

On Factorizations of Totally Positive Matrices

M. Gasca and J. M. Peña

Abstract. Different approaches to the decomposition of a nonsingular totally positive matrix as a product of bidiagonal matrices are studied. Special attention is paid to the interpretation of the factorization in terms of the Neville elimination process of the matrix and in terms of corner cutting algorithms of Computer Aided Geometric Design. Conditions of uniqueness for the decomposition are also given.

§1. Introduction

Totally positive matrices (TP matrices in the sequel) are real, nonnegative matrices whose all minors are nonnegative. They have a long history and many applications (see the paper by Allan Pinkus in this volume for the early history and motivations) and have been studied mainly by researchers of those applications. In spite of their interesting algebraic properties, they have not yet received much attention from linear algebraists, including those working specifically on nonnegative matrices. One of the aims of the masterful survey [1] by T. Ando, which presents a very complete list of results on TP matrices until 1986, was to attract this attention.

In some papers by M. Gasca and G. Mühlbach ([13] for example) on the connection between interpolation formulas and elimination techniques it became clear that what they called *Neville elimination* had special interest for TP matrices. It is a procedure to make zeros in a column of a matrix by adding to each row an appropriate multiple of the precedent one and had been already used in some of the first papers on TP matrices [33]. However, in the last years, we have developed a better knowledge of the properties of Neville elimination which has allowed us to improve many previous results on those matrices [14–22]. In this paper we shall use this elimination technique to get the factorization of a nonsingular TP matrix as a product of bidiagonal matrices. This provides a useful representation of such matrices which allows us to identify some important subclasses, as for example that of *strictly totally positive matrices* (that is, TP matrices whose minors are all positive). Under some conditions on the zero pattern of the bidiagonal matrices that representation is unique.

A direct consequence of the well-known Cauchy-Binet identity for determinants is that the product of TP matrices is again a TP matrix. Consequently, one of the topics in the literature of TP matrices has been their decomposition as products of simpler TP matrices. In particular, in view of applications, the most interesting factorization seems to be in terms of bidiagonal nonnegative matrices which, obviously, are always TP matrices. Let us give a brief overview of some of the different approximations to this question.

Square TP matrices of order n form a multiplicative semigroup S_n , and the nonsingular matrices of S_n form a semigroup s_n of the group of all real nonsingular square matrices of order n . In [31], Loewner used some notions from the theory of Lie groups which we briefly recall for the study of S_n and s_n . If $U(t)$ is a differentiable matrix function of the real parameter t in an interval $[0, t_0]$, representing for each t an element of s_n , and $U(0)$ is the identity matrix I (which belongs to s_n), then the matrix $(\frac{dU(t)}{dt})_{t=0}$ is called an *infinitesimal element* of s_n . The first task in [31] was to prove that the set σ_n of all infinitesimal elements of s_n consists of the Jacobi (i.e., tridiagonal) matrices with nonnegative off-diagonal elements.

As in Lie-group theory, it can be shown that if $\Omega(t)$ ($0 \leq t \leq t_0$) is any one-parameter family of elements of σ_n which is piecewise continuous in t , the differential equation

$$\frac{dU(t)}{dt} = \Omega(t)U(t) \quad (1.1)$$

has a unique continuous solution $U(t)$ in s_n satisfying $U(0) = I$. In this case we say that $U(t_0)$ is generated by the infinitesimal elements $\Omega(t)$ ($0 \leq t \leq t_0$). In general, a semigroup cannot be completely generated by its infinitesimal elements. However, Loewner proved in [31] that this is not the case for s_n . He used the following reformulation of a result due to Whitney [33].

Let E_{ij} ($1 \leq i, j \leq \dots, n$) be the $n \times n$ matrix with all elements zero with the exception of a one at the place (i, j) and denote $F_{ij}(\omega) = I + \omega E_{ij}$. Then every nonsingular TP matrix U can be written as a product

$$U = U_1 U_2 \cdots U_{n-1} D V_1 V_2 \cdots V_{n-1}, \quad (1.2)$$

where, for $i = 1, 2, \dots, n-1$,

$$U_i = F_{n,n-1}(\omega_{n,n-1}^i) F_{n-1,n-2}(\omega_{n-1,n-2}^i) \cdots F_{i+1,i}(\omega_{i+1,i}^i), \quad (1.3)$$

$$V_i = F_{n-i,n-i+1}(\omega_{n-i,n-i+1}^i) F_{n-i+1,n-i+2}(\omega_{n-i+1,n-i+2}^i) \cdots F_{n-1,n}(\omega_{n-1,n}^i), \quad (1.4)$$

with all the ω -s nonnegative, and D represents a diagonal matrix with positive diagonal elements.

Observe that the matrices U_i and V_i are products of bidiagonal elementary totally positive matrices but neither U_i nor V_i are bidiagonal: (1.2) uses $n(n-1)$ bidiagonal factors.

The conclusion of [31] is that, by using infinitesimal generators, any non-singular TP matrix of order n can be generated from the identity by the solutions of the differential equation (1.1).

In 1979 Frydman and Singer ([12], Theorem 1) showed that the class of transition matrices for the finite state time-inhomogeneous birth and death processes coincides with the class of nonsingular TP stochastic matrices. This result was based upon a factorization of nonsingular TP stochastic matrices P in terms of bidiagonal matrices ([12], Theorem 1') similar to (1.2) without the diagonal matrix D :

$$P = U_1 U_2 \cdots U_{n-1} V_1 V_2 \cdots V_{n-1}, \quad (1.5)$$

and with the elementary matrices F scaled to be stochastic. As in (1.2) the matrices U_i, V_i are not bidiagonal and (1.5) contains $n(n-1)$ bidiagonal factors. The fact that those transition matrices for birth and death processes are totally positive had been pointed out in 1959 by Karlin and Mc Gregor [27, 28] with probabilistic arguments. All these results have been surveyed in 1986 by G. Goodman [23], who extended them to compound matrices, that is matrices whose elements are the values of the minors of a certain order m of a given matrix A .

In [9], Remark 4.1, Cryer pointed out that a matrix A is TP iff it can be written in the form

$$A = \prod_{r=1}^N L_r \prod_{s=1}^M U_s$$

where each L_r (resp. U_s) is a TP bidiagonal lower (upper) triangular matrix. In that remark the author did not give any relation between N, M , and the order n of A

On the other hand, factorizations of TP matrices as product of bidiagonal matrices are important in Computer Aided Geometric Design and, in particular, in corner cutting algorithms. In [24], Goodman and Micchelli showed that the existence of a corner cutting algorithm transforming a control polygon of a curve into another one with the same number of vertices was equivalent to the fact that both polygons were related by a nonsingular stochastic matrix. The key tool to obtain this result was again (see [24], Theorem 1) the characterization of a nonsingular TP stochastic matrix of order n as the product of $n-1$ bidiagonal lower triangular stochastic matrices by other $n-1$ matrices which are bidiagonal upper triangular, with a zero pattern which will be precised in Sections 3 and 4. Observe that in this case the factorization is formed by $2n-2$ bidiagonal matrices and compare with (1.5). What has happened is that the set of all elementary matrices which appeared in (1.5), by replacing the factors U_i, V_i by their corresponding decompositions (1.3),(1.4), has been reordered, as we shall see in Section 3, to give rise to a short number of bidiagonal (in general nonelementary) matrices. In [32], Theorem 3.1, Micchelli and Pinkus obtained a factorization theorem for rectangular TP matrices as a

product of bidiagonal matrices in order to extend the previous interpretation to general corner cutting algorithms. For more details related with this matter and for concrete factorizations associated with corner cutting algorithms to the Bezier polygon, see [25],[26], [5] and [6]. The use of Neville elimination was crucial to prove the results on optimality obtained in these last two papers. See also the paper by Carnicer and Peña on optimal bases in this volume.

As we shall see in the following sections, in [19] we proved the uniqueness of factorizations of the type just mentioned above for nonsingular TP matrices, under certain conditions on the zero pattern of the bidiagonal factors. In [20], we interpreted these last results in terms of corner cutting algorithms.

Finally, factorizations of biinfinite TP matrices in terms of bidiagonal matrices have also important applications. Motivated mainly by some problems of Approximation Theory, Cavaretta et al. proved in [7], Theorem 1, the existence of such factorization for biinfinite strictly m -banded TP matrices. In [4], Theorem B, de Boor and Pinkus proved that every finite nonsingular TP m -banded matrix is the product of m TP bidiagonal matrices and, in Theorem C of the same paper, they extended this result to biinfinite m -banded TP matrices with linearly independent rows and columns. Recently, Dahmen and Micchelli [10] have obtained factorization results in terms of bidiagonal matrices for biinfinite TP matrices with a special structure of zeros (which they called 2-slanted matrices) and they applied these results to multi-resolution analysis and wavelet theory. This study has been extended in [21] to p -slanted matrices ($p \geq 2$), and some applications of these results to spline interpolation and locally finite decomposition of spline spaces have been also provided.

This paper is organized in the following way. In Section 2 we describe carefully Neville elimination for nonsingular matrices and study the properties of the bidiagonal elementary matrices associated to it. We pay special attention to the nonsingular matrices which can be transformed into diagonal form by Neville elimination without permutations of rows or columns. This property is referred to as the WRC condition. Section 3 is devoted to the decomposition of matrices which satisfy the WRC condition as products of bidiagonal matrices with a prescribed zero pattern in the places $(i, i - 1)$ for the bidiagonal lower triangular matrices and in the places $(i, i + 1)$ for the bidiagonal upper triangular ones. Since any nonsingular TP matrix A satisfies the WRC condition, in Section 4 we apply to these matrices the results of the precedent section and, depending on the choice of the zero pattern of the bidiagonal matrices, we arrive to form 16 different factorizations (some of them coincident for some special classes of TP matrices). Each of these decomposition is unique under the prescribed conditions and has a different interpretation with respect to the Neville elimination process of the matrix A . At our knowledge, the uniqueness of the different factorizations, which is a consequence of the uniqueness of the elimination process, is a novelty in this type of results. In the last section we relate the previously studied factorizations of A with other characterizations of nonsingular TP matrices. In

particular we discuss determinantal characterizations, that is necessary and sufficient conditions on the signs of a short number of minors of a matrix in order to classify it as a TP matrix or as a matrix belonging to some special classes of TP matrices. We also mention some of the characterizations of a nonsingular TP matrix by its LU decomposition or, as we recently saw in [16], by its QR decomposition.

§2. Neville elimination

The following notation will be convenient. For $k, n \in \mathbb{N}$, $1 \leq k \leq n$, $Q_{k,n}$ will denote the set of all increasing sequences of k natural numbers not greater than n . For $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$, $\beta = (\beta_1, \beta_2, \dots, \beta_k) \in Q_{k,n}$ and A an $n \times n$ real matrix, we denote by $A[\alpha|\beta]$ the $k \times k$ submatrix of A containing rows $\alpha_1, \dots, \alpha_k$ and columns β_1, \dots, β_k of A . $Q_{k,n}^0$ will denote the set of increasing sequences of k consecutive natural numbers not greater than n .

Neville elimination was precisely described for general finite matrices in [14]. Here, following [22], we restrict ourselves to the case of nonsingular matrices. For a nonsingular matrix A of order n this elimination procedure consists of $n - 1$ successive steps, resulting in a sequence of matrices as follows:

$$A = \tilde{A}^{(1)} \rightarrow A^{(1)} \rightarrow \tilde{A}^{(2)} \rightarrow A^{(2)} \dots \rightarrow \tilde{A}^{(n)} = A^{(n)} = U,$$

where U is an upper triangular matrix. For each t , $1 \leq t \leq n$, the matrix $A^{(t)} = (a_{ij}^{(t)})_{1 \leq i, j \leq n}$ has zeros below its main diagonal in the first $t - 1$ columns, and also one has

$$a_{it}^{(t)} = 0, \quad i \geq t \quad \Rightarrow \quad a_{ht}^{(t)} = 0 \quad \forall h \geq i. \quad (2.1)$$

$A^{(t)}$ is obtained from the matrix $\tilde{A}^{(t)}$ by moving to the bottom the rows with a zero entry in the column t , if necessary, in order to get (2.1). The rows are moved and placed with the same relative order as they appear in $\tilde{A}^{(t)}$. To get $\tilde{A}^{(t+1)}$ from $A^{(t)}$ we produce zeros in the column t below the main diagonal by subtracting a multiple of the i th row to the $(i + 1)$ th for $i = n - 1, n - 2, \dots, t$, according to the formula

$$\tilde{a}_{ij}^{(t+1)} = \begin{cases} a_{ij}^{(t)} & \text{if } i \leq t \\ a_{ij}^{(t)} - \left(a_{it}^{(t)} / a_{i-1,t}^{(t)} \right) a_{i-1,j}^{(t)} & \text{if } i \geq t + 1 \text{ and } a_{i-1,t}^{(t)} \neq 0 \\ a_{ij}^{(t)} & \text{if } i \geq t + 1 \text{ and } a_{i-1,t}^{(t)} = 0. \end{cases} \quad (2.2)$$

Observe that in the third case $a_{i-1,t}^{(t)} = 0$ implies that $a_{it}^{(t)} = 0$. In this process one has $\tilde{A}^{(n)} = A^{(n)} = U$, and when no row exchanges are needed, then $\tilde{A}^{(t)} = A^{(t)}$ for all t . This happens, for example, when

$$\det A[\alpha|1, 2, \dots, t] \neq 0, \quad 1 \leq t \leq n, \quad \forall \alpha \in Q_{t,n}^0 \quad (2.3)$$

Theorem 2.1. *A nonsingular $n \times n$ matrix A satisfies the WR condition if and only if it can be factorized in the form (2.13) with the $m_{i,j}$'s satisfying (2.6). If A satisfies that condition, the factorization is unique and m_{ij} is the (i, j) multiplier of the Neville elimination of A .*

When no row exchanges are needed in the Neville elimination of A and U^T we say that the process is possible without row or column exchanges and, for brevity, that A satisfies the WRC condition. In this case, accordingly to Theorem 2.1, A can be written in the form

$$\begin{aligned} A = & (E_n(m_{n1})E_{n-1}(m_{n-1,1}) \cdots E_2(m_{21})) \cdot (E_n(m_{n2}) \cdots E_3(m_{32})) \cdots \\ & \cdot (E_n(m_{n,n-1}))D(E_n^T(m'_{n,n-1}))(E_{n-1}^T(m'_{n-1,n-2})E_n^T(m'_{n,n-2})) \cdots \\ & \cdot (E_2^T(m'_{21}) \cdots E_{n-1}^T(m'_{n-1,1})E_n^T(m'_{n1})) \end{aligned} \quad (2.14)$$

with m_{ij} and m'_{ij} satisfying (2.6) and D a diagonal matrix. Moreover this representation is unique, and m_{ij} (respectively m'_{ij}) is the (i, j) multiplier of the Neville elimination of A (resp. U^T). Here E_i^T denotes the transpose of E_i .

By transposition of the right-hand side of (2.14) we deduce that A^T satisfies the WRC condition too and the multipliers of the Neville elimination of A^T are those of U^T . Therefore, if we denote

$$m_{ij} = m'_{ji} \quad \text{for all } i > j, \quad (2.15)$$

we can say that, for any $(i, j), i \neq j$, m_{ij} is the (i, j) multiplier of the *complete* Neville elimination of A , in the sense that, if $i > j$, it is the multiplier of the Neville elimination of A and, if $i < j$, it is the (j, i) multiplier of the Neville elimination of A^T (or, equivalently, of U^T). On the contrary, what we have called *pivots* are different, in general, for the Neville elimination of U^T and A^T . However, due to the fact that U^T is lower triangular, the (i, i) entry of D in (2.14) is the same as of U^T and both are the (i, i) pivot of the Neville elimination of A (and of A^T).

All these questions can be summarized in the following:

Theorem 2.2. *A nonsingular $n \times n$ matrix A satisfies the WRC condition if and only if it can be factorized in the form*

$$\begin{aligned} A = & (E_n(m_{n1})E_{n-1}(m_{n-1,1}) \cdots E_2(m_{21})) \cdot (E_n(m_{n2}) \cdots E_3(m_{32})) \cdots \\ & \cdot (E_n(m_{n,n-1}))D(E_n^T(m_{n-1,n}))(E_{n-1}^T(m_{n-2,n-1})E_n^T(m_{n-2,n})) \cdots \\ & \cdot (E_2^T(m_{12}) \cdots E_{n-1}^T(m_{1,n-1})E_n^T(m_{1n})) \end{aligned} \quad (2.16)$$

with D a diagonal matrix and the m_{ij} 's satisfying

$$\begin{aligned} m_{ij} = 0 & \Rightarrow m_{hj} = 0 \quad \forall h > i \quad \text{if } i > j, \\ m_{ij} = 0 & \Rightarrow m_{ik} = 0 \quad \forall k > j \quad \text{if } i < j. \end{aligned} \quad (2.17)$$

Moreover, under this condition the factorization is unique, m_{ij} is the (i, j) multiplier of the complete Neville elimination of A and the (i, i) entry of D is the (i, i) pivot of the Neville elimination of A .

§4. Totally positive matrices

In [19], Corollary 5.5, it was proved that a nonsingular matrix A is totally positive if and only if there are no row or column exchanges in the complete Neville elimination of A and all the pivots are nonnegative. Furthermore it is clear that the diagonal pivots must be different from zero. With the notations of Section 2 and taking into account that, according to (2.5), the multipliers of the elimination process are quotient of pivots, we can reformulate this result as

Theorem 4.1. *A nonsingular $n \times n$ matrix A is totally positive if and only if it can be factorized in the form (2.16) with D a diagonal matrix with positive diagonal entries and the $m_{i,j}$'s nonnegative numbers satisfying (2.17). Moreover, under this condition the factorization is unique, $m_{i,j}$ is the (i, j) multiplier of the complete Neville elimination of A and the (i, i) entry of D is the (i, i) pivot of the Neville elimination of A .*

Analogously, following Section 3 a nonsingular TP matrix can be expressed in two different forms as a product of bidiagonal matrices.

Theorem 4.2. *A nonsingular $n \times n$ matrix A is totally positive if and only if it can be factorized in the form (3.2) (respectively (3.10)) with D a diagonal matrix with positive diagonal entries, F_i, G_i like in (3.3),(3.4), (resp. H_i, K_i like in (3.11),(3.12)) and the $m_{i,j}$'s ($\alpha_{i,j}$'s) nonnegative numbers satisfying (2.17) ((3.13)). Under this condition, the factorization is unique.*

Since (3.2) comes from a simple reordering of (2.15), the interpretation of the numbers $m_{i,j}$ is the same as in Theorem 4.1. The $\alpha_{i,j}$'s are the same numbers, in general in a different ordering. However, when all the multipliers $m_{i,j}$ are different from zero, (2.17) and (3.13) do not apply and by uniqueness both factorizations are coincident.

Example: The matrix

$$A = \begin{pmatrix} 1 & 1 & 3 & 6 & 0 \\ 1 & 2 & 7 & 15 & 2 \\ 3 & 6 & 23 & 52 & 13 \\ 0 & 0 & 4 & 15 & 17 \\ 0 & 0 & 2 & 10 & 24 \end{pmatrix}$$

is totally positive because its complete Neville elimination can be carried out without row or column exchanges, with multipliers $m_{51} = m_{41} = 0$, $m_{31} = 3$, $m_{21} = 1$, $m_{52} = m_{42} = m_{32} = 0$, $m_{53} = 2$, $m_{43} = 1/2$, $m_{54} = 5/2$, $m_{15} = 0$, $m_{14} = 2$, $m_{13} = 3$, $m_{12} = 1$, $m_{25} = 2$, $m_{24} = m_{23} = 1$, $m_{35} = 1$, $m_{34} = 1/2$, $m_{45} = 0$, and diagonal pivots $p_{11} = p_{22} = p_{44} = 1$, $p_{33} = 2$, $p_{55} = 8$. Its factorization (3.2) is

$$\begin{aligned}
A &= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1/2 & 1 \end{pmatrix} \\
&\cdot \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 5/2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 8 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1/2 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \\
&\cdot \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.
\end{aligned}$$

In order to get the factorization (3.10), the first, third and fourth factors above must be replaced, respectively, by

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1/2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 5/2 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1/2 & 1 \end{pmatrix}.$$

As it has been recalled in Section 1 a matrix is said to be *strictly totally positive* (STP for brevity) if all its minors are positive. In [19], Theorem 4.1, it was proved that a square matrix is STP if and only if it satisfies the WRC condition and the multipliers and diagonal pivots of its complete Neville elimination are all positive. This can be expressed in a theorem similar to Theorem 4.2:

Theorem 4.3. *A nonsingular $n \times n$ matrix A is strictly totally positive if and only if it can be factorized in the form (3.2) with D a diagonal matrix with positive diagonal entries, F_i, G_i like in (3.3),(3.4), and the $m_{i,j}$'s positive numbers. This factorization is unique.*

Observe that the condition (2.17) does not apply here and that the factorizations (3.2) and (3.10) are now coincident.

An important class of TP matrices is that of *almost strictly totally positive* matrices (see [15]). A nonsingular TP matrix A of order n is called almost strictly totally positive (ASTP for brevity) if it satisfies the following property: for any $\alpha, \beta \in Q_{k,n}^0$

$$\det A[\alpha|\beta] > 0 \iff a_{\alpha_h, \beta_h} > 0, \quad h = 1, 2, \dots, k. \quad (4.1)$$

In other words, a minor of A formed from consecutive rows and columns is positive if and only if all its diagonal entries are positive. It was proved in [15] that if A is ASTP then (4.1) holds for any $\alpha, \beta \in Q_{k,n}$ and, consequently, for this type of matrices, we know exactly the minors which are positive and the ones which are zero. Important examples of ASTP matrices are the collocation matrices of B-splines (see [2]) and Hurwitz matrices (see [30]). Obviously, STP matrices form a subclass of ASTP matrices. The following theorem ([22], Theorem 4.1 and Remark 4.2) allows us the characterization of ASTP matrices in terms of their factorizations:

Theorem 4.4. *A nonsingular $n \times n$ matrix A is ASTP if and only if it can be factorized in the form (3.2) with D a diagonal matrix with positive diagonal entries, F_i, G_i like in (3.3),(3.4), and with $m_{ij} \geq 0$ satisfying*

$$\begin{aligned} m_{ij} = 0 \text{ and } i > j &\Rightarrow m_{rs} = 0 \quad \forall (r, s) \text{ with } r \geq i, s \leq j \\ m_{ij} = 0 \text{ and } i < j &\Rightarrow m_{rs} = 0 \quad \forall (r, s) \text{ with } r \leq i, s \geq j. \end{aligned} \quad (4.2)$$

Let us observe that (2.17) is a part of (4.2) and therefore the factorization is unique. Moreover, since (3.13) is also a part of (4.2) the factorizations (3.2) and (3.10) are coincident for ASTP matrices.

Remark. A *signature sequence* of order n is a sequence $\varepsilon = (\varepsilon_i)_{1 \leq i \leq n}$ of real numbers with $|\varepsilon_i| = 1$ for all i . An $n \times m$ matrix A is called *sign-regular with signature ε* if ε is a signature sequence of order $h = \min\{n, m\}$ such that, for $k = 1, 2, \dots, h$,

$$\varepsilon_k \det A[\alpha|\beta] \geq 0 \quad (4.3)$$

for any $\alpha \in Q_{k,n}, \beta \in Q_{k,m}$. A is called *strictly sign-regular with signature ε* if \geq is replaced by $>$ in (4.3). TP and STP matrices are examples of sign-regular and strictly sign-regular matrices, respectively, with $\varepsilon_i = 1$ for all i . Neville elimination provides an algorithmic characterization of strictly sign-regular matrices as can be found in [17] and also a factorization of them as a product of bidiagonal matrices. In the same paper, we gave the algorithms corresponding to Theorems 4.1 and 4.3 above to check the total positivity or strict total positivity of a matrix via Neville elimination.

Some other factorizations of nonsingular totally positive matrices of order n as a product of $2n - 2$ bidiagonal matrices by a diagonal matrix can be considered. The *converse* $A^\#$ of a matrix $A = (a_{ij})_{0 \leq i, j \leq n}$ is the matrix whose (i, j) entry ($0 \leq i, j \leq n$) is $a_{n-i, n-j}$ (see [1]). It is easy to prove that $(AB)^\# = A^\#B^\#$. According to Theorem 4.2, if A is a nonsingular TP matrix, then the upper triangular matrix $K = K_1 \cdots K_{n-2}K_{n-1}$ of (3.10) is TP. Consequently its converse $K^\#$ is lower triangular and TP and again by Theorem 4.2 it can be decomposed as in (3.10)

$$K^\# = \tilde{H}_{n-1}\tilde{H}_{n-2} \cdots \tilde{H}_1, \quad (4.4)$$

where the matrices \tilde{H}_i have the form (3.11) with the bidiagonal elements satisfying (3.13). If we denote $U_i := \tilde{H}_i^\#$, one has

$$K = U_{n-1}U_{n-2} \cdots U_1.$$

$$A = \hat{U}_{n-1}\hat{U}_{n-2}\cdots\hat{U}_1\hat{D}\hat{H}_{n-1}\cdots\hat{H}_2\hat{H}_1, \quad (4.12)$$

$$A = \hat{K}_1\hat{K}_2\cdots\hat{K}_{n-1}\hat{D}\hat{L}_1\hat{L}_2\cdots\hat{L}_{n-1}, \quad (4.13)$$

$$A = \hat{K}_1\hat{K}_2\cdots\hat{K}_{n-1}\hat{D}\hat{H}_{n-1}\cdots\hat{H}_2\hat{H}_1. \quad (4.14)$$

In summary, from (3.10) we have got eight different factorizations of A as a product of bidiagonal matrices. Other eight factorizations, with the same types of factors as above but with different conditions for the subdiagonal elements are obtained by starting from the factorization (2.16) with the condition (2.17). Nevertheless, recall that for ASTP matrices, and in particular for STP matrices, (2.16) and (3.10) are coincident and therefore only the eight first factorizations must be considered for these matrices.

A matrix with nonnegative entries is called *stochastic* if the row sums are equal to one. Matrices which are nonsingular, stochastic and totally positive are of particular interest in Computer Aided Geometric Design. An *elementary corner cutting* is a transformation which maps any polygon $P_1P_2\cdots P_n$ into another one $Q_1Q_2\cdots Q_n$ defined by

$$\begin{aligned} Q_j &= P_j \quad j \neq i \\ Q_i &= \lambda P_i + (1 - \lambda)P_{i+1} \end{aligned} \quad (4.15)$$

for some $1 \leq i \leq n - 1$ or

$$\begin{aligned} Q_j &= P_j \quad j \neq i \\ Q_i &= \lambda P_i + (1 - \lambda)P_{i-1} \end{aligned} \quad (4.16)$$

for some $2 \leq i \leq n$.

A *corner cutting algorithm* is any composition of elementary corner cutting transformations. An elementary corner cutting is defined by a bidiagonal, nonsingular, totally positive and stochastic matrix, which is upper (respectively lower) triangular in the case (4.15) (resp. (4.16)). In [24] Goodman and Micchelli proved that a matrix which is nonsingular, totally positive and stochastic, can be written as a product of bidiagonal matrices of the same type and therefore describes a corner cutting algorithm.

Observe that a nonsingular, totally positive matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ can be factorized in the form

$$A = DB \quad (4.17)$$

with $D = (d_{ij})_{1 \leq i, j \leq n}$ a diagonal matrix with positive diagonal entries $d_{ii} = 1/(\sum_{i=1}^n a_{ij})$ and B a nonsingular, totally positive, stochastic matrix. The factorization is obviously unique. By using (4.17) for each factor, a product of nonsingular, totally positive matrices $A = A_1A_2\cdots A_m$ can be written

$$A = DB_1B_2\cdots B_m \quad (4.18)$$

with B_1, B_2, \dots, B_m totally positive and stochastic and D a diagonal matrix with positive diagonal. Moreover, the matrix A is stochastic if and only if D is the identity matrix I .

Hence, Theorem 4.2 can be slightly changed to characterize nonsingular, stochastic, TP matrices in terms of factorizations with bidiagonal, stochastic factors:

for a nonsingular matrix A to be TP is that the signs of the minors involved in the calculation of multipliers and diagonal pivots of the complete Neville elimination of A ensure that the conditions of Theorem 4.1, for instance, are accomplished.

In consequence, Theorem 4.1 leads to the following determinantal characterization, which corresponds to [16], Theorem 3.1:

Theorem 5.1. *Let A be a nonsingular matrix of order n . Then A is TP if and only if it satisfies simultaneously, for each $k \in \{1, 2, \dots, n\}$, the following conditions:*

$$\det A[\alpha|1, 2, \dots, k] \geq 0 \quad \forall \alpha \in Q_{k,n} \quad (5.1)$$

$$\det A[1, 2, \dots, k|\beta] \geq 0 \quad \forall \beta \in Q_{k,n}, \quad (5.2)$$

$$\det A[1, 2, \dots, k] > 0. \quad (5.3)$$

This characterization reduces considerably the number of minors used in the characterization obtained by Cryer in [8], namely: a nonsingular matrix A is TP if and only if

$$\det A[\alpha|\beta] \geq 0 \quad \forall \alpha \in Q_{k,n}^0, \beta \in Q_{k,n}, k \in \{1, \dots, n\}. \quad (5.4)$$

On the other hand, observe the symmetry of the conditions of Theorem 5.1 with respect of rows and columns, compared with the unnatural asymmetry of (5.4). In the case of general TP matrices, nonnecessarily regular, (5.4) is replaced in [8] by a similar condition which depends on the rang of A . These matrices can be characterized too by their Neville elimination as it can be seen in [14], Theorem 5.4.

With respect to STP matrices, Theorem 4.3 corresponds to the following result, which was contained in [14], Theorem 4.1:

Theorem 5.2. *Let A be an $n \times m$ matrix. Then A is STP if and only if*

$$\det A[\alpha|1, \dots, k] > 0 \quad \forall \alpha \in Q_{k,n}^0, k \in \{1, \dots, n\}, \quad (5.5)$$

$$\det A[1, \dots, k|\beta] > 0 \quad \forall \beta \in Q_{k,n}^0, k \in \{1, \dots, n\}. \quad (5.6)$$

This result improves the classical characterization of STP matrices due to Fekete in [11] (another proof thereof can be found in Theorem 2.5 of [1]): an $n \times m$ matrix is STP if and only if all its minors formed by consecutive rows and columns are strictly positive.

Observe that, for example, in the case $n = m$, $n(n+1)(2n+1)/6$ minors should be checked according to Fekete's criterion, while they are only $n(n+1)$ according to Theorem 5.2.

Concerning ASTP matrices, Theorem 4.4 is closely related with the following result, which is contained in [22], Theorem 3.3. First, we must introduce some notations related with the zero pattern of an ASTP matrix A .

For an $n \times n$ matrix A let us denote:

$$i_0 = 1, \quad j_0 = 1;$$

for $t = 1, \dots, l$:

$$i_t = \max\{i | a_{i, j_{t-1}} \neq 0\} + 1 \quad (\leq n + 1),$$

$$j_t = \max\{j | a_{i_t, j} = 0\} + 1 \quad (\leq n + 1),$$

where l is given in this recurrent definition by $i_l = n + 1$. Analogously we denote:

$$\hat{j}_0 = 1, \quad \hat{i}_0 = 1;$$

for $t = 1, \dots, r$:

$$\hat{j}_t = \max\{j | a_{\hat{i}_{t-1}, j} \neq 0\} + 1,$$

$$\hat{i}_t = \max\{i | a_{i, \hat{j}_t} = 0\} + 1,$$

where $\hat{j}_r = n + 1$. In other words, the entries below the places $(i_1 - 1, j)$ with $j_0 \leq j < j_1$, $(i_2 - 1, j)$ with $j_1 \leq j < j_2$, \dots , $(i_{l-1} - 1, j)$ with $j_{l-2} \leq j < j_{l-1}$ are zero. So are the entries to the right of the places $(i, \hat{j}_1 - 1)$ with $\hat{i}_0 \leq i < \hat{i}_1$, $(i, \hat{j}_2 - 1)$ with $\hat{i}_1 \leq i < \hat{i}_2$, \dots , $(i, \hat{j}_{r-1} - 1)$ with $\hat{i}_{r-2} \leq i < \hat{i}_{r-1}$. On the other hand, the entries of both lists, those above the first list and those to the left of the last list are nonzero. We shall express this by saying that the matrix A has a zero pattern given by $I = \{i_0, i_1, \dots, i_l\}$, $J = \{j_0, j_1, \dots, j_l\}$, $\hat{I} = \{\hat{i}_0, \hat{i}_1, \dots, \hat{i}_r\}$ and $\hat{J} = \{\hat{j}_0, \hat{j}_1, \dots, \hat{j}_r\}$. Only matrices with these patterns of zeros and all the other entries positive can be ASTP, as it is explained in [22] with the following result:

Theorem 5.3. *Let A be a nonsingular $n \times n$ totally positive matrix with a zero pattern given by I, J, \hat{I}, \hat{J} as above, satisfying*

$$\begin{aligned} i_t &> j_t & t = 1, \dots, l - 1 \\ \hat{j}_t &> \hat{i}_t & t = 1, \dots, r - 1. \end{aligned} \tag{5.7}$$

Then A is an ASTP matrix if and only if it satisfies the following conditions simultaneously:

i) For $1 \leq t \leq l$,

$$\text{for } j_{t-1} \leq h \dot{y} < j_t,$$

$$\det A[i_t - 1 - h + j_k, \dots, i_t - 1 | j_k, j_k + 1, \dots, h] > 0,$$

where $j_k = \max\{j_s | s \leq t - 1, h - j_s < i_t - i_s\}$.

ii) For $1 \leq t \leq r$,

$$\text{for } \hat{i}_{t-1} \leq h \dot{y} < \hat{i}_t,$$

$$\det A[\hat{i}_k, \hat{i}_k + 1 \dots, h | \hat{j}_t - 1 - h + \hat{i}_k, \dots, \hat{j}_t - 1] > 0$$

where $\hat{i}_k = \max\{\hat{i}_s | s \leq t - 1, h - \hat{i}_s < \hat{j}_t - \hat{j}_s\}$.

On the other hand, the decompositions studied in the precedent section can be expressed in a more compact form if we replace all the lower triangular bidiagonal matrices by their product and, analogously, all the upper triangular ones by their product. In this form we get the LDU factorization of the matrix A , and so Theorem 4.1 can be reformulated (see [9], Theorem 7.1 for a first version of this result)

Theorem 5.4. *A nonsingular $n \times n$ matrix A is totally positive if and only if it can be decomposed in the form LDU with D a diagonal matrix with positive diagonal entries, L a lower triangular, unit diagonal TP matrix and U an upper triangular, unit diagonal TP matrix.*

Analogously, the factorization of a strictly totally positive matrix which has been given in Theorem 4.3 can be rewritten in the form LDU , but, obviously, the triangular matrices L and U are not STP. If B is an $n \times n$ lower (resp., upper) triangular matrix such that $\det B[\alpha|\beta] > 0$ for any $\alpha, \beta \in Q_{p,n}$, $1 \leq p \leq n$, with $\beta_k \leq \alpha_k$ (resp, $\beta_k \geq \alpha_k$) then B is called Δ STP matrix. Observe that a Δ STP matrix is an ASTP matrix. Hence we get (see [8], Theorem 1.1)

Theorem 5.5. *An $n \times n$ matrix A is strictly totally positive if and only if it can be decomposed in the form LDU with D a diagonal matrix with positive diagonal entries, L a lower triangular, unit diagonal Δ STP matrix and U an upper triangular, unit diagonal Δ STP matrix.*

Another class of important TP matrices has not yet been considered in this paper: that of the matrices A such that A^m is STP for some positive integer m . They are called *oscillatory*. A triangular matrix A is called Δ -oscillatory if A is TP and A^m is Δ STP for some positive integer m . Simultaneously to the result of Theorem 5.5, it was shown in Theorem 1.1 of [8] that a square matrix A is oscillatory if and only if it has an LU -factorization such that L and U are Δ oscillatory.

Let us give two final remarks. The QR factorization of a matrix, that is its decomposition as the product of an orthogonal matrix by an upper triangular matrix is important in Numerical Analysis. Its application to TP matrices has been recently considered by us in [16]. By introducing some new classes of matrices related to that of TP matrices, whose definitions are omitted here for brevity, nonsingular TP and STP matrices are characterized in terms of their QR factorization in Theorem 4.7 of that paper.

The second and last remark concerns to the solution of linear systems with TP coefficient matrices. Some years ago, in [3], deBoor and Pinkus proved that partial pivoting is not necessary when Gauss elimination is used to solve them. Recently, in [18], we have studied scaled partial pivoting with respect to l_∞ -norm and Euclidean norm for Gauss and Neville elimination applied to the same systems. It has been proved that in exact arithmetic they do not need row exchanges. The same result holds, for sufficiently high precision arithmetic, in Gauss elimination and also, for almost strictly totally positive matrices, in Neville elimination.

References

1. Ando, T., Totally positive matrices, Linear Algebra Appl. **90** (1987), 165–219.
2. de Boor, C., Total positivity of the spline collocation matrix, Indiana Univ. J. Math. **25** (1976), 541–551.

3. de Boor, C. and Pinkus, A. Backward error analysis for totally positive linear systems, *Numer. Math.* **27** (1977), 485–490.
4. de Boor, C. and Pinkus, A. The approximation of a totally positive band matrix by a strictly banded totally positive one, *Linear Algebra Appl.* **42** (1982), 81–98.
5. Carnicer, J. M., and Peña, J. M., Shape preserving representations and optimality of the Bernstein basis, *Advances in Computational Mathematics* **1** (1993), 173–196.
6. Carnicer, J. M., and Peña, J. M., Totally positive bases for shape preserving curve design and optimality of B-splines. *Computer-Aided Geom. Design* **11** (1994), 633–654.
7. Cavaretta, A. S., Dahmen, W., Micchelli, C. A., and Smith, P. W., A factorization theorem for banded matrices, *Linear Algebra Appl.* **39** (1981), 229–245.
8. Cryer, C. W., The *LU*-factorization of totally positive matrices, *Linear Algebra Appl.* **7** (1973), 83–92.
9. Cryer, C. W., Some properties of totally positive matrices, *Linear Algebra Appl.* **15** (1976), 1–25.
10. Dahmen, W., and Micchelli, C.A., Banded matrices with banded inverses II: Locally finite decomposition of spline spaces, *Constructive Approximation* **9** (1993), 263–282.
11. Fekete, M., and Polya, G., Uber ein Problem von Laguerre, *Rend. C.M. Palermo* **34** (1912), 89–120.
12. Frydman H., and Singer B., Total positivity and the embedding problem for Markov chains, *Math. Proc. Cambridge Philos. Soc.* **85** (1979), 339–344.
13. Gasca, M., and Mühlbach, G., Generalized Schur complements and a Test for Total Positivity, *Applied Numer. Math.* **3** (1987), 215–232.
14. Gasca, M., and Peña, J. M., Total positivity and Neville elimination, *Linear Algebra Appl.* **165** (1992), 25–44.
15. Gasca, M., Micchelli, C. A., and Peña, J. M., Almost strictly totally positive matrices, *Numerical Algorithms* **2** (1992), 225–236.
16. Gasca, M., and Peña, J. M., Total positivity, *QR*-factorization and Neville elimination, *SIAM J. on Matrix Anal. Appl.* **14** (1993), 1132–1140.
17. Gasca, M., and Peña, J. M., Sign-regular and totally positive matrices: an algorithmic approach, in *International Multivariate Approximation: From CAGD to wavelets*, Jetter, K. and Utreras, L. (eds.), World Scientific Publishers, 1993, 131–146.
18. Gasca, M., and Peña, J. M., Scaled pivoting in Gauss and Neville elimination for totally positive systems, *Applied Numerical Mathematics*, **13** (1993), 345–356.
19. Gasca, M., and Peña, J. M., A matricial description of Neville elimination with applications to total positivity, *Linear Algebra Appl.* **202** (1994), 33–54.
20. Gasca, M., and Peña, J. M., Corner cutting algorithms and totally positive matrices, in *Curves and Surfaces in Geometric Design*, P. J. Laurent,

- A. Le Méhauté and L. L. Schumaker (eds.), AKPeters, 1994, 177–184.
21. Gasca, M., Micchelli, C. A., and Peña, J. M., Banded matrices with banded inverses III: p -slanted matrices, in *Wavelets, Images and Surface fitting*, P. J. Laurent, A. Le Méhauté and L. L. Schumaker (eds.), AKPeters, 1994, 245–268.
 22. Gasca, M., and Peña, J. M., On the characterization of Almost Strictly Totally positive matrices, to appear in *Advances in Computational Mathematics*.
 23. Goodman, G. A probabilistic representation of totally positive matrices, *Advances in Applied Mathematics* **7** (1986), 236–252.
 24. Goodman, T. N. T., and Micchelli, C. A., Corner cutting algorithms for the Bezier representation of free form curves, *Linear Alg. Appl.* **99** (1988), 225–252.
 25. Goodman, T. N. T., Shape preserving representations, in *Mathematical Methods in Computer Aided Geometric Design*, T. Lyche and L. Schumaker (eds.), Academic Press, New York, 1989, 333–357.
 26. Goodman, T. N. T., and Said, H. B., Shape preserving properties of the generalized Ball basis, *Computer-Aided Geom. Design* **8** (1991), 115–121.
 27. Karlin, S., and McGregor, J. L., Coincidence probabilities of birth and death processes, *Pacific J. Math.* **9** (1959) 1109–1140.
 28. Karlin, S., and McGregor, J. L., Coincidence probabilities, *Pacific J. Math.* **9** (1959) 1141–1164.
 29. Karlin, S., *Total positivity*, Vol. I, Stanford University Press, Standford, 1968.
 30. Kempermann, J. H. B., A Hurwitz matrix is totally positive, *SIAM J. Math. Anal.* **13** (1982), 331–341.
 31. Loewner, C., On totally positive matrices, *Math. Z.* **63** (1955), 338–340.
 32. Micchelli, C.A., and Pinkus A., Descartes systems from corner cutting, *Constr. Approx.*, (1991) **7**, 195–208.
 33. Whitney, A. M., A reduction theorem for totally positive matrices, *J. d'Analyse Math.* **2** (1952), 88–92.

Acknowledgements. Both authors have been partially supported by the DGICYT Spain Research Grant PB93-0310.

Mariano Gasca
 Departamento de Matemática Aplicada
 Facultad de Ciencias, Universidad de Zaragoza,
 50009 Zaragoza, Spain.
 gasca@cc.unizar.es

Juan M. Peña
 Departamento de Matemática Aplicada
 Facultad de Ciencias, Universidad de Zaragoza,
 50009 Zaragoza, Spain.
 jmpena@cc.unizar.es