $\frac{1}{2}(|\delta|^2 + |\delta^c|^2 - |\gamma|^2 - |\gamma^c|^2)$. Notice that since $\max(|\gamma|, |\gamma^c|) < \max(|\delta|, |\delta^c|)$ and $|\delta| + |\delta^c| = |\gamma| + |\gamma^c|$ it follows that $|\delta|^2 + |\delta^c|^2 > |\gamma|^2 + |\gamma^c|^2$. Thus the degree of t in the ratio $\frac{(\gamma)(\gamma^c)}{(\delta)(\delta^c)}$ is positive, and hence the ratio is unbounded with respect to the totally nonnegative matrices. ■

The previous lemma can be extended as follows.

LEMMA 4.3.10 *Let γ and δ be two nonempty index sets of $N = \{1, 2, \dots, n\}$. If the ratio $\frac{(\gamma)(\gamma^c)}{(\delta)(\delta^c)}$ is bounded with respect to the totally nonnegative matrices, then $g(\gamma, L) \geq g(\delta, L)$, for every interval L of N .*

Proof. Suppose there exists an interval L such that $g(\gamma, L) < g(\delta, L)$. Consider the new ratio $\frac{(\gamma_1)(\gamma_1^c)}{(\delta_1)(\delta_1^c)}$, where $\gamma_1 = \gamma \cap L$, $\gamma_1^c = \gamma^c \cap L$, $\delta_1 = \delta \cap L$, and $\delta_1^c = \delta^c \cap L$. Then the condition $g(\gamma, L) < g(\delta, L)$ is equivalent to $\max(|\gamma_1|, |\gamma_1^c|) < \max(|\delta_1|, |\delta_1^c|)$. Thus by Lemma 4.3.9 this new ratio is unbounded with respect to the totally nonnegative matrices. Hence there exists a sequence of totally nonnegative matrices $\{A_i\}$ (of size $|L|$ -by- $|L|$) so that the ratio $\frac{(\gamma_1)(\gamma_1^c)}{(\delta_1)(\delta_1^c)}$ increases without bound when evaluated at this sequence. Define $B_i = I \oplus A_i \oplus I$, such that $B_i[L] = A_i$ for all i . Then B_i is totally nonnegative and the ratio $\frac{(\gamma)(\gamma^c)}{(\delta)(\delta^c)}$ evaluated at B_i is equal to $\frac{(\gamma_1)(\gamma_1^c)}{(\delta_1)(\delta_1^c)}$ evaluated at A_i , which implies the ratio $\frac{(\gamma)(\gamma^c)}{(\delta)(\delta^c)}$ is unbounded with respect to the totally nonnegative matrices. ■

Recall that we have already established the following inequalities with respect to the totally nonnegative matrices:

1. $(1, 2)(\phi) \leq (1)(2)$ (Koteljanskii),

2. $(1, 2)(3) \leq (1, 3)(2)$ (Example 4.1.2), and
3. $(1, 4)(2, 3) \leq (1, 3)(2, 4)$ (Proposition 4.3.6).

The above inequalities will serve as the base cases for the next proof, which uses induction on the number of indices.

THEOREM 4.3.11 *Let γ and δ be two nonempty index sets of $N = \{1, 2, \dots, n\}$.*

Then the ratio $\frac{(\gamma)(\gamma^c)}{(\delta)(\delta^c)}$ is bounded with respect to the totally nonnegative matrices if and only if $g(\gamma, L) \geq g(\delta, L)$, for every interval L of N .

Proof. In Lemma 4.3.10 we proved that the condition $g(\gamma, L) \geq g(\delta, L)$, for every interval L of N is necessary for this ratio to be bounded. Moreover, we note that this condition is sufficient for the aforementioned inequalities above. To verify sufficiency in general suppose $\frac{(\gamma)(\gamma^c)}{(\delta)(\delta^c)}$ is a given ratio and that $g(\gamma, L) \geq g(\delta, L)$, for every interval L of N . Recall that by Proposition 4.3.8 that if δ interlaces δ^c , then the ratio $\frac{(\gamma)(\gamma^c)}{(\delta)(\delta^c)}$ must satisfy $g(\gamma, L) \geq g(\delta, L)$. Define the following new sets: $\gamma_1 = \gamma \cap \delta$, $\gamma_2 = \gamma \cap \delta^c$, $\delta_1 = \gamma^c \cap \delta$, and $\delta_2 = \gamma^c \cap \delta^c$. Then we can write the ratio $\frac{(\gamma)(\gamma^c)}{(\delta)(\delta^c)}$ as $\frac{(\gamma_1, \gamma_2)(\delta_1, \delta_2)}{(\gamma_1, \delta_2)(\delta_1, \gamma_2)}$. Consider the following decomposition (or factorization) of $\frac{(\gamma)(\gamma^c)}{(\delta)(\delta^c)}$:

$$\frac{(\gamma_1, \gamma_2)(\delta_1, \delta_2)}{(\gamma_1, \delta_2)(\delta_1, \gamma_2)} = \frac{(\gamma_1, \gamma_2)(\gamma_{11}, \delta_{12}, \delta_2)}{(\gamma_1, \delta_2)(\gamma_{11}, \delta_{12}, \gamma_2)} \cdot \frac{(\gamma_{11}, \delta_{12}, \gamma_2)(\delta_1, \delta_2)}{(\gamma_{11}, \delta_{12}, \delta_2)(\delta_1, \gamma_2)}, \quad (4.3)$$

where $\gamma_1 = \{\gamma_{11}, \gamma_{12}\}$ and $\delta_1 = \{\delta_{11}, \delta_{12}\}$. Observe that the case γ_1 and δ_1 are both empty cannot occur since the sets γ and δ are assumed to be nonempty. Consider the

first factor $\frac{(\gamma_1, \gamma_2)(\delta_{11}, \delta_{12}, \delta_2)}{(\gamma_1, \delta_2)(\delta_{11}, \delta_{12}, \gamma_2)}$ where $\gamma_1 = \{\gamma_{11}, \gamma_{12}\}$. Then the boundedness of this ratio is equivalent to the boundedness of $\frac{(\gamma_{12}, \gamma_2)(\delta_{12}, \delta_2)}{(\gamma_{12}, \delta_2)(\delta_{12}, \gamma_2)}$ since deletion of common indices (here we deleted the common set γ_{11}) from every index set preserves the boundedness of a ratio (see Proposition 4.2.9). Similarly, the second factor is bounded if and only if $\frac{(\gamma_{11}, \gamma_2)(\delta_{11}, \delta_2)}{(\gamma_{11}, \delta_2)(\delta_{11}, \gamma_2)}$ is bounded. We now make the following key claim. Suppose we are able to decompose $\gamma_1 = \{\gamma_{11}, \gamma_{12}\}$ and $\delta_1 = \{\delta_{11}, \delta_{12}\}$ such that $\frac{(\gamma_1)(\delta_1)}{(\gamma_{11}, \delta_{12})(\delta_{11}, \gamma_{12})}$ satisfies condition (M), namely,

$$\max(|\gamma_1 \cap L|, |\delta_1 \cap L|) \geq \max(|(\gamma_{11} \cup \delta_{12}) \cap L|, |(\delta_{11} \cup \gamma_{12}) \cap L|).$$

Then each of the factors, $\frac{(\gamma_{12}, \gamma_2)(\delta_{12}, \delta_2)}{(\gamma_{12}, \delta_2)(\delta_{12}, \gamma_2)}$ and $\frac{(\gamma_{11}, \gamma_2)(\delta_{11}, \delta_2)}{(\gamma_{11}, \delta_2)(\delta_{11}, \gamma_2)}$, must satisfy condition (M).

To prove this claim fix an interval L . Then there are three possible cases: (1):

$|\gamma \cap L| > |\gamma^c \cap L|$; (2): $|\gamma \cap L| < |\gamma^c \cap L|$; or (3): $|\gamma \cap L| = |\gamma^c \cap L|$. Suppose case (1) holds.

Then since $g(\gamma, L) \geq g(\delta, L)$ it follows that $|\gamma_1 \cap L| \geq |\delta_1 \cap L|$ and $|\gamma_2 \cap L| \geq |\delta_2 \cap L|$.

However, if $\frac{(\gamma_1)(\delta_1)}{(\gamma_{11}, \delta_{12})(\delta_{11}, \gamma_{12})}$ satisfies condition (M) and $|\gamma_1 \cap L| \geq |\delta_1 \cap L|$, then applying

similar reasoning reveals that $|\gamma_{11} \cap L| \geq |\delta_{11} \cap L|$ and $|\gamma_{12} \cap L| \geq |\delta_{12} \cap L|$. Thus the

following four inequalities hold: (i): $|\gamma_1 \cap L| \geq |\delta_1 \cap L|$; (ii): $|\gamma_2 \cap L| \geq |\delta_2 \cap L|$; (iii):

$|\gamma_{11} \cap L| \geq |\delta_{11} \cap L|$; and (iv): $|\gamma_{12} \cap L| \geq |\delta_{12} \cap L|$. But the inequalities (i)-(iv) imply

that both of the ratios $\frac{(\gamma_{12}, \gamma_2)(\delta_{12}, \delta_2)}{(\gamma_{12}, \delta_2)(\delta_{12}, \gamma_2)}$ and $\frac{(\gamma_{11}, \gamma_2)(\delta_{11}, \delta_2)}{(\gamma_{11}, \delta_2)(\delta_{11}, \gamma_2)}$ satisfy condition (M), for the

fixed interval L . Similar analysis holds for cases (2) and (3) when $|\gamma \cap L| \leq |\gamma^c \cap L|$

and is omitted here. This completes the proof of the claim. The proof of the result

will be complete if we can show that there always exists a choice for γ_1 and δ_1 so that we may decompose them and ensure that the ratio $\frac{(\gamma_1)(\delta_1)}{(\gamma_{11}, \delta_{12})(\delta_{11}, \gamma_{12})}$ satisfies condition (M). It is not difficult to show that no such decomposition of γ_1 and δ_1 exists only if γ_1 and δ_1 interlace. Consider the original ratio $\frac{(\gamma_1, \gamma_2)(\delta_1, \delta_2)}{(\gamma_1, \delta_2)(\delta_1, \gamma_2)}$, and suppose that γ_1 and δ_1 (and hence that γ_2 and δ_2) interlace independent of the choice for γ_1 and δ_1 . In this case we may assume, (by complementation) without loss of generality, that $\{\gamma_1, \delta_2\}$ consists of all the odd integers of N and $\{\delta_1, \gamma_2\}$ consists of all the even integers of N . Since if this was not the case, then there would exist (at least) one consecutive pair in at least one of the sets $\{\gamma_1, \delta_2\}$ or $\{\delta_1, \gamma_2\}$, and since $g(\gamma, L) \geq g(\delta, L)$ there must exist (at least) one consecutive pair in γ or δ . Hence it follows that there exists a choice for γ_1 and δ_1 so that they do not interlace, contradicting our assumption.

Thus we assume that $\{\gamma_1, \delta_2\}$ consists of all the odd integers of N and $\{\delta_1, \gamma_2\}$ consists of all the even integers of N , and that γ_1 and δ_1 interlace, and γ_2 and δ_2 interlace. We also assume that $1 \in \gamma_1$. Then there are two cases to consider: (1): $2 \in \delta_1$; (2): $2 \in \gamma_2$.

Case (1): $2 \in \delta_1$. Then $\gamma_1 = \{1, 3, 5, \dots, 2k - 1, *\}$, for some $k \geq 1$; $\gamma_2 = \{2l + 2, *\}$, for some $l \geq 1$; $\delta_1 = \{2, 4, \dots, 2l, *\}$; and $\delta_2 = \{2k + 1, *\}$. There are two subcases to consider: (1a): $l \geq k$; or (1b): $l < k$. For the case (1a) let $\gamma_{11} = \{1\}$, $\gamma_{12} = \{3, 5, \dots, 2k - 1, *\}$, $\delta_{11} = \{2\}$, and $\delta_{12} = \{4, \dots, 2l, *\}$. Then we claim that the sets $\{\gamma_{12}, \delta_2\}$ and $\{\delta_{12}, \gamma_2\}$ interlace and $\{\gamma_{11}, \delta_2\}$ and $\{\delta_{11}, \gamma_2\}$ interlace. Note that if this was the case then the factors $\frac{(\gamma_{12}, \gamma_2)(\delta_{12}, \delta_2)}{(\gamma_{12}, \delta_2)(\delta_{12}, \gamma_2)}$ and $\frac{(\gamma_{11}, \gamma_2)(\delta_{11}, \delta_2)}{(\gamma_{11}, \delta_2)(\delta_{11}, \gamma_2)}$ would satisfy condition (M). To verify that the above sets interlace observe that

$\{\gamma_{11}; \delta_2\}, \{\delta_{11}; \gamma_2\} = \{1; 2k + 1, *\}\{2; 2l + 2, *\}$. Since $l \geq k$ we have that $2k + 1 < 2l + 2$, hence interlacing follows, since γ_2 and δ_2 interlace by assumption. Similarly, $\{\gamma_{12}; \delta_2\}, \{\delta_{12}; \gamma_2\} = \{3, 5, \dots, 2k - 1, *; 2k + 1, *\}\{4, \dots, 2l; 2l + 2, *\}$, and again interlacing follows since $l \geq k$ and γ_2 and δ_2 interlace. For case (1b), let $\gamma_{11} = \{1, 3, 5, \dots, 2l + 1\}$, $\gamma_{12} = \gamma_1 \cap \gamma_{11}^c$, and let $\delta_{11} = \{2, \dots, 2l\}$, $\delta_{12} = \delta_1 \cap \delta_{11}^c$. Then in this case $\{\gamma_{11}; \delta_2\}, \{\delta_{11}; \gamma_2\} = \{1, 3, 5, \dots, 2l + 1; 2k + 1, *\}\{2, 4, \dots, 2l; 2l + 2, *\}$. and since $l < k$ and γ_2 and δ_2 interlace it follows that $\{\gamma_{11}, \delta_2\}$ and $\{\delta_{11}, \gamma_2\}$ interlace. Similarly, $\{\gamma_{12}, \delta_2\}$ and $\{\delta_{12}, \gamma_2\}$ interlace since we have just discarded the first $2l + 1$ consecutive integers. Thus in either subcase the factors in the decomposition given in (4.3) satisfy condition (M).

Case (2): $2 \in \gamma_2$. In this case note that $3 \in \delta_2$, since otherwise $3 \in \gamma_1$, and hence the triple $\{1, 2, 3\}$ is contained in γ , and therefore there is no choice for γ_1 and δ_1 so that they interlace which contradicts our assumption. For simplicity we relabel the sets in question as follows: $\gamma_1 = \delta'_2$, $\gamma_2 = \delta'_1$, $\delta_1 = \gamma'_2$, and $\delta_2 = \gamma'_1$. So using the new labels we have that $1 \in \delta'_2$, $2 \in \delta'_1$ and $3 \in \gamma'_1$. Moreover, $\gamma'_1 = \{3, 5, \dots, 2k - 1, *\}$, for some $k \geq 2$, $\gamma'_2 = \{2l + 1, *\}$ for some $l \geq 1$, $\delta'_1 = \{2, 4, \dots, 2l, *\}$, and $\delta'_2 = \{1, 2k + 1, *\}$. Now we let $\gamma'_{11} = \{3\}$, $\gamma'_{12} = \{5, \dots, 2k - 1, *\}$, $\delta'_{11} = \{2\}$, and $\delta'_{12} = \{4, \dots, 2l, *\}$. As before it now follows that $\{\gamma'_{11}, \delta'_2\}$ and $\{\delta'_{11}, \gamma'_2\}$ interlace, and that $\{\gamma'_{12}, \delta'_2\}$ and $\{\delta'_{12}, \gamma'_2\}$ interlace.

Thus in all cases each of the factors in (4.3) satisfy condition (M) and since both factors involve fewer indices, we have that both are bounded ratios by induction. As noted before the base cases for this induction argument are the inequalities:

$(1, 2)(\phi) \leq (1)(2)$, $(1, 2)(3) \leq (1, 3)(2)$, and $(1, 4)(2, 3) \leq (1, 3)(2, 4)$, which have been previously verified. This completes the proof. ■

Proof of Theorem 4.3.7: Note that the ratio $\frac{(\alpha_1)(\alpha_2)}{(\beta_1)(\beta_2)}$ is bounded with respect to the totally nonnegative matrices if and only if $\frac{(S)(S^c)}{(T)(T^c)}$ is bounded with respect to the totally nonnegative matrices, for some S and T obtained from α_1, α_2 and β_1, β_2 , respectively, by deleting common indices and shifting (as these operations preserve boundedness). ■

There are many very useful consequences to Theorems 4.3.7 and 4.3.11, which we state here.

COROLLARY 4.3.12 *Let γ and δ be two nonempty index sets of $N = \{1, 2, \dots, n\}$.*

Then the ratio $\frac{(\gamma)(\gamma^c)}{(\delta)(\delta^c)}$ is bounded with respect to the totally nonnegative matrices if and only if it is bounded by one.

The proof of the above result follows directly from Theorem 4.3.11 and the fact that the base case ratios are bounded by one.

COROLLARY 4.3.13 *Let $\alpha_1, \alpha_2, \beta_1$ and $\beta_2 \subseteq \{1, 2, \dots, n\}$. Then the ratio $\frac{(\alpha_1)(\alpha_2)}{(\beta_1)(\beta_2)}$*

is bounded with respect to the totally nonnegative matrices if and only if it is bounded by one.

This result follows from Corollary 4.3.12 and the fact that applying the operations of deletion and shifting preserves the inequality.

COROLLARY 4.3.14 *Let γ and δ be two nonempty index sets of $N = \{1, 2, \dots, n\}$*

with $n \geq 5$. If the ratio $\frac{(\gamma)(\gamma^c)}{(\delta)(\delta^c)}$ is bounded, then it can be factored into ratios of the

same type with fewer indices.

For the next consequence we let $(odds) = (1, 3, 5, \dots)$ (i.e., the minors consisting of all the odd integers of N) and $(evens) = (2, 4, 6, \dots)$ (the minor consisting of all the even integers of N). Then the next result follows directly from Theorem 4.3.11.

COROLLARY 4.3.15 *For $n \geq 2$, and $\gamma \subseteq \{1, 2, \dots, n\}$,*

$$(1, 2, 3, \dots, n)(\phi) \leq (\gamma)(\gamma^c) \leq (odds)(evens),$$

for any totally nonnegative matrix.

Of course the first inequality is Fischer's inequality.

The next two results are very useful for studying the relationships between extremal ratios for specific values of n (see the next section for more information).

COROLLARY 4.3.16 *Suppose γ and δ are nonempty subsets of $\{1, 2, \dots, n\}$, and assume that $j \notin \gamma \cup \delta$. Then $\frac{(\gamma)(\delta \cup \{j\})}{(\delta)(\gamma \cup \{j\})} \leq 1$, with respect to the totally nonnegative matrices if and only if $|\delta \cap L| \geq |\gamma \cap L|$, for every interval L with $j \in L$.*

Proof. Assume that γ and δ are nonempty subsets of $\{1, 2, \dots, n\}$, and assume that $j \notin \gamma \cup \delta$. Then, by Theorem 4.3.11, the ratio $\frac{(\gamma)(\delta \cup \{j\})}{(\delta)(\gamma \cup \{j\})} \leq 1$ if and only if $\max(|\gamma \cap L|, |(\delta \cup \{j\}) \cap L|) \geq \max(|(\gamma \cup \{j\}) \cap L|, |\delta \cap L|)$, for each interval L . Observe that if $j \notin L$, then equality holds for the maxima above. If $j \in L$, then

$\max(|\gamma \cap L|, |(\delta \cup \{j\}) \cap L|) \geq \max(|(\gamma \cup \{j\}) \cap L|, |\delta \cap L|)$ holds if and only if $|\delta \cap L| \geq |\gamma \cap L|$. ■

COROLLARY 4.3.17 *Suppose γ and δ are nonempty subsets of $\{1, 2, \dots, n\}$, and assume that $i, j \notin \gamma \cup \delta$. Then $\frac{(\gamma \cup \{i\})(\delta \cup \{j\})}{(\delta \cup \{i\})(\gamma \cup \{j\})} \leq 1$, with respect to the totally nonnegative matrices if and only if the following conditions hold.*

(i) *If $i \in L$ and $j \notin L$, then $|\gamma \cap L| \geq |\delta \cap L|$,*

(ii) *If $j \in L$ and $i \notin L$, then $|\delta \cap L| \geq |\gamma \cap L|$,*

for each such interval L .

4.4 General Case for $n \leq 5$

DEFINITION 4.4.1 A collection C of bounded ratios is referred to as *generators* if and only if any bounded ratio can be written as products of positive powers of ratios from C .

The idea used throughout this section is to assign the weights of $0, 1, t, t^{-1}$, where t is a nonnegative variable, which we make arbitrarily large, to the variables $\{l_i\}$ $\{u_j\}$. We do this in different combinations to produce as many necessary conditions for boundedness as possible. For such a weighting each principal minor will then be a function in terms of t . For a ratio of principal minors to be bounded, the degree of t in the denominator must be greater than or equal to the degree of t in the numerator.

Therefore, we compute the degree of t in each minor and note that the *total* degree of t in the ratio must be less than 0 for the ratio to be bounded.

Firstly, we consider the 3-by-3 case in full detail. Every ratio on 3 indices can be written in the following way: $(1)^{x_1}(2)^{x_2}(3)^{x_3}(12)^{x_{12}}(13)^{x_{13}}(23)^{x_{23}}(123)^{x_{123}}$, where x_{α_i} is the degree of (α_i) in the ratio (that is, x_{α_i} is negative when there are more copies of (α_i) in the denominator of the given ratio). Let y_{α_i} be the degree of t in (α_i) . Then the expression $y_1x_1+y_2x_2+y_3x_3+y_{12}x_{12}+y_{13}x_{13}+y_{23}x_{23}+y_{123}x_{123}$, represents the degree of t in this ratio. Since by assumption $(1) = (12) = (123) = 1$, $y_1 = y_{12} = y_{123} = 0$, and it is sufficient to consider the expression $y_2x_2 + y_3x_3 + y_{13}x_{13} + y_{23}x_{23}$. For this ratio to be bounded, $y_2x_2 + y_3x_3 + y_{13}x_{13} + y_{23}x_{23} \leq 0$. Using diagrams with edges set to various values, we develop a list of conditions, each of the form $y_2x_2 + y_3x_3 + y_{13}x_{13} + y_{23}x_{23} \leq 0$. Observe that these conditions are linear inequalities in the variables x_{α_i} and represent hyperplanes in the space given by the variables $\{x_{\alpha_i}\}$. We also note here that a given ratio satisfies ST0 (for $n = 3$) if and only if: $x_1 + x_{12} + x_{13} + x_{123} = 0$, $x_2 + x_{12} + x_{23} + x_{123} = 0$, and $x_3 + x_{13} + x_{23} + x_{123} = 0$.

EXAMPLE 4.4.2 Let A be a 3-by-3 nonsingular totally nonnegative matrix, with the corresponding diagram in Figure 4.3.

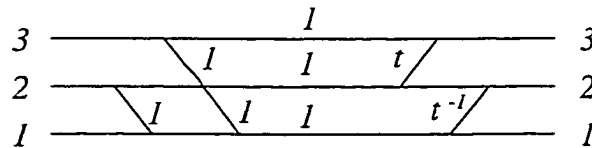


Figure 4.3: 3-by-3 example diagram.

If an edge does not appear, it is assumed to have a weight of 0. Using this diagram we compute all possible principal minors. Recall that $(1) = (12) = (123) = 1$. There are three possible paths for (2) . First, it can go straight across the diagram and yield a weight of 1. Second, this path can use the first edge (joining 1 and 2), with a weight of 1, and then traverse the edge with weight t^{-1} , yielding a total weight of t^{-1} . Finally, this path can use the second edge (joining 1 and 2) and then traverse the edge with t^{-1} . Again this yields a total weight of t^{-1} . Therefore, $(2) = 1 + 2t^{-1}$. Employing similar techniques we can also show that $(3) = 1 + t$, $(13) = 1 + t$, and $(23) = 2 + 2t^{-1}$.

Using this information, it follows that $y_2 = 0$, $y_3 = 1$, $y_{13} = 1$, and $y_{23} = 0$. Therefore, the expression $x_3 + x_{13}$ represents the total degree of t in any ratio for the diagram in Figure 4.3. Thus in order for a ratio to be bounded, it must satisfy the condition $x_3 + x_{13} \leq 0$. For completeness we note that in this case

$$A = \begin{bmatrix} 1 & t^{-1} & 0 \\ 2 & 1 + 2t^{-1} & t \\ 1 & 1 + t^{-1} & 1 + t \end{bmatrix}.$$

By considering various diagrams, we derive many necessary conditions which a ratio must satisfy in order for it to be bounded.

Each ratio can be written as a vector, with each entry representing the degree of a specific principal minor in the ratio. We label the entries lexicographically. Since $(1) = (12) = (123) = 1$, we disregard these minors, because they do not affect whether or not the ratio is bounded. This vector, therefore, takes the form: $[x_2, x_3, x_{13}, x_{23}]^T$.

If a ratio is bounded, its corresponding vector will lie in the convex hull of the cone formed by the intersection of these hyperplanes.

EXAMPLE 4.4.3 By Fischer's inequality, the ratio $\frac{(13)}{(1)(3)}$ is bounded. Therefore, the vector $[0, -1, 1, 0]^T$ will lie in the cone of bounded ratios (or, in our case it even is an extremal of our cone).

Note that some specific weightings will produce redundant conditions. So, we search for conditions which are positively linearly independent (in other words, they cannot be written as a nonnegative linear combination of the remaining conditions). The intersection of these hyperplanes forms the cone of interest, and the extremals of this cone happen to be the generators.

In the 3-by-3 case, four necessary conditions were found (see Figure 4.4) to define the cone of bounded ratios. Because the edges of the left side of each diagram are all weighted 1, it is sufficient to consider the right side only.

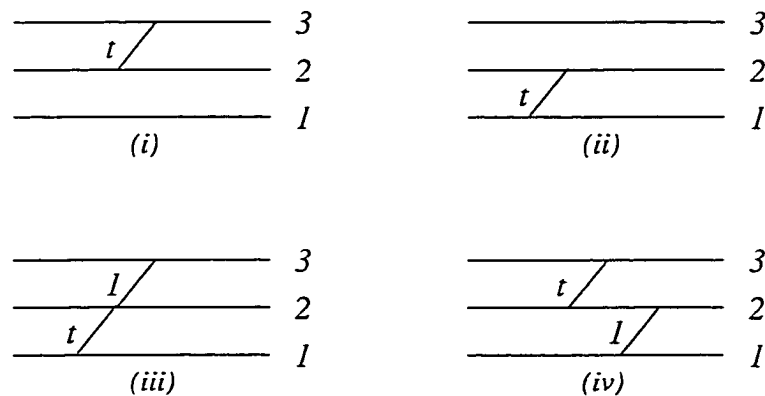


Figure 4.4: Necessary diagrams for 3-by-3 case.

Each of these graphs (in Figure 4.4) gives rise to a new hyperplane:

4.4.i $x_3 + x_{13} \leq 0,$

4.4.ii $x_2 + x_{23} \leq 0,$

$$4.4.iii \quad x_2 + x_3 + x_{23} \leq 0,$$

$$4.4.iv \quad x_3 + x_{13} + x_{23} \leq 0.$$

Suppose we are given a list of inequalities in n variables. A set $C \subset \mathbb{R}^n$ is a *polyhedral cone* if and only if $C = \{x \in \mathbb{R}^n : Ax \leq 0\}$, where A is an m -by- n real matrix.

THEOREM 4.4.4 *Suppose α/β is a ratio of principal minors on three indices. Then α/β is bounded with respect to the totally nonnegative matrices if and only if α/β can be written as a product of positive powers of the following bounded ratios:*

$$\frac{(13)(\emptyset)}{(1)(3)}, \quad \frac{(123)(2)}{(12)(23)}, \quad \frac{(12)(3)}{(13)(2)}, \quad \frac{(23)(1)}{(13)(2)}.$$

Proof. We first note that the boundedness of the above four ratios has already been verified. The two ratios on the left are Koteljanskii ratios; the third was shown to be bounded in Example 4.1.2; and the boundedness of the final ratio follows from the third by reversing the indices. The four diagrams in Figure 4.4 give rise to the following four necessary conditions:

$$(a) \quad x_3 + x_{13} \leq 0;$$

$$(b) \quad x_2 + x_{23} \leq 0;$$

$$(c) \quad x_2 + x_3 + x_{23} \leq 0; \text{ and}$$

(d) $x_3 + x_{13} + x_{23} \leq 0$.

Recall that the necessary conditions (STO) are equivalent to the following three equations: $x_1 + x_{12} + x_{13} + x_{123} = 0$, $x_2 + x_{12} + x_{23} + x_{123} = 0$, and $x_3 + x_{13} + x_{23} + x_{123} = 0$. For simplicity of the analysis we use the equations above to convert the four inequalities to “greater than” inequalities, namely,

(a) $x_{23} + x_{123} \geq 0$;

(b) $x_{12} + x_{123} \geq 0$;

(c) $x_{12} + x_{13} + x_{23} + 2x_{123} \geq 0$; and

(d) $x_{123} \geq 0$.

Fix $x_{123} = k \geq 0$. Then the intersection given by the linear inequalities: $x_{23} \geq -k$, $x_{12} \geq -k$, and $x_{12} + x_{13} + x_{23} \geq -2k$, forms a polyhedral cone. Translate the variables $(x_{12}, x_{23}, x_{13})^T$ to $(x_{12} + k, x_{23} + k, x_{13})^T = (z_{12}, z_{23}, z_{13})^T$ so that the corresponding inequalities, in terms of the z 's, are translated to the origin. The extremals of this polyhedral cone are determined by the rays formed from the intersection of any two of the three hyperplanes. For example, the intersection of $z_{12} = z_{23} = 0$ is the ray given by $\{(0, 0, t)^T : t \geq 0\}$. The remaining extreme rays are determined similarly. In this case the polyhedral cone C formed by the intersection of the above hyperplanes is given by $C = \{t_1(0, 0, 1)^T + t_2(0, 1, -1)^T + t_3(1, 0, -1)^T : t_i \geq 0, \sum_i t_i = 1\}$. Therefore any vector $v \in C$, may be written as $v = (z_{12}, z_{23}, z_{13})^T = (t_3, t_2, t_1 - t_2 - t_3)^T$. Writing these equations in terms of the x 's gives $x_{12} = t_3 - k$, $x_{23} = t_2 - k$, and $x_{13} = t_1 - t_2 - t_3$.

Using the equations given by (ST0) we may now solve for the remaining variables, for example, $x_1 = -x_{12} - x_{13} - x_{123} = t_2 - k$. Similarly, it follows that $x_2 = k - t_2 - t_3$, and $x_3 = t_3 - t_1$. Finally, we substitute the above x -values back into the original ratio, that is

$$\begin{aligned} & (123)^{x_{123}}(12)^{x_{12}}(13)^{x_{13}}(23)^{x_{23}}(1)^{x_1}(2)^{x_2}(3)^{x_3} \\ &= (123)^k(12)^{t_3-k}(13)^{t_1-t_2-t_3}(23)^{t_2-k}(1)^{t_2-t_1}(2)^{k-t_2-t_3}(3)^{t_3-t_1} \\ &= \left[\frac{(123)(2)}{(12)(23)} \right]^k \left[\frac{(13)(\phi)}{(1)(3)} \right]^{t_1} \left[\frac{(23)(1)}{(13)(2)} \right]^{t_2} \left[\frac{(12)(3)}{(13)(2)} \right]^{t_3}. \end{aligned}$$

where $k, t_1, t_2, t_3 \geq 0$. Observe that the two ratios on the left are bounded (in fact by one) as they are special Koteljanskii ratios, and the latter two ratios are bounded (again by one) by Theorem 4.3.7. This completes the proof. ■

COROLLARY 4.4.5 *The cone of bounded ratios on three indices with respect to the totally nonnegative matrices is generated by*

$$\frac{(13)(\emptyset)}{(1)(3)}, \frac{(123)(2)}{(12)(23)}, \frac{(12)(3)}{(13)(2)}, \frac{(23)(1)}{(13)(2)}.$$

COROLLARY 4.4.6 *Any ratio α/β of principal minors on three indices is bounded with respect to the totally nonnegative matrices if and only if it is bounded by one.*

The next result provides some extra information concerning the relationships among the list of generators for bounded ratios on three indices.

PROPOSITION 4.4.7 *Each of the 3-by-3 generators can be derived from $\frac{(13)(\emptyset)}{(1)(3)}$ and $\frac{(12)(3)}{(13)(2)}$, by using the operators which preserve bounded ratios in Propositions 4.2.3 and 4.2.7.*

$$A_2\left(\frac{(13)(\emptyset)}{(1)(3)}\right) = \frac{(123)(2)}{(12)(23)}$$

$$A_2\left(\frac{(13)(\emptyset)}{(1)(3)}\right) = \frac{(123)(2)}{(12)(23)}$$

$$R\left(\frac{(12)(3)}{(13)(2)}\right) = \frac{(23)(1)}{(13)(2)}.$$

For the cases $n = 4$ and 5 the analysis is similar but not surprisingly more involved. The number of necessary conditions increase, the number of variables increase, and the number of generators (or extremals) increases. However, the mathematics in question is very similar, and to aid our computations we use a C^{++} program (see [Fu]) specifically designed to compute the extreme rays of a given polyhedral cone (not to mention other tasks associated with linear programming). Thus most of the technical details are hidden and taken care of, by this program. Our plan here is to describe the idea of how the conditions were found, and how they were used to eventually determine the extremals, which in our case are ratios of principal minors (see [FGG]).

We begin by taking the four diagrams from the 3-by-3 case in Figure 4.4 and systematically append a fourth vertex, and some extra edges, to generate various conditions again using only the four possible weights $0, 1, t, t^{-1}$ (as in the 3-by-3 case). We set up the problem in a similar manner, namely, we know that any ratio on four indices may be represented by $(1234)^{x_{1234}}(123)^{x_{123}}(124)^{x_{124}} \dots (2)^{x_2}(1)^{x_1}$, where x_α is the multiplicity of (α) in the ratio. We also let (as before) y_α denote the degree of

t in the minor (α) . Recall that we assume the above ratio satisfies (ST0), which is equivalent to the following four equations: (1): $x_1 + x_{12} + x_{13} + x_{14} + x_{123} + x_{124} + x_{134} + x_{1234} = 0$; (2): $x_2 + x_{12} + x_{23} + x_{24} + x_{123} + x_{124} + x_{234} + x_{1234} = 0$; (3): $x_3 + x_{13} + x_{23} + x_{34} + x_{123} + x_{134} + x_{234} + x_{1234} = 0$; and (4): $x_4 + x_{14} + x_{24} + x_{34} + x_{124} + x_{134} + x_{234} + x_{1234} = 0$. Using these four equations we are able to reduce the number of variables in our system from 15 to 11.

After a sufficient list of conditions, in the form of linear inequalities in the variables x_α , has been determined we then input these conditions in the form of a matrix system of inequalities $Ax \leq 0$ into the program [Fu], and analyze the output. The output will be in the form of a list of vectors each of which are the extreme rays of the polyhedral cone $\{x \in \mathbb{R}^{11} : Ax \leq 0\}$. For each extremal we must verify whether or not it is a bounded ratio with respect to the totally nonnegative matrices. If it is, then we have found a generator for all of the bounded ratios on four indices. If we prove that the extreme ratio is not bounded, then we must develop another condition (or geometrically, cut the existing polyhedral cone by another hyperplane) to exclude this ratio. Consider the following example. Suppose for simplicity, the following ratio was outputted from a run of this program, $\frac{(1,3)(2,4)}{(1,4)(2,3)}$. We conclude this ratio is not bounded since it does not satisfy the condition given in Proposition 4.1.8 for the set $\{2, 3\}$. However, we need to derive conditions in terms of the variables x_α to exclude this ratio. To do this we use the symbolic software package Maple[©]. This package aids in designing a diagram for which the degree of t in the numerator of the above ratio is larger than the degree of t in the denominator. We note here that these diagrams are

by no means unique and many different diagrams may give rise to the same condition. For this particular ratio we might come up with the following diagram (see Figure 4.5).

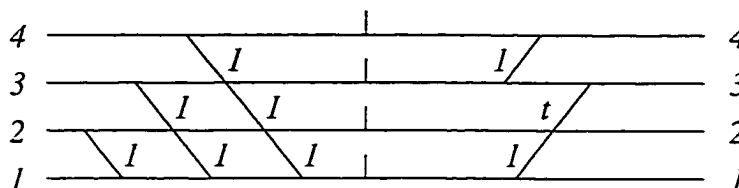


Figure 4.5: Particular Diagram

Using this diagram we obtain the inequality $x_3 + x_{13} + x_{34} + x_{134} \leq 0$. We note here that the suspect ratio above does not satisfy this inequality, and hence adding this inequality to our list will discard this suspect ratio from the list of extremals. One of the drawbacks to this type of search and destroy method is that we may (and often do) develop redundant constraints or conditions. However, one of the by-products of the program [Fu] is that a list of the redundant condition is also outputted, and eventually can be discarded. The final diagrams (twelve in total) produced for the 4-by-4 case are shown in Figure 4.6 at the end of this chapter, and as in the 3-by-3 case the left-hand side of the diagram is omitted as all edges are assigned the weight of 1, and any missing edges are assumed to have weight 0.

The corresponding inequalities are as follows:

$$(i): x_4 + x_{14} + x_{24} + x_{124} \leq 0,$$

$$(ii): x_3 + x_4 + x_{13} + x_{14} + x_{23} + x_{24} + x_{34} + x_{124} + x_{134} + x_{234} \leq 0,$$

$$(iii): x_2 + x_3 + x_4 + x_{13} + x_{14} + x_{34} + x_{134} \leq 0,$$

$$(iv): x_4 + x_{14} + x_{24} + x_{34} + x_{124} + x_{134} \leq 0,$$

$$(v): x_3 + x_{13} + x_{34} + x_{134} \leq 0,$$

$$(vi): x_2 + x_3 + x_{23} + x_{24} + x_{34} + x_{234} \leq 0,$$

$$(vii): x_3 + x_4 + x_{13} + x_{14} + x_{23} + x_{24} + x_{234} \leq 0,$$

$$(viii): x_2 + x_{23} + x_{24} + x_{234} \leq 0,$$

$$(ix): x_4 + x_{14} + x_{24} + x_{34} + x_{124} + x_{134} + x_{234} \leq 0,$$

$$(x): x_3 + x_{13} + x_{23} + x_{34} + x_{134} + x_{234} \leq 0,$$

$$(xi): x_2 + x_4 + x_{14} + x_{23} + 2x_{24} + x_{34} + x_{124} + x_{234} \leq 0,$$

$$(xii): x_3 + x_4 + x_{13} + x_{14} + x_{23} + x_{24} + 2x_{34} + x_{134} + x_{234} \leq 0.$$

The above twelve linear inequalities may be represented via the following matrix system:

$$Ax = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 2 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_2 \\ x_3 \\ x_4 \\ x_{13} \\ x_{14} \\ x_{23} \\ x_{24} \\ x_{34} \\ x_{124} \\ x_{134} \\ x_{234} \end{bmatrix} \leq 0.$$

Using the above twelve inequalities we obtained the following result. (See Proposition 4.4.10 for demonstration of the boundedness of the listed ratios.)

THEOREM 4.4.8 *Suppose α/β is a ratio of principal minors on four indices. Then α/β is bounded with respect to the totally nonnegative matrices if and only if α/β can be written as a product of positive powers of the following bounded ratios:*

$$\frac{(14)(\emptyset)}{(1)(4)}, \frac{(2)(124)}{(12)(24)}, \frac{(3)(134)}{(13)(34)}, \frac{(23)(1234)}{(123)(123)}, \frac{(12)(3)}{(13)(2)}, \frac{(1)(24)}{(2)(14)}, \frac{(2)(34)}{(3)(24)},$$

$$\frac{(4)(13)}{(3)(14)}, \frac{(12)(134)}{(13)(124)}, \frac{(13)(234)}{(23)(134)}, \frac{(34)(124)}{(24)(134)}, \frac{(24)(123)}{(23)(124)}, \frac{(14)(23)}{(13)(24)}.$$

We can make conclusions that are remarkably similar to the consequences for the case when $n = 3$. One surprising similarity between the above list of generators for $n = 4$, and the list for $n = 3$, is that each generator consists of two sets over two sets.

COROLLARY 4.4.9 *The cone of bounded ratios on four indices with respect to the totally nonnegative matrices is generated by*

$$\frac{(14)(\emptyset)}{(1)(4)}, \frac{(2)(124)}{(12)(24)}, \frac{(3)(134)}{(13)(34)}, \frac{(23)(1234)}{(123)(123)}, \frac{(12)(3)}{(13)(2)}, \frac{(1)(24)}{(2)(14)}, \frac{(2)(34)}{(3)(24)},$$

$$\frac{(4)(13)}{(3)(14)}, \frac{(12)(134)}{(13)(124)}, \frac{(13)(234)}{(23)(134)}, \frac{(34)(124)}{(24)(134)}, \frac{(24)(123)}{(23)(124)}, \frac{(14)(23)}{(13)(24)}.$$

The next result provides some extra information concerning the relationships among the list of generators for bounded ratios on four indices. Recall that the following inequalities are valid for any totally nonnegative matrix: $(1, 2)(\phi) \leq (1)(2)$ (Kotljanskii), $(1, 2)(3) \leq (1, 3)(2)$ (Example 4.1.2), and $(1, 4)(2, 3) \leq (1, 3)(2, 4)$ (Proposition 4.3.6).

PROPOSITION 4.4.10 *Each of the 4-by-4 generators can be derived from $\frac{(14)(\emptyset)}{(1)(4)}$, $\frac{(12)(3)}{(13)(2)}$, and $\frac{(14)(23)}{(13)(24)}$, by using the operators which preserve bounded ratios in Propositions 4.2.3 and 4.2.7.*

Since the ratios $\frac{(14)(\emptyset)}{(1)(4)}$, $\frac{(12)(3)}{(13)(2)}$, and $\frac{(14)(23)}{(13)(24)}$ are not only bounded, but bounded by one with respect to the totally nonnegative matrices, we have the following result.

COROLLARY 4.4.11 *Any ratio α/β of principal minors on four indices is bounded with respect to the totally nonnegative matrices if and only if it is bounded by one.*

As discussed previously we used similar analysis to tackle the $n = 5$ case. However, for this case the computation became increasingly more involved and more time consuming. We now have 26 variables x_α and many more possibilities for diagrams, and hence constraints. Never the less, using the program [Fu] we were able to completely characterize all the bounded ratios on five indices. (An interested reader is encouraged to consult the REU report [FGG] for more details on the $n = 5$ case including diagrams and analysis.)

THEOREM 4.4.12 *Suppose α/β is a ratio of principal minors on five indices. Then α/β is bounded with respect to the totally nonnegative matrices if and only if α/β can be written as a product of positive powers of the following bounded ratios:*

$$\begin{aligned} & \frac{(15)(\emptyset)}{(1)(5)}, \frac{(2)(125)}{(12)(25)}, \frac{(3)(135)}{(13)(35)}, \frac{(4)(145)}{(14)(45)}, \frac{(23)(1235)}{(235)(123)}, \frac{(24)(1245)}{(124)(245)}, \frac{(34)(1345)}{(134)(345)}, \frac{(234)(12345)}{(1234)(2345)}, \\ & \frac{(12)(3)}{(2)(13)}, \frac{(13)(4)}{(3)(14)}, \frac{(14)(5)}{(15)(4)}, \frac{(1)(25)}{(2)(15)}, \frac{(2)(35)}{(3)(25)}, \frac{(3)(45)}{(4)(35)}, \frac{(12)(135)}{(13)(125)}, \frac{(13)(235)}{(23)(135)}, \frac{(14)(245)}{(24)(145)}, \\ & \frac{(13)(145)}{(14)(135)}, \frac{(23)(245)}{(24)(235)}, \frac{(123)(24)}{(124)(23)}, \frac{(124)(25)}{(24)(125)}, \frac{(34)(124)}{(24)(134)}, \frac{(35)(134)}{(34)(135)}, \frac{(45)(135)}{(35)(145)}, \frac{(35)(125)}{(25)(135)}, \\ & \frac{(24)(345)}{(34)(245)}, \frac{(123)(1245)}{(124)(1235)}, \frac{(124)(1345)}{(134)(1245)}, \frac{(134)(2345)}{(234)(1345)}, \frac{(235)(1234)}{(234)(1235)}, \frac{(245)(1235)}{(235)(1245)}, \\ & \frac{(345)(1245)}{(245)(1345)}, \frac{(14)(23)}{(13)(24)}, \frac{(15)(24)}{(14)(25)}, \frac{(25)(34)}{(24)(35)}, \frac{(125)(134)}{(124)(135)}, \frac{(135)(234)}{(134)(235)}, \frac{(145)(235)}{(135)(245)} \end{aligned}$$

Again we note that this list of generators (38 in total) all consist of two sets over two sets just as in the $n = 3$ and $n = 4$ cases. Moreover, this list really only depends on three “basic” ratios (also similar to the $n = 3$ and $n = 4$ cases).

PROPOSITION 4.4.13 *Each of the 5-by-5 generators can be derived from $\frac{(15)(0)}{(1)(5)}$, $\frac{(12)(3)}{(13)(2)}$, and $\frac{(14)(23)}{(13)(24)}$, by using the operators which preserve bounded ratios in Propositions 4.2.3 and 4.2.7.*

We also have the following result that parallels the previous cases.

COROLLARY 4.4.14 *Any ratio α/β of principal minors on five indices is bounded with respect to the totally nonnegative matrices if and only if it is bounded by one.*

4.5 Summary and Further Discussion

In this chapter we completely characterized all multiplicative principal minor inequalities through $n = 5$. Moreover, we have demonstrated many particular classes of bounded ratios, for example, $\frac{(\alpha_1)(\alpha_2)}{(\beta_1)(\beta_2)}$ for all values of n . However, as with most topics in mathematics many unresolved issues still remain. It is the purpose of this section to address some of these issues.

We begin with the topic of the existence of a bounded ratio with respect to the totally nonnegative matrices that is not bounded by one. First we remark that in the case of bounded ratios with respect to the positive definite matrices, there do exist bounded ratios which are not bounded by one (see [BJ2]). On the other hand, any

ratio that is bounded with respect to the M -matrices is necessarily bounded by one (see [FHJ]). Throughout this chapter we have demonstrated many bounded ratios and in every case these ratios were bounded by one. In fact, we do not know of a single ratio that is bounded with respect to the totally nonnegative matrices and is **not** bounded by one. Furthermore, a consequence of the results in section 4 is that any bounded ratio on at most 5 indices is, indeed, bounded by one. Moreover, we conjecture this to be the case in general, namely, that any bounded ratio with respect to the totally nonnegative matrices is necessarily bounded by one.

Another remarkable characteristic of bounded ratios for totally nonnegative matrices is the notion of *subtraction free expressions* in the nonnegative variables from a bidiagonal factorization of an arbitrary totally nonnegative matrix. Recall that in Proposition 4.3.6 and Example 4.1.5 we demonstrated the boundedness of the ratios $\frac{(14)(23)}{(13)(24)}$ and $\frac{(12)(3)}{(13)(2)}$, respectively, by showing that $(13)(24) - (14)(23)$, and $(13)(2) - (12)(3)$ are subtraction free expressions in the nonnegative variables $\{l_i\}$, $\{u_j\}$ and $\{d_k\}$ from a bidiagonal factorization of an arbitrary totally nonnegative matrix. Moreover, this phenomenon holds for every extremal (or generator) through $n = 5$, from which it follows that any bounded ratio with respect to the totally nonnegative matrices on at most five indices also satisfies this phenomenon. At present every ratio that is known to be bounded by one satisfies this property. Clearly, if a given ratio satisfies the property that the difference between the denominator and numerator is a subtraction free expression in nonnegative variables, then this ratio is bounded by one. The resolution of the issue of whether or not a ratio can be bounded with respect

to the totally nonnegative matrices, and the difference between the denominator and numerator is a subtraction free expression will definitely shed some light on whether there exist bounded ratios with respect to the totally nonnegative matrices that are not bounded by one.

A natural next step in studying these types of determinantal inequalities is to classify all such inequalities for $n = 6$, and so on. We have begun such analysis for the case $n = 6$, however, we have encountered some difficulties with sufficient (or lack thereof) computer support. For the case $n = 6$, there are 57 unknown variables and at present we have approximately 400 necessary conditions. Using some (rudimentary) prediction criteria we anticipate roughly 1300 necessary conditions and around a 130 generators. In order to continue the analysis on the $n = 6$ case we have to develop some different techniques and ideas. We are currently pursuing many different ideas along these lines.

The final topic we address here is the notion of *extremality*. At present, we know very little about extremal ratios, and predicting when a given ratio is extremal or not. Consider the Koteljanskii ratio $\frac{(1,n)(\phi)}{(1)(n)}$. Through $n = 5$, this ratio is extremal, however, we cannot prove this in general, nor show that it is not extremal for some n . (Although we do conjecture that this ratio is extremal for all n .) Also through $n = 5$, it can be shown that if α/β is extremal, then $A_i(\alpha/\beta)$ and $D_j(\alpha/\beta)$ (when defined) are extremal ratios. Currently, this is also a mystery. Finally, we comment on the appearance of the extremal ratios through $n = 5$. Each such extremal is of the form $\frac{(\alpha_1)(\alpha_2)}{(\beta_1)(\beta_2)}$. This was unexpected and for the most part still remains unexplained.

However, during the computation in the $n = 6$ case, the following ratio was outputted,

$$\frac{(1346)(1256)(124)(34)(25)}{(1246)(356)(134)(125)(24)}.$$

This ratio seems to represent a “new” class of bounded ratios with respect to the totally nonnegative matrices. The proof that this ratio is bounded is very computational and is omitted here. However, given that this ratio is bounded we may conclude that this ratio is extremal. Thus our initial guess that all extremals are of the form $\frac{(\alpha_1)(\alpha_2)}{(\beta_1)(\beta_2)}$ is incorrect. Another curious open problem concerning this ratio is that we do not know if this ratio is bounded by one. Part of the reason for this is various computer limitations, although we were able to demonstrate boundedness. Perhaps this is an example of a bounded ratio that is not bounded by one.

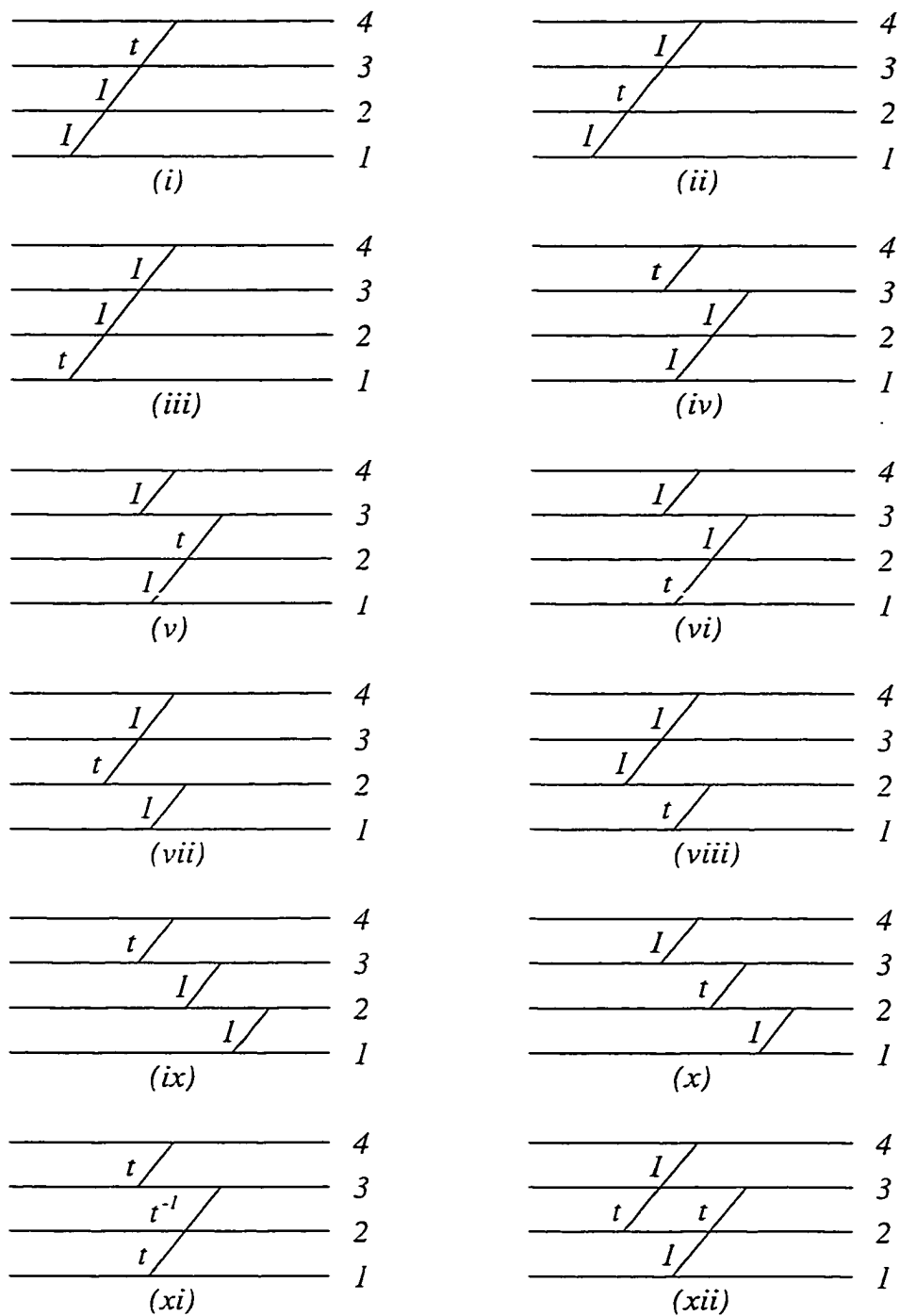


Figure 4.6: The 4-by-4 Diagrams.

Chapter 5

Hadamard Products of Totally Nonnegative Matrices

In this chapter we investigate the entry-wise (or Hadamard) product of totally nonnegative matrices. Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two m -by- n matrices. Then the *Hadamard product* of A and B , denoted by $A \circ B$, is the m -by- n matrix whose $(i, j)^{th}$ entry is given by $a_{ij}b_{ij}$. See [HJ1, HJ2] for a more detailed discussion of the Hadamard product and its role in matrix theory.

The Hadamard product of certain TN matrices has come up before in the study of the roots of a given polynomial. We discuss this role briefly now for background and motivation. Let $f(x) = \sum_{i=0}^n a_i x^i$ be an n^{th} degree polynomial in x . By the *Routh-Hurwitz matrix* we mean the n -by- n matrix given by

$$A = \begin{bmatrix} a_1 & a_3 & a_5 & a_7 & \cdots & 0 & 0 \\ a_0 & a_2 & a_4 & a_6 & \cdots & 0 & 0 \\ 0 & a_1 & a_3 & a_5 & \cdots & 0 & 0 \\ 0 & a_0 & a_2 & a_4 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & a_{n-1} & 0 \\ 0 & 0 & 0 & 0 & \cdots & a_{n-2} & a_n \end{bmatrix}.$$

A polynomial $f(x)$ is said to be *stable* if all the zeros of $f(x)$ have negative real parts. It is proved in [As], for example, that $f(x)$ is stable if and only if the Routh-Hurwitz matrix formed from f is totally nonnegative. Suppose $f(x) = \sum_{i=0}^n a_i x^i$ and $g(x) = \sum_{i=0}^m b_i x^i$ are two polynomials of degree n and m , respectively. Then the *Hadamard product of f and g* is defined to be the polynomial $(f \circ g)(x) = \sum_{i=0}^k a_i b_i x^i$, where $k = \min(m, n)$. In [GW1] it is proved that the Hadamard product of stable polynomials is stable. Thus a consequence of this result (which is also noted in [GW1]) is that the Hadamard product of two totally nonnegative Routh-Hurwitz matrices is in turn a totally nonnegative matrix. See also [GW2] for a list of other subclasses of TN matrices that are closed under Hadamard multiplication. This list includes such classes as: totally nonnegative Vandermonde matrices; tridiagonal and inverse tridiagonal TN matrices, and others. The reader is also referred to [Wa] where yet another class of totally nonnegative matrices is shown to be closed under Hadamard multiplication.

Our interest lay in studying when in general the Hadamard product of two totally nonnegative matrices is again totally nonnegative. Recall the celebrated classical result of I. Schur (see [HJ1], for example) which states that the Hadamard product

of two positive semidefinite matrices is again a positive semidefinite matrix. Unfortunately, this is not the case for two general TN matrices. Consider the following example (see also [J] for another example).

EXAMPLE 5.0.1 Let $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, and let $B = A^T$. Then A and B are TN. Furthermore, $A \circ B = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$, and $\det(A \circ B) = -1 < 0$. Thus $A \circ B$ is not TN.

In light of this example and previous remarks a characterization of when the Hadamard product of two TN matrices is a TN matrix is delicate and difficult.

We close this introduction with an overview of this chapter. We begin with definitions and a preliminary discussion of the Hadamard dual of totally nonnegative matrices. In the next section we present some general results and completely characterize the Hadamard dual for 3-by- n totally nonnegative matrices. We then show that the matrices in the Hadamard dual satisfy the classical inequality of Oppenheim concerning $\det(A \circ B)$. Also considered are sign patterns of matrices in the Hadamard dual.

5.1 Notation and General Set-up

We begin with the definition of the Hadamard dual of the m -by- n totally nonnegative matrices.

DEFINITION 5.1.1 The *Hadamard dual* for the totally nonnegative matrices is defined to be the set of all m -by- n matrices A such that $A \circ B$ is totally nonnegative, for all m -by- n totally nonnegative matrices B .

In symbols,

$$\{A : A \circ B \in \text{TN}, \forall B \in \text{TN}\}.$$

Often we may just refer to the Hadamard dual to mean the Hadamard dual for TN matrices. We now present various elementary properties for matrices in the Hadamard dual.

Observe that J , the m -by- n matrix of all ones is the Hadamard identity, that is $A \circ J = J \circ A = A$. Thus, it is clear that J is in the Hadamard dual. Also note that J is TN. This simple fact turns out to be quite useful for proving other results about matrices in the Hadamard dual, as demonstrated in the next result.

COROLLARY 5.1.2 *Let A be an m -by- n matrix in the Hadamard dual. Then A is totally nonnegative.*

Proof. Since J is TN and A is in the Hadamard dual, we have $A \circ J = A$ is TN.

■

Thus the Hadamard dual is nonempty and is contained among the totally nonnegative matrices, for all m and n .

PROPOSITION 5.1.3 *Suppose A and B are two m -by- n matrices in the Hadamard dual. Then $A \circ B$, the Hadamard product of A and B , is in the Hadamard dual.*

Proof. Let C be any m -by- n TN matrix. Then $B \circ C$ is TN since B is in the Hadamard dual. Hence $A \circ (B \circ C)$ is TN. But $A \circ (B \circ C) = (A \circ B) \circ C$. Thus $A \circ B$ is in the Hadamard dual, since C was arbitrary. ■

Recall that if $D = [d_{ij}]$ is a diagonal matrix, then $\det DA[\alpha|\beta] = \det D[\alpha] \det A[\alpha|\beta]$. Hence if A is TN, then DA is TN, for every entry-wise nonnegative (and hence totally nonnegative) diagonal matrix D . Moreover, observe that $D(A \circ B) = DA \circ B = A \circ DB$, from which it follows that DA is in the Hadamard dual whenever D is a TN diagonal matrix and A is in the Hadamard dual. The above facts aid in the proof of the next proposition.

PROPOSITION 5.1.4 *Any rank one totally nonnegative matrix is in the Hadamard dual.*

Proof. Let A be a rank one TN matrix, say $A = xy^T$, where $x = [x_i] \in \mathbb{R}^m$ and $y = [y_i] \in \mathbb{R}^n$ are entry-wise nonnegative vectors. Let $D = \text{diag}(x_i)$ and $E = \text{diag}(y_i)$. Then it is easy to show that $A = DJE$. (Observe that $J = ee^T$, where e is the vector of all ones. Then $DJE = D(ee^T)E = (De)(e^TE) = xy^T = A$.) Since J is in the Hadamard dual, we have that DJE is in the Hadamard dual. In other words A is in the Hadamard dual. ■

Note that Example 5.0.1 implies that not all rank two TN matrices are in the Hadamard dual, and in fact by directly summing the matrix A in Example 5.0.1 with an identity matrix it follows that there exist TN matrices of all ranks greater than one that are not in the Hadamard dual. We now state a very useful result concerning an inheritance property for matrices in the Hadamard dual.

PROPOSITION 5.1.5 *An m -by- n totally nonnegative matrix A is in the Hadamard dual if and only if every submatrix of A is in the corresponding Hadamard dual.*

Proof. The sufficiency is trivial since A is a submatrix of itself. To prove necessity, suppose there exists a submatrix, say $A[\alpha|\beta]$, that is not in the Hadamard dual. Then there exists a TN matrix B such that $A[\alpha|\beta] \circ B$ is not TN. Embed B into an m -by- n matrix $C = [c_{ij}]$ such that $C[\alpha|\beta] = B$, and $c_{ij} = 0$ otherwise. It is not difficult to show that C is TN. Now consider $A \circ C$. Since $A[\alpha|\beta] \circ B$ is a submatrix of $A \circ C$ and $A[\alpha|\beta] \circ B$ is not TN, we have that $A \circ C$ is not TN. This completes the proof. ■

We conclude this section with a result concerning the column vectors that can be appended to a given matrix in the Hadamard dual in such a way so that the resulting matrix remains in the Hadamard dual.

PROPOSITION 5.1.6 *The set of columns (or rows) that can be appended to an m -by- n TN matrix in the Hadamard dual so that the resulting matrix remains in the Hadamard dual is a nonempty convex set.*

Proof. Suppose A is an m -by- n TN matrix in the Hadamard dual. Let S denote the set of columns that can be appended to A so that the new matrix remains in the Hadamard dual. Since $0 \in S$, we have that $S \neq \emptyset$. Let $x, y \in S$. Then the augmented matrices $[A|x]$ and $[A|y]$ are both in the Hadamard dual. Suppose $t \in [0, 1]$ and consider the matrix $[A|tx + (1-t)y]$. Let $[B|z]$ be any m -by- $(n+1)$ TN

matrix. Then

$$[A|tx + (1 - t)y] \circ [B|z] = [A \circ B|t(x \circ z) + (1 - t)(y \circ z)].$$

Since A is in the Hadamard dual any submatrix of $A \circ B$ is TN. Therefore we only need to consider the submatrices of $[A|tx + (1 - t)y] \circ [B|z]$ that involve column $n + 1$. Let $[A'|tx' + (1 - t)y'] \circ [B'|z']$ denote any such square submatrix of $[A|tx + (1 - t)y] \circ [B|z]$.

Then

$$\begin{aligned} & \det ([A'|tx' + (1 - t)y'] \circ [B'|z']) \\ &= \det ([A' \circ B'|t(x' \circ z')]) \\ & \quad + \det ([A' \circ B'|(1 - t)(y' \circ z')]) \\ &= t \det ([A' \circ B'|x' \circ z']) \\ & \quad + (1 - t) \det ([A' \circ B'|y' \circ z']) \\ &= t \det ([A'|x'] \circ [B'|z']) \\ & \quad + (1 - t) \det ([A'|y'] \circ [B'|z']) \\ &\geq 0, \end{aligned}$$

since both $[A|x]$ and $[A|y]$ are in the Hadamard dual. This completes the proof. ■

5.2 Elements of the Hadamard Dual

In this section we are concerned with describing various elements in the Hadamard dual. This description is two fold, either we will describe elements in the Hadamard dual for certain values of m, n ; or we will describe certain special subclasses of totally nonnegative matrices that are contained in the Hadamard dual for all m, n . Recall, for example, that all rank one TN matrices are members of the Hadamard dual (see Proposition 5.1.4). Firstly, we explore other possible subclasses that are contained in the Hadamard dual.

Observe that an n -by- n diagonal matrix D is TN if and only if D is entry-wise nonnegative. This fact is a simple exercise and we do not bother proving it here. A consequence of this fact is the next result.

PROPOSITION 5.2.1 *Any n -by- n diagonal totally nonnegative matrix is in the Hadamard dual.*

Proof. Let D be any n -by- n diagonal TN matrix, and let B be an arbitrary n -by- n TN matrix. Then $D \circ B$ is an entry-wise nonnegative diagonal matrix, and hence is TN. ■

Recall that an n -by- n matrix is called an *upper (lower) bidiagonal matrix* if the only possible nonzero entries occur on the main diagonal and on the super- (sub-) diagonal, and all remaining entries are necessarily zero. We claim that an n -by- n upper bidiagonal (similar arguments apply for the case of a lower bidiagonal matrix) matrix is TN if and only if it is entry-wise nonnegative. We offer the following argument

for completeness. The necessity of this claim is clear and verification of sufficiency for $n \leq 2$ is trivial. Thus suppose the sufficiency of the claim holds for all t -by- t upper bidiagonal matrices with $t < n$. Let U be as in the claim. Then, by induction, $U[\{1, 2, \dots, n-1\}]$ is TN. Hence $U[\{1, 2, \dots, n-1\}|\{1, 2, \dots, n\}]$ is TN since it is obtained from $U[\{1, 2, \dots, n-1\}]$ by appending a zero column and (possibly) increasing the $(n-1, n)^{th}$ entry (see Proposition 2.4.4). Then U is obtained from $U[\{1, 2, \dots, n-1\}|\{1, 2, \dots, n\}]$ by appending a zero row and (possibly) increasing the (n, n) entry. An immediate consequence of this claim is the following fact.

PROPOSITION 5.2.2 *Every n -by- n totally nonnegative bidiagonal matrix is a member of the Hadamard dual.*

Proof. Similar to the proof of Proposition 5.2.1. ■

An n -by- n matrix $A = [a_{ij}]$ is said to be a *tridiagonal matrix* if $a_{ij} = 0$ whenever, $|i - j| > 1$. A non-obvious, but well-known fact is the next proposition. The proof presented here offers a nice application of k -subdirect sums that is discussed in Chapter 2.

PROPOSITION 5.2.3 *Let T be an n -by- n tridiagonal matrix. Then T is totally nonnegative if and only if T is an entry-wise nonnegative P_0 -matrix.*

Proof. The necessity is obvious. The proof of sufficiency is by induction on the size of T . Observe that this result is trivial for $n \leq 2$. Suppose the result holds for all such tridiagonal matrices of size k , where $k < n$. Let T be as in the statement of

the proposition. Notice that T may be partitioned as follows

$$T = \left[\begin{array}{cc|c} t_{11} & t_{12} & 0 \\ t_{21} & t_{22} & x_{23} \\ \hline 0 & y_{32} & T_{33} \end{array} \right],$$

where T_{33} is $(n - 2)$ -by- $(n - 2)$ and t_{ij} are scalars for $1 \leq i, j \leq 2$. Hence T is in the special 1-subdirect sum form. Since T is assumed to be a P_0 -matrix, by Theorem 2.8.2, T can be written as $T = T_1 \oplus_1 T_2$, where T_1 is 2-by-2 and T_2 is $(n - 1)$ -by- $(n - 1)$, and both T_1 and T_2 are P_0 -matrices. Since both T_1 and T_2 are also entry-wise nonnegative tridiagonal P_0 -matrices, by induction T_1 and T_2 are both TN matrices. Now, by Theorem 2.8.3, $T_1 \oplus_1 T_2 = T$ is a TN matrix. This completes the proof. ■

The above fact was probably first proved in [GK2], where they refer to tridiagonal matrices as Jacobi matrices. We are now in a position to extend the result of Garloff and Wagner [GW2] concerning the Hadamard product of tridiagonal matrices.

THEOREM 5.2.4 *Let T be an n -by- n totally nonnegative tridiagonal matrix. Then T is in the Hadamard dual.*

Proof. It is enough to prove this result for the case when T is irreducible, otherwise apply the following argument to each irreducible block. Let B be an arbitrary n -by- n TN matrix. Similarly we may assume B is irreducible, which implies $b_{ij} > 0$ for all i, j such that $|i - j| \leq 1$, i.e., B has positive “tri-diagonal part” (see Lemma 3.0.8). Since pre- and postmultiplication by positive diagonal matrices does not affect the property of being TN or whether or not a matrix is in the Hadamard dual, we may assume that $b_{ii} = 1$, for $i = 1, 2, \dots, n$ and that $b_{ij} = b_{ji}$ for all i, j with $|i - j| = 1$. Notice that if $S = \text{diag}(1, -1, 1, -1, \dots, \pm 1)$, then STS has nonpositive

off-diagonal entries, and since T was TN it follows that STS is a (possibly singular) M -matrix (see [HJ2]). In this case there exists a positive diagonal matrix D_1 such that $D_1STS D_1^{-1} = S(D_1TD_1^{-1})S$ is a row diagonally dominant matrix (see [HJ2]). (A matrix $C = [c_{ij}]$ is said to be *row diagonally dominant* if $|b_{ii}| \geq \sum_{j \neq i} |b_{ij}|$, for $i = 1, 2, \dots, n$.) Let $C = [c_{ij}] = S(D_1TD_1^{-1})S \circ B = S(D_1TD_1^{-1} \circ B)S$. Since B is TN with $b_{ii} = 1$ and $b_{ij} = b_{ji}$ whenever $|i - j| = 1$, it follows that $0 < b_{ij} \leq 1$ for all i, j with $|i - j| = 1$. Hence $D_1TD_1^{-1} \circ B$ is row diagonally dominant. Now since $D_1TD_1^{-1} \circ B$ is tridiagonal we have that $S(D_1TD_1^{-1} \circ B)S$ has nonpositive off-diagonal entries. By a result in [HJ2] we have that $S(D_1TD_1^{-1} \circ B)S$ is a (possibly singular) M -matrix. Therefore $D_1TD_1^{-1} \circ B$ is a nonnegative tridiagonal P_0 -matrix. Hence, by Proposition 5.2.3, $D_1TD_1^{-1} \circ B$ is a TN matrix, and hence $T \circ B$ is a TN matrix. Thus T is in the Hadamard dual. ■

We obtain a result of Garloff and Wagner [GW2] as a special case.

COROLLARY 5.2.5 *The Hadamard product of any two n -by- n tridiagonal totally nonnegative matrices is again totally nonnegative.*

We now consider describing the Hadamard dual for some specific values of m and n . We begin with the simple case $m = n = 2$.

LEMMA 5.2.6 *Let $A = [a_{ij}]$ be any 2-by-2 totally nonnegative matrix. Then A is in the Hadamard dual.*

Proof. Let $B = [b_{ij}]$ be any 2-by-2 TN matrix. Then $A \circ B$ is entry-wise nonnegative and $\det(A \circ B) = a_{11}a_{22}b_{11}b_{22} - a_{12}a_{21}b_{12}b_{21} \geq 0$, since $a_{11}a_{22} \geq a_{12}a_{21}$

and $b_{11}b_{22} \geq b_{12}b_{21}$. ■

COROLLARY 5.2.7 *Let A be an 2-by- n totally nonnegative matrix. Then A is in the Hadamard dual.*

Proof. If $n \leq 2$, then we are done. Thus assume $n \geq 3$. By Proposition 5.1.5 it is enough to check that all of the 2-by-2 submatrices of A are in the Hadamard dual. By Lemma 5.2.6 all such submatrices are members of the Hadamard dual. Hence A is in the Hadamard dual. ■

5.3 The 3-by- n Case

The analysis for the 3-by-3 Hadamard dual differs significantly from the 2-by-2 case, and unfortunately unlike the 2-by-2 case, not all 3-by-3 totally nonnegative matrices are in the Hadamard dual. Recall from Example 5.0.1 that the matrix $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ does not belong in the Hadamard dual. We let W denote the above 3-by-3 matrix. We will see that W plays an important role concerning matrices in the Hadamard dual. We begin our analysis of the 3-by-3 Hadamard dual with a preliminary lemma concerning a special class of 3-by-3 totally nonnegative matrices in the Hadamard dual.

LEMMA 5.3.1 *Let $A = \begin{bmatrix} 1 & 1 & a \\ 1 & 1 & a \\ a & a & 1 \end{bmatrix}$. Then A is in the Hadamard dual if and only if A is totally nonnegative.*

Proof. The necessity follows from Proposition 5.1.2. To verify sufficiency suppose A is TN. Let $B = [b_{ij}]$ be any 3-by-3 TN matrix. By virtue of the 2-by-

2 case it is enough to show that $\det(A \circ B) \geq 0$. We make use of Sylvester's identity for determinants (see (1.3)). Note that we may assume that $b_{22} > 0$, otherwise B is reducible in which case verification of $\det(A \circ B) \geq 0$ is trivial. Using Sylvester's identity (see chapter 1) we see that $\det B \geq 0$ is equivalent to $\frac{(b_{11}b_{22}-b_{12}b_{21})(b_{22}b_{33}-b_{23}b_{32})}{b_{22}} \geq \frac{(b_{12}b_{23}-b_{22}b_{13})(b_{21}b_{32}-b_{31}b_{22})}{b_{22}}$. Since A is TN, $0 \leq a \leq 1$. Observe that

$$\begin{aligned} & \frac{(b_{11}b_{22} - b_{12}b_{21})(b_{22}b_{33} - b_{23}b_{32}a^2)}{b_{22}} \\ & \geq \frac{(b_{11}b_{22} - b_{12}b_{21})(b_{22}b_{33} - b_{23}b_{32})}{b_{22}}, \text{ since } 0 \leq a \leq 1 \\ & \geq \frac{(b_{12}b_{23} - b_{22}b_{13})(b_{21}b_{32} - b_{31}b_{22})}{b_{22}}, \text{ since } \det B \geq 0 \\ & \geq a^2 \frac{(b_{12}b_{23} - b_{22}b_{13})(b_{21}b_{32} - b_{31}b_{22})}{b_{22}}, \text{ since } 0 \leq a \leq 1. \end{aligned}$$

Therefore $\frac{(b_{11}b_{22}-b_{12}b_{21})(b_{22}b_{33}-b_{23}b_{32}a^2)}{b_{22}} \geq a^2 \frac{(b_{12}b_{23}-b_{22}b_{13})(b_{21}b_{32}-b_{31}b_{22})}{b_{22}}$, which implies $\det(A \circ B) \geq 0$, and hence A is in the Hadamard dual. ■

LEMMA 5.3.2 *Let A be a 3-by-3 singular totally nonnegative matrix. If $A \circ W$ and $A \circ W^T$ are both totally nonnegative, then A is in the Hadamard dual.*

Proof. In light of the 2-by-2 case we may assume that A is irreducible. Moreover, up to positive diagonal equivalence we may also assume A is in the following form,

$$A = \begin{bmatrix} 1 & a & c \\ a & 1 & b \\ d & b & 1 \end{bmatrix}.$$
 Since A is singular, $\det A = 1 + abc + abd - a^2 - b^2 - cd = 0$, or

$1 + abc + abd = a^2 + b^2 + cd$. By hypothesis $A \circ W$ and $A \circ W^T$ are both totally nonnegative, hence

1. $\det(A \circ W) = 1 + abd - a^2 - b^2 \geq 0$, and

2. $\det(A \circ W^T) = 1 + abc - a^2 - b^2 \geq 0$.

Since A is a singular TN matrix it follows that equality must hold in both 1. and 2.

above. This gives gives rise to one of the following four cases:

1. $c = 0$, and $ab = c$,
2. $c = 0$, and $d = 0$,
3. $d = 0$, and $ab = d$,
4. $ab = d$, and $ab = c$.

Suppose B is an arbitrary 3-by-3 TN matrix, as with A , we may assume that B is in the following form $B = \begin{bmatrix} 1 & \alpha & \gamma \\ \alpha & 1 & \beta \\ \delta & \beta & 1 \end{bmatrix}$.

Observe that cases (1) and (3) cannot occur since A was assumed to be irreducible.

In case (2) A is tridiagonal, and hence is in the Hadamard dual by Theorem 5.2.4.

Finally, consider case (4) and suppose that $ab = d$, and $ab = c$. Then $\det A = 1 + (ab)^2 - a^2 - b^2 = (1 - a^2)(1 - b^2) = 0$. Therefore either $a = 1$ or $b = 1$. In either case A is of the form in Lemma 5.3.1 and hence is in the Hadamard dual. ■

We are now in a position to characterize all 3-by-3 TN matrices that are members of the Hadamard dual.

THEOREM 5.3.3 *Let A be a 3-by-3 totally nonnegative matrix. Then A is in the Hadamard dual if and only if $A \circ W$ and $A \circ W^T$ are both totally nonnegative, where*

$$W = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} .$$

Proof. The necessity is clear since W (and hence W^T) is TN. Assume that $A \circ W$ and $A \circ W^T$ are both TN. By Lemma 5.3.2 it suffices to assume that A is nonsingular. In this case by Theorem 2.2.2 A admits a bidiagonal factorization which up to positive diagonal factor we may assume is

$$\begin{aligned} A &= \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & b & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ c & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & d & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & e \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & f & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & f+d & de \\ a+c & (a+c)f + (a+c)d + 1 & ((a+c)d + 1)e \\ bc & bcf + bcd + b & 1 + (bcd + b)e \end{bmatrix}, \end{aligned}$$

where $a, b, c, d, e, f \geq 0$. Let B be any nonsingular TN matrix. Similarly, we assume that B is in the form

$$B = \begin{bmatrix} 1 & z+x & xy \\ u+w & (u+w)z + (u+w)x + 1 & ((u+w)x + 1)y \\ vw & vwz + vwx + v & 1 + (vwx + v)y \end{bmatrix},$$

where $u, v, w, x, y, z \geq 0$. A routine computation reveals that $\det(A \circ W) = 1 - abde$, hence $1 \geq abde$, since $A \circ W$ is TN. Similarly, one can show that $\det(A \circ W^T) = 1 - bcef$, and hence $1 \geq bcef$. Now we compute $\det(A \circ B)$. A routine but lengthy computation shows that

$$\begin{aligned} \det(A \circ B) &= zwyv - zwyvabde + xuyv - xuyvbcef + (\text{nonnegative terms}) \\ &\geq zwyv(1 - abde) + xuyv(1 - bcef) \geq 0. \end{aligned}$$

Hence $A \circ B$ is TN for all nonsingular TN matrices B (the 2-by-2 submatrices follow from Lemma 5.2.6). The fact that $A \circ B$ is TN for all 3-by-3 TN matrices B follows by a routine continuity argument since any TN matrix is the limit of nonsingular (in particular TP) TN matrices (see Theorem 2.1.6 and [W]). ■

We now present many different consequences and equivalent statements of Theorem 5.3.3.

COROLLARY 5.3.4 Let $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ be a 3-by-3 totally nonnegative matrix. Then A is in the Hadamard dual if and only if

$$aei + gbf \geq afh + dbi,$$

$$aei + dch \geq afh + dbi.$$

EXAMPLE 5.3.5 Let $P = \begin{bmatrix} 1 & p^2 & p^4 \\ p^2 & 1 & p^2 \\ p^4 & p^2 & 1 \end{bmatrix}$, where $p \in (0, 1)$. Then A is TN and we claim that A is in the Hadamard dual. To see this observe that the inequalities in Corollary 5.3.4 are equivalent to $1 + p^8 \geq p^4 + p^4$, which is of course equivalent to $(1 - p^4)^2 \geq 0$.

EXAMPLE 5.3.6 (Polya Matrix) Let $q \in (0, 1)$. Define the n -by- n Polya matrix Q whose $(i, j)^{th}$ entry is equal to q^{-2ij} . Then it is well-known (see [W]) that Q is

totally positive for all n (in fact Q is diagonally equivalent to a TP Vandermonde matrix). Suppose Q represents the 3-by-3 Polya matrix. We wish to determine when (if ever) Q is in the Hadamard dual. By Corollary 5.3.4 and the fact that Q is symmetric implies Q is in the Hadamard dual if and only if $q^{-28} + q^{-22} \geq q^{-26} + q^{-26}$, which is equivalent to $q^{-28}(1 - q^2 - q^2(1 - q^4)) \geq 0$. This inequality holds if and only if $1 - q^2 \geq q^2(1 - q^4) = q^2(1 - q^2)(1 + q^2)$. Thus q must satisfy $q^4 + q^2 - 1 \leq 0$. It is easy to check that the inequality holds for $q^2 \in (0, 1/\mu)$, where $\mu = \frac{1+\sqrt{5}}{2}$ (the golden mean). Hence Q is in the Hadamard dual for all $q \in (0, \sqrt{1/\mu})$.

COROLLARY 5.3.7 *Let $A = [a_{ij}]$ be a 3-by-3 totally nonnegative matrix. Suppose $B = [b_{ij}]$ is the unsigned classical adjoint matrix. Then A is in the Hadamard dual if and only if*

1. $a_{11}b_{11} - a_{12}b_{12} \geq 0$, and
2. $a_{11}b_{11} - a_{21}b_{21} \geq 0$.

Even though Corollary 5.3.7 is just a recapitulation of Corollary 5.3.4, stated in this form aides in the proof of the following result on the inverse of nonsingular TN matrices in the Hadamard dual. Recall that if A is a nonsingular TN matrix, then $SA^{-1}S$ is a TN matrix, where $S = \text{diag}(1, -1, 1, -1, \dots, \pm 1)$.

THEOREM 5.3.8 *Suppose A is a 3-by-3 nonsingular TN matrix in the Hadamard dual. Then $SA^{-1}S$ is in the Hadamard dual.*

Proof. Observe that $SA^{-1}S$ is TN and, furthermore $SA^{-1}S = \frac{1}{\det A}B$, where $B = [b_{ij}]$ is the unsigned classical adjoint of A . Hence $SA^{-1}S$ is in the Hadamard dual if and only if B is a member of the Hadamard dual. Observe that the inequalities (1) and (2) in Corollary 5.3.7 are symmetric in A and B . Thus B is in the Hadamard dual. This completes the proof. ■

COROLLARY 5.3.9 *Let A be a 3-by-3 totally nonnegative matrix whose inverse is tridiagonal. Then A is in the Hadamard dual.*

Proof. Proof follows from Theorems 5.3.3 and 5.3.8. ■

Garloff and Wagner [GW1] noted that the set of all inverse tridiagonal totally nonnegative matrices is closed under Hadamard multiplication. The above result extends this fact in the 3-by-3 case. However, the general case is still unresolved, although we conjecture that the set of all inverse tridiagonal totally nonnegative matrices is contained in the Hadamard dual for all n .

For $3 \leq k \leq n$, let $W^{(k)} = (w_{ij}^{(k)})$ be the 3-by- n totally nonnegative matrix consisting of entries:

$$w_{ij}^{(k)} = \begin{cases} 0 & \text{if } i = 1, j \geq k, \\ 1 & \text{otherwise.} \end{cases}$$

For $1 \leq k \leq n-2$, let $U^{(k)} = (u_{ij}^{(k)})$ be the 3-by- n totally nonnegative matrix consisting of entries:

$$u_{ij}^{(k)} = \begin{cases} 0 & \text{if } i = 3, 1 \leq j \leq k, \\ 1 & \text{otherwise.} \end{cases}$$

For example, if $n = 5$ and $k = 3$, then $W^{(3)} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$, and

$$U^{(3)} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

THEOREM 5.3.10 *Let A be a 3-by- n ($n \geq 3$) totally nonnegative matrix. Then A is in the Hadamard dual if and only if $A \circ W^{(k)}$ is totally nonnegative for $3 \leq k \leq n$ and $A \circ U^{(j)}$ is totally nonnegative for $1 \leq j \leq n - 2$.*

Proof. The necessity is obvious, since $W^{(k)}$ and $U^{(j)}$ are both TN. Observe that it is enough to show that every 3-by-3 submatrix of A is in the Hadamard dual, by Proposition 5.1.5. Let B be any 3-by-3 submatrix of A . Consider the matrices $A \circ W^{(k)}$ and $A \circ U^{(j)}$. By hypothesis $A \circ W^{(k)}$ and $A \circ U^{(j)}$ are TN. Hence by considering appropriate submatrices, it follows that $B \circ W$ and $B \circ W^T$ are both TN. Therefore B is in the Hadamard dual by Theorem 5.3.3. Thus A is in the Hadamard dual. ■

Of course by transposition we may obtain a similar characterization of the n -by-3 Hadamard dual. At present no characterization of the Hadamard dual for 4-by-4 totally nonnegative matrices is known, but we offer the following ideas and conjectures on this issue.

CONJECTURE 5.3.11 *A 4-by-4 totally nonnegative matrix, A is in the Hadamard dual if and only if $A \circ V_4$, $A \circ V_4^T$, $A \circ V_5$, and $A \circ V_5^T$ are all totally nonnegative matrices, where*

$$V_4 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad V_5 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

After examination of the 3-by-3 and 3-by- n test matrices, a list of potential 4-by-4 test matrices can be proposed. This list includes the following six matrices as well as their transposes:

$$\begin{array}{ccc} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}. \end{array}$$

We refer to these matrices as V_1 thru V_6 , respectively. By observation, we are able to eliminate two of these matrices almost immediately, for $\det(A \circ V_2) \geq \det(A)$ and $\det(A \circ V_6) \geq \det(A)$.

Instead of using the remaining four matrices and their transposes, we first assume that all of the 3-by-3 submatrices of A (where we want to decide if A is in the Hadamard dual) have passed the 3-by-3 tests. Thus, assuming this information, the following inequalities are apparent:

$$\begin{array}{l} \det(A \circ V_1) \leq \det(A \circ V_2) \quad \text{and} \quad \det(A \circ V_3) \leq \det(A \circ V_2), \\ \det(A \circ V_1) \geq \det(A \circ V_4) \quad \text{and} \quad \det(A \circ V_2) \geq \det(A \circ V_4^T), \\ \det(A \circ V_1) \geq \det(A \circ V_3) \quad \text{and} \quad \det(A \circ V_2) \geq \det(A \circ V_3), \\ \det(A) \geq \det(A \circ V_1) \geq \det(A \circ V_3) \quad \text{and} \quad \det(A) \geq \det(A \circ V_2) \geq \det(A \circ V_3). \end{array}$$

Here we are assuming $\det A[\{2, 3\}] > 0$. Using the above inequalities implies that we may eliminate two more of these potential test matrices as well as their transposes. Thus only V_4 and V_5 remain. Unfortunately, we have been unable to determine other relevant inequalities relating these two matrices to each other, or to A .

Additionally, we would like to note that the methods employed in proving Theorem 5.3.3 have not proved fruitful for the 4-by-4 case. In conclusion, although the Hadamard dual is not completely described in general, considerable information concerning characteristics of the Hadamard dual members is known. Further research on this problem will hopefully lead to a complete description of the Hadamard dual, or perhaps more general qualities that members of the Hadamard dual must possess.

5.4 (0,+)-Patterns

In this section we consider $(0, +)$ -patterns (see also Chapter 2) of totally nonnegative matrices in the Hadamard dual. Recall that an m -by- n $(0, +)$ -sign pattern is an m -by- n array of symbols chosen from $\{+, 0\}$, and a realization of a sign pattern, S , is a real m -by- n matrix A such that:

$$a_{ij} > 0 \text{ when } s_{ij} = +; \text{ and } a_{ij} = 0 \text{ when } s_{ij} = 0.$$

As seen in chapter 2 there are two natural mathematical notions associated with sign-pattern problems. They are the notions of *require* and *allow*. We say an m -by- n sign pattern S *requires property P* if every realization of S has property P . On the other hand we say a sign pattern S *allows property P* if there exists a realization of

S with property P . We begin our analysis here by completely characterizing all the sign patterns S that require matrices in the Hadamard dual of the totally nonnegative matrices.

DEFINITION 5.4.1 Given an m -by- n sign pattern, S , we say that S *requires Hadamard duality of a TN matrix* if any totally nonnegative realization of S is in the Hadamard dual.

Observe that in order for a given sign pattern, S , to require Hadamard duality, it is necessary that S is of double echelon form described in Theorem 2.5.4

EXAMPLE 5.4.2 Any 1-by-1, 2-by-2 or 3-by-3 pattern other than any one of

$$\begin{bmatrix} + & + & + \\ + & + & + \\ + & + & + \end{bmatrix}, \begin{bmatrix} + & + & 0 \\ + & + & + \\ + & + & + \end{bmatrix}, \text{ or } \begin{bmatrix} + & + & + \\ + & + & + \\ 0 & + & + \end{bmatrix}$$

requires Hadamard duality of a TN matrix.

To verify this, first observe that by Examples 5.0.1 and 5.3.6 there exist matrices with the above sign patterns that are not in the Hadamard dual. Thus, suppose S is a 3-by-3 sign pattern different from the three patterns above. Then, S is a (with possibly more zeros) tridiagonal pattern, and hence S requires Hadamard duality of a TN matrix, by Theorem 5.2.4.

LEMMA 5.4.3 *Suppose A is an m -by- n totally nonnegative matrix, and let \mathcal{X} be any 1-by- n sign pattern. Then, there exists a realization, x of \mathcal{X} , such that $\begin{bmatrix} A \\ x \end{bmatrix}$ is*

totally nonnegative if and only if $\begin{bmatrix} A \\ \mathcal{X} \end{bmatrix}$ is of double echelon form.

Proof. The above condition is obviously necessary. Now suppose $\begin{bmatrix} A \\ \mathcal{X} \end{bmatrix}$ is of double echelon form. Assume without loss of generality that $0 \neq \mathcal{X}$ (because otherwise we are adding a zero row to a TN matrix, which preserves the property of being TN see chapter 2). Assume that \mathcal{X} is in following form: $\mathcal{X} = \begin{bmatrix} 0 & \dots & 0, & + & \dots & +, & 0 & \dots & 0 \end{bmatrix}$, where the plus signs span columns j to $j+k \leq n$. Observe that if $j+k < n$, then columns $j+k+1, \dots, n$ of A must be all zero columns since $\begin{bmatrix} A \\ \mathcal{X} \end{bmatrix}$ is in double echelon form. Thus, since zero columns do not affect total nonnegativity, it is enough to prove this lemma for the case $j+k = n$. Hence, $\mathcal{X} = \begin{bmatrix} 0 & \dots & 0, & + & \dots & + \end{bmatrix}$, where the first plus sign occurs in the j^{th} column. Let x be any realization of \mathcal{X} , and let $C = \begin{bmatrix} A \\ x \end{bmatrix}$. We will choose values for $x_i, i \geq j$ sequentially.

By Proposition 2.4.4, we may choose x_j positive and large enough so that

$$C \left[\{1, 2, \dots, m+1\} \mid \{1, \dots, j\} \right] = \begin{bmatrix} A[\{1, 2, \dots, m\} | \{1, \dots, j\}] \\ 0 \quad \dots \quad 0, \quad x_j \end{bmatrix}$$

is TN. Similarly, choose x_{j+1} so that $C \left[\{1, 2, \dots, m+1\} \mid \{1, \dots, j+1\} \right]$ is TN.

Continuing in this manner until we choose x_n such that $C = \begin{bmatrix} A \\ x \end{bmatrix}$ is TN. Observe

that at each stage, x_i ($i \geq j$) enters positively into each minor that includes x_i . Thus, there is no upper bound for the choice of x_i at each stage. This completes the proof.

■

THEOREM 5.4.4 *Let S be an n -by- n sign pattern that is in double echelon form. Then, S requires Hadamard duality of a TN matrix if and only if S does not contain any one of the following sign patterns as a subpattern:*

$$\begin{bmatrix} + & + & + \\ + & + & + \\ + & + & + \end{bmatrix}, \begin{bmatrix} + & + & 0 \\ + & + & + \\ + & + & + \end{bmatrix}, \text{ or } \begin{bmatrix} + & + & + \\ + & + & + \\ 0 & + & + \end{bmatrix}.$$

Proof. Suppose S is in double echelon form for which

$$\begin{bmatrix} + & + & 0 \\ + & + & + \\ + & + & + \end{bmatrix}$$

is a subpattern. The analysis is similar for the other two patterns. First observe that we may assume this subpattern occurs as a contiguous pattern, since S is in double echelon form. Suppose this 3-by-3 subpattern is indexed by rows $j, j+1, j+2$ of S . Let B be a 3-by-3 totally nonnegative matrix with sign pattern $\begin{bmatrix} + & + & 0 \\ + & + & + \\ + & + & + \end{bmatrix}$ that is not in the Hadamard dual (Recall Example 5.0.1). Extend B to an 3-by- n TN matrix \bar{B} by any one of the following operations depending on the sign pattern in rows $j, j+1, j+2$ of S :

1. By repeating columns,

2. By adding columns of the form $\begin{bmatrix} 0 & 0 & \delta \end{bmatrix}^T$ or $\begin{bmatrix} \delta & 0 & 0 \end{bmatrix}^T$, $\delta > 0$ or

3. By invoking Lemma 5.4.3 applied to the transpose of the matrix constructed thus far.

Now, by Lemma 5.4.3 we may construct an n -by- n TN matrix \tilde{B} with sign pattern S . However, \tilde{B} is not in the Hadamard dual since \tilde{B} contains a submatrix that is not in the Hadamard dual (see Proposition 5.1.5).

Suppose S is of double echelon form and does not contain

$$\begin{bmatrix} + & + & + \\ + & + & + \\ + & + & + \end{bmatrix}, \quad \begin{bmatrix} + & + & 0 \\ + & + & + \\ + & + & + \end{bmatrix}, \quad \text{or} \quad \begin{bmatrix} + & + & + \\ + & + & + \\ 0 & + & + \end{bmatrix}$$

as a subpattern. We proceed by using induction on n . This claim has already been verified for $n \leq 3$, so assume the result is true for all such patterns of size less than or equal to $n - 1$. Let S be as assumed above. Observe that, by induction, any TN realization of S has all of its proper submatrices in the Hadamard dual. Thus, we only need to check the $\det(A \circ B)$, where A is any realization of S and B is TN. We consider three cases:

Case 1: Suppose the i^{th} diagonal entry of S is zero. Then, S contains a zero block of size $n - i + 1 + i = n + 1$. Hence, $A \circ B$ has a zero block of size $n + 1$ for any realization A of S . But, in this case, $\det(A \circ B) = 0$ (see [Ry]). Thus, A is in the Hadamard dual.

Case 2: Suppose S has positive main diagonal entries, but that some entry on the superdiagonal is zero (similar arguments hold if an entry on the subdiagonal is zero). Assume the $(i, i + 1)^{\text{st}}$ entry of S is zero. Since S has positive main

diagonal entries in addition to being in double echelon form, it follows that S contains a block of zeros of size $n - i + i = n$. Hence, S is block triangular, and by induction, we have $\det(A \circ B) \geq 0$, for any realization A of S and B is TN

Case 3: Finally, suppose S has positive main super and subdiagonal entries. Since S does not contain any of the three subpatterns (by assumption), it follows that the $(i, i + 2)$ and $(i + 2, i)$ entries of S must be zero for $i = 1, \dots, n - 2$. Since S is in double echelon form, it follows that S is a tridiagonal pattern. Thus, any realization A of S is in the Hadamard dual by Theorem 5.2.4. ■

Note that if A is m -by- n with $n \geq m$ (without loss of generality), then A is in the Hadamard dual if and only if every m -by- m submatrix of A is in the Hadamard dual. This follows from Proposition 5.1.5. The above remark combined with Theorem 5.4.4 gives rise to the following corollary:

COROLLARY 5.4.5 *Let S be any m -by- n sign pattern that is in double echelon form. Then, S requires Hadamard duality of a TN matrix if and only if S does not contain*

$$\begin{bmatrix} + & + & + \\ + & + & + \\ + & + & + \end{bmatrix}, \quad \begin{bmatrix} + & + & 0 \\ + & + & + \\ + & + & + \end{bmatrix}, \quad \text{or} \quad \begin{bmatrix} + & + & + \\ + & + & + \\ 0 & + & + \end{bmatrix}$$

as a subpattern.

The corresponding “allow” question seems much more difficult. One reason for this is that we do not have a general description of the matrices in the Hadamard dual for larger sizes. However, at the very least we can say the following.

THEOREM 5.4.6 *Let S be any 3-by- n $(0, +)$ double echelon pattern. Then S allows a totally nonnegative matrix in the Hadamard dual.*

Proof. It is not difficult (given Corollary 5.3.4) to show that any 3-by-3 $(0, +)$ double echelon pattern allows a TN matrix in the Hadamard dual. Suppose by induction that any 3-by- $(n - 1)$ $(0, +)$ double echelon pattern allows a TN matrix in the Hadamard dual. Let S be any 3-by- n $(0, +)$ double echelon pattern. Let B be a realization of the first $n - 1$ columns of S that is in the Hadamard dual. Such a B exists by the induction hypothesis. Let $x = (x_1, x_2, x_3)^T$ be a column vector with the same $(0, +)$ pattern as the n^{th} column of S . If $x = 0$, then we are done. So suppose $x \neq 0$. It is clear that if x_1 is a fixed nonnegative number, then we may choose x_2 large enough (if $x_2 = 0$, then $x_1 = 0$ by the assumption that S is double echelon) so that the first two rows of $[B|x]$ are in the Hadamard dual. Observe that x_3 lies in the $(3,3)$ entry of any submatrix including row 3 and column n . Since B is TN x_3 enters positively into the determinant of any such 3-by-3 submatrix. Thus x_3 can be chosen large enough so that all of the conditions of Theorem 5.3.10 (which are really just determinantal conditions) hold and so that $[B|x]$ is TN. This completes the proof. ■

The corresponding “allow” question for larger sizes is currently unresolved, however, we conjecture that any m -by- n $(0, +)$ double echelon sign pattern allows a TN matrix in the Hadamard dual. Certainly, the full pattern does, since J is in the Hadamard dual.

5.5 Oppenheim's Inequality

Suppose A and B are two n -by- n positive semidefinite matrices. Then by a classical result of Schur (see [HJ1]), $A \circ B$ is again positive semidefinite. Therefore, in particular, $\det(A \circ B) \geq 0$ in this case. However, even more is true here. Oppenheim proved that if A and B are positive semidefinite, then $\det(A \circ B) \geq \det B \prod_{i=1}^n a_{ii}$ (see [HJ1]).

For the case in which A and B are n -by- n totally nonnegative matrices it is certainly not true that $\det(A \circ B) \geq 0$ (see Example 5.0.1). However, if A is totally nonnegative and in the Hadamard dual, then $A \circ B$ is totally nonnegative (whenever B is TN) and $\det(A \circ B) \geq 0$. Furthermore, Oppenheim's inequality holds in this case.

THEOREM 5.5.1 *Let A be an n -by- n totally nonnegative matrix in the Hadamard dual, and suppose B is any n -by- n totally nonnegative matrix. Then*

$$\det(A \circ B) \geq \det B \prod_{i=1}^n a_{ii}.$$

Proof. If B is singular, then there is nothing to show, since $\det(A \circ B) \geq 0$, as A is in the Hadamard dual. Assume B is nonsingular. If $n = 1$, then the inequality is trivial. Suppose, by induction, that Oppenheim's inequality holds for all $(n - 1)$ -by- $(n - 1)$ TN matrices A and B with A in the Hadamard dual. Suppose A and B

are n -by- n TN matrices and assume that A is in the Hadamard dual. Let A_{11} (B_{11}) denote the principal submatrix obtained from A (B) by deleting row and column 1. Then by induction $\det(A_{11} \circ B_{11}) \geq \det B_{11} \prod_{i=2}^n a_{ii}$. Since B is nonsingular, by Fischer's inequality B_{11} is nonsingular. Consider the matrix $\tilde{B} = B - xE_{11}$, where $x = \frac{\det B}{\det B_{11}}$ and E_{11} is the (1,1) standard basis matrix. Then $\det \tilde{B} = 0$, and \tilde{B} is TN (see Lemma 2.6.2). Therefore $A \circ \tilde{B}$ is TN and $\det(A \circ \tilde{B}) \geq 0$. Observe that $\det(A \circ \tilde{B}) = \det(A \circ B) - xa_{11}\det(A_{11} \circ B_{11}) \geq 0$. Thus

$$\begin{aligned} \det(A \circ B) &\geq xa_{11}\det(A_{11} \circ B_{11}) \\ &\geq xa_{11}\det B_{11} \prod_{i=2}^n a_{ii} \\ &= \det B \prod_{i=1}^n a_{ii}, \end{aligned}$$

as desired. ■

We close this section with some further remarks concerning Oppenheim's inequality. In the case when A and B are n -by- n positive semidefinite matrices it is known that the following inequality holds: $\det(A \circ B) \geq \max \left\{ \det B \prod_{i=1}^n a_{ii}, \det A \prod_{i=1}^n b_{ii} \right\}$. However, in the case when A is in the Hadamard dual and B is an n -by- n TN matrix it is not true in general that $\det(A \circ B) \geq \det A \prod_{i=1}^n b_{ii}$. Consider the following example.

EXAMPLE 5.5.2 Let A be any 3-by-3 totally positive matrix in the Hadamard dual, and let $B = W$, the 3-by-3 totally nonnegative matrix equal to $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.

Then since the (1,3) entry of A enters positively into $\det A$ it follows that

$$\det(A \circ B) < \det A = \det A \prod_{i=1}^3 b_{ii}.$$

The next remark settles the issue of the possibility that Oppenheim's inequality offers a characterization of all TN matrices in the Hadamard dual, namely, if a given TN matrix A satisfies Oppenheim's inequality, then A is in the Hadamard dual.

Note that this is true for $n \leq 3$. For $n = 4$ consider the following matrix. Let

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}. \text{ Suppose } B \text{ is any 4-by-4 TN matrix. Then since the (1,4)}$$

enters negatively into $\det B$ it follows that $\det(A \circ B) \geq \det B = \det B \prod_{i=1}^4 a_{ii}$. Hence A satisfies Oppenheim's inequality, but A is not in the Hadamard dual since A contains a submatrix ($A[\{1, 2, 3\}|\{2, 3, 4\}]$) that is not in the dual.

Bibliography

- [A] T. Ando, Totally Positive Matrices, *Linear Algebra Appl.* **90**:165-219 (1987).
- [As] B.A. Asner, On the Total Nonnegativity of the Hurwitz Matrix, *SIAM J. Appl. Math.* **18** No. 2:407-414 (1970).
- [BJ1] W. W. Barrett and C. R. Johnson, Possible Spectra of Totally Positive Matrices, *Linear Algebra Appl.* **62**:231-233 (1984).
- [BJ2] W. W. Barrett and C. R. Johnson, Determinantal Inequalities for Positive-Definite Matrices, *Discrete Math.* **115**:97-106 (1993).
- [BFZ] A. Berenstein, S. Fomin and A. Zelevinsky, Parameterizations of Canonical Bases and Totally Positive Matrices, *Advances in Math.*, **122**:49-149 (1996).
- [BoP] C. de Boor and A. Pinkus, The Approximation of a Totally Positive Band Matrix by a Strictly Totally Positive One, *Linear Algebra Appl.* **42**:81-98 (1982).
- [B] F. Brenti, Combinatorics and Total Positivity, *J. Comb. Theory, Series A*, **71**:175-218 (1995).

- [Ca] D. Carlson, Weakly, Sign-Symmetric Matrices and Some Determinantal Inequalities, *Colloq. Math.* **17**:123-129 (1967).
- [CP1] J.M. Carnicer and J.M. Peña, Totally Positive Bases for Shape Preserving Curve Design and Optimality of B-Splines, *Comput. Aided Geom. Design* **11**:633-654 (1994).
- [CP2] J.M. Carnicer and J.M. Peña, Bidiagonalization of Oscillatory Matrices, *Linear and Multilinear Algebra* **42**:365-376 (1997).
- [C1] C.W. Cryer, The LU -factorization of Totally Positive Matrices, *Linear Algebra Appl.* **7**:83-92 (1973).
- [C2] C.W. Cryer, Some Properties of Totally Positive Matrices, *Linear Algebra Appl.* **15**:1-25 (1976).
- [Ev] S.P. Eveson, The Eigenvalue Distribution of Oscillatory and Strictly Sign-Regular Matrices, *Linear Algebra Appl.* **246**:17-21 (1996).
- [FGG] S.M. Fallat, M.I. Gekhtman and C. Goldbeck, Ratios of principal minors of totally nonnegative matrices, An unpublished paper from an National Science Foundation Research Experiences for Undergraduates program held at the College of William and Mary in the summer of 1998.
- [FHJ] S.M. Fallat, H.T. Hall and C.R. Johnson, Characterization of Product Inequalities for Principal Minors of M -matrices and Inverse M -matrices, *Quart. J. Math.* **49** No. 196: 451-458 (1998).

- [FJ1] S.M. Fallat and C.R. Johnson, Sub-direct Sums and Positivity Classes of Matrices, *Linear Algebra Appl.* **288**:149-173 (1999).
- [FJ2] S.M. Fallat and C.R. Johnson, Olga, Matrix Theory and the Tausky Unification Problem, *Linear Algebra Appl.* **280**:39-49 (1998).
- [FJS1] S.M. Fallat, C.R. Johnson and R.L. Smith, Inserting Rows or Columns and Preserving Total Positivity, manuscript.
- [FJS2] S.M. Fallat, C.R. Johnson and R.L. Smith, Certain Totally Positive Matrix Completion Problems, manuscript.
- [F] M. Fekete, Uber ein Problem von Laguerre, *Rend. Conti. Palermo* **34**:89-100, 110-120 (1913).
- [FZ] S. Fomin and A. Zelevinsky, Double Bruhat Cells and Total Positivity, *Journal of the AMS* **12**:335-380 (1999).
- [Fr] S. Friedland, Weak Interlacing Properties of Totally Positive Matrices, *Linear Algebra Appl.* **71**:247-266 (1985).
- [Fu] K. Fukuda, *cdd version 0.61 and cdd+ version 0.75*, Institute for Operations Research, Zurich, Switzerland (1997). Program and synopsis is available from the anonymous ftp site:

ftp.ifor.math.ethz.ch in the directory pub/fukuda/cdd .
- [GK1] F.R. Gantmacher and M.G. Krein, Sur les Matrices Complement Non-negatives et Oscillatoires, *Comp. Math.* **4**:445-476 (1937).

- [GK2] F.R. Gantmacher and M.G. Krein, *Oszillationsmatrizen, Oszillationskerne und kleine Schwingungen Mechanischer Systeme*, Akademie-Verlag, Berlin, 1960.
- [G1] J. Garloff, Majorization Between the Diagonal Elements and the Eigenvalues of an Oscillating Matrix, *Linear Algebra Appl.* **47**:181-184 (1982).
- [G2] J. Garloff, An Inverse Eigenvalue Problem for Totally Nonnegative Matrices, *Linear and Multilinear Algebra* **17**:19-23 (1985).
- [G3] J. Garloff, Vertex Implications for Totally Nonnegative Matrices. *Total Positivity and its Applications*, Mathematics and its Applications, Vol. 359 (M. Gasca and C.A. Micchelli, Eds.), Kluwer Academic, 103-107 1996.
- [GW1] J. Garloff and D.G. Wagner, Hadamard Products of Stable Polynomials are Stable, *J. Math. Anal. Appl.* **202**:797-809 (1996).
- [GW2] J. Garloff and D.G. Wagner, Preservation of Total Nonnegativity under Hadamard Products and Related Topics, *Total Positivity and its Applications*, Mathematics and its Applications, Vol. 359 (M. Gasca and C.A. Micchelli, Eds.), Kluwer Academic, 97-102 1996.
- [GM] M. Gasca and C.A. Micchelli, *Total Positivity and its Applications*, Mathematics and its Applications, Vol. 359, Kluwer Academic 1996.
- [GP1] M. Gasca and J.M. Peña, Total Positivity and Neville Elimination, *Linear Algebra Appl.* **165**:25-44 (1992).

- [GP2] M. Gasca and J.M. Peña, Total Positivity, QR Factorization, and Neville Elimination, *SIAM J. Matrix Anal. Appl.* **14**:1132-1140 (1993).
- [GP3] M. Gasca and J.M. Peña, A Matricial Description of Neville Elimination with Applications to Total Positivity, *Linear Algebra Appl.* **202**:33-53 (1994).
- [GP4] M. Gasca and J.M. Peña, On Factorizations of Totally Positive Matrices. *Total Positivity and its Applications*, Mathematics and its Applications, Vol. 359 (M. Gasca and C.A. Micchelli, Eds.), Kluwer Academic, 109-130 1996.
- [Go] G.S. Goodman, A Probabilistic Representation of Totally Positive Matrices, *Advances Appl. Math.* **7**:236-252 (1986).
- [He] B. Heiligers, Totally Nonnegative Moment Matrices, *Linear Algebra Appl.* **199**:213-227 (1994).
- [HJ1] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, New York, 1985.
- [HJ2] R.A. Horn and C.R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, New York, 1991.
- [J] C.R. Johnson, Closure Properties of Certain Positivity Classes of Matrices under Various Algebraic Operations *Linear Algebra Appl.* **97**:243-247 (1987).
- [K1] S. Karlin, Oscillation Properties of Eigenvectors of Strictly Totally Positive Matrices, *J. Analyse Math.* **14**:247-266 (1965).

- [K2] S. Karlin, *Total Positivity*, Vol. I, Stanford University Press, Stanford, 1968.
- [Ko1] D. M. Kotljanskiĭ, The Theory of Nonnegative and Oscillating Matrices (Russian) *Ukrain. Mat. Ž.* **2**:94-101 (1950); English transl.: *Translations of the AMS*, Series 2 **27**:1-8 (1963).
- [Ko2] D. M. Kotljanskiĭ, A Property of Sign-Symmetric Matrices (Russian), *Mat. Nauk (N.S.)* **8**:163-167 (1953); English transl.: *Translations of the AMS*, Series 2 **27**:19-24 (1963).
- [L] C. Loewner, On Totally Positive Matrices, *Math. Z.* **63**:338-340 (1955).
- [M1] T.L. Markham, On Oscillatory Matrices, *Linear Algebra Appl.* **3** No. 2:143-156 (1970).
- [M2] T.L. Markham, A Semigroup of Totally Nonnegative Matrices, *Linear Algebra Appl.* **3** No. 2:157-164 (1970).
- [P1] J.M. Peña, On the Schur and Singular Decompositions of Oscillatory Matrices, *Linear Algebra Appl.* **261**:307-315 (1997).
- [P2] J.M. Peña, On the Relationship Between Graphs and Totally Positive Matrices, *SIAM J. Matrix Anal. Appl.* **19** No. 2:369-377 (1997).
- [Pi1] A. Pinkus, An Interlacing Property of Eigenvalues of Strictly Totally Positive Matrices, *Linear Algebra Appl.* **279**:201-206 (1998).

- [Pr] H.S. Price, Monotone and Oscillatory Matrices Applied to Finite Difference Approximations, *Math. Comp.* **22**:489-516 (1968).
- [R] C.E. Radke, Classes of Matrices with Distinct, Real Characteristic Values, *SIAM J. Appl. Math.* **16**:1192-1207 (1968).
- [Ry] H.J. Ryser, *Combinatorial Mathematics*, The Mathematical association of America, Rahway, 1963.
- [RH] J.W. Rainey and G.J. Halbetler, Tridiagonalization of Completely Nonnegative Matrices, *Math. Comp.* **26**:121-128 (1972).
- [Sc] I.J. Schoenberg, Uber Variationsvermindernde Lineare Transformationen, *Math. Z.* **32**:321-328 (1930).
- [SS] B.Z. Shapiro and M.Z. Shapiro, On the Boundary of Totally Positive Upper Triangular Matrices, *Linear Algebra Appl.* **231**:105-109 (1995).
- [So] A.V. Sobolev, Totally Positive Operators, *Siberian Math. J.* **16**:636-641 (1975).
- [St] B. Sturmfels, Totally Positive Matrices and Cyclic Polytopes, *Linear Algebra Appl.* **107**:275-281 (1988).
- [Wa] D.G. Wagner, Total Positivity of Hadamard Products, *J. Math. Anal. Appl.* **163**:459-483 (1992).

- [W] A. Whitney, A Reduction Theorem for Totally Positive Matrices, *J. Analyse Math.* **2**:88-92 (1952).

Vita

Shaun Michael Fallat

Born in Fraser Lake, British Columbia, Canada on November 27, 1972. Valedictorian of Fraser Lake Elementary Secondary School, 1990. Received a Bachelor's of Science Degree (with honours) in Statistics from the University of Victoria in May 1994 and a Masters of Science Degree from the University of Victoria in June 1996. Was awarded the Lieutenant Governor's Silver Medal, the highest possible honour for a Masters student at the University of Victoria. Entered the College of William and Mary Applied Science program as a research assistant in August of 1996 to begin work on a Degree of Doctor of Philosophy.