

TESIS DE LA UNIVERSIDAD
DE ZARAGOZA

2022

198

Héctor Orera Hernández

Numerical methods and accurate computations with structured matrices

Director/es

Peña Ferrández, Juan Manuel
Delgado Gracia, Jorge

<http://zaguan.unizar.es/collection/Tesis>

ISSN 2254-7606



Premsas de la Universidad
Universidad Zaragoza

© Universidad de Zaragoza
Servicio de Publicaciones

ISSN 2254-7606



Universidad
Zaragoza

Tesis Doctoral

NUMERICAL METHODS AND ACCURATE
COMPUTATIONS WITH STRUCTURED MATRICES

Autor

Héctor Orera Hernández

Director/es

Peña Ferrández, Juan Manuel
Delgado Gracia, Jorge

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

Programa de Doctorado en Matemáticas y Estadística

2022

NUMERICAL METHODS AND ACCURATE COMPUTATIONS WITH STRUCTURED MATRICES



Universidad
Zaragoza

Héctor Orera Hernández

Tesis doctoral en matemáticas

Universidad de Zaragoza

Directores de tesis:

Dr. D. Juan Manuel Peña Ferrández

Dr. D. Jorge Delgado Gracia

This doctoral thesis is presented as a compendium of the following research articles:

- [17] J. Delgado, H. Orera and J. M. Peña. Accurate computations with Laguerre matrices. *Numer. Linear Algebra Appl.* 26 (2019), e2217, 10 pp.
- [16] J. Delgado, H. Orera and J. M. Peña. Accurate algorithms for Bessel matrices. *J. Sci. Comput.* 80 (2019), 1264-1278.
- [79] H. Orera and J. M. Peña. Accurate inverses of Nekrasov Z-matrices. *Linear Algebra Appl.* 574 (2019), 46-59.
- [81] H. Orera and J. M. Peña. Infinity norm bounds for the inverse of Nekrasov matrices using scaling matrices. *Appl. Math. Comput.* 358 (2019), 119-127.
- [80] H. Orera and J. M. Peña. B_{π}^R -tensors. *Linear Algebra Appl.* 581 (2019), 247-259.
- [18] J. Delgado, H. Orera and J. M. Peña. Accurate bidiagonal decomposition and computations with generalized Pascal matrices. *J. Comput. Appl. Math.* 391 (2021), Paper No. 113443, 10 pp.
- [20] J. Delgado, H. Orera and J. M. Peña. High relative accuracy with matrices of q-integers. *Numer. Linear Algebra Appl.* 28 (2021), Paper No. e2383, 20 pp.
- [19] J. Delgado, H. Orera and J. M. Peña. Optimal properties of tensor product of B-bases. *Appl. Math. Lett.* 121 (2021), Paper No. 107473, 5 pp.
- [21] J. Delgado, H. Orera and J. M. Peña. Characterizations and accurate computations for tridiagonal Toeplitz matrices, *Linear and Multilinear Algebra* (2021), Published online, DOI: 10.1080/03081087.2021.1884180.
- [82] H. Orera and J. M. Peña. Accurate determinants of some classes of matrices. *Linear Algebra Appl.* 630 (2021), 1-14.
- [83] H. Orera and J. M. Peña. Error bounds for linear complementarity problems of B_{π}^R -matrices. *Comput. Appl. Math.* 40 (2021), Paper No. 94, 13 pp.

The dissertation's author has been employed under a Gobierno de Aragón predoctoral contract from 01/08/2018 to 09/06/2019. From 10/06/2019, his predoctoral contract has been funded by the FPU program (Ministerio de Ciencia, Innovación y Universidades), FPU17/03769.

In addition, the author's work has also been supported by the following research grants:

- Análisis de la representación de curvas y superficies, cálculos precisos con matrices estructuradas y aplicaciones, PGC2018-096321-B-I00. Ministerio de Ciencia, Innovación y Universidades. 01/01/2019 - 31/12/2022. PI: Juan Manuel Peña Ferrández.
- Análisis Numérico, Optimización y Aplicaciones. Gobierno de Aragón, E41_20R. 01/01/2020 - 31/12/2022. PI: Juan Manuel Peña Ferrández.

Acknowledgments

I would like to start by expressing my deepest appreciation to my thesis supervisors, Prof. Dr. Juan Manuel Peña and Prof. Dr. Jorge Delgado, for their invaluable guidance and support throughout the realization of this thesis. They have shared with me both their enthusiasm for research and their vast knowledge, and they have always offered me their advice and encouragement during my time as a doctoral student.

I wish to thank my colleagues from the Department of Applied Mathematics for providing me with a great work environment and assisting me on this journey.

I am grateful to Professor Dr. Tomas Sauer for his hospitality and support during my research stay at the University of Passau. I would also like to extend my thanks to all the people at FORWISS. They made me feel at home during my time in Passau.

Finally, I cannot forget my family and friends. Their influence, love and encouragement have been fundamental during this period. They have helped me navigate through a pandemic while keeping the illusion of moving forward with this project. I cannot begin to express my thanks to my parents, Jorge and Carmen, and to my brother, David, for their unconditional support. To them, and to everyone who has shared a part of this journey with me, thank you.

Abstract

The main topic of this doctoral thesis is Numerical Linear Algebra, with a special emphasis on two structured classes of matrices: totally positive matrices and M -matrices. For some subclasses of these matrices, it is possible to develop algorithms to solve numerically to high relative accuracy some of the most common problems in linear algebra independently of the traditional condition number. The key to achieve accurate computations lies in the use of a different parametrization that captures the special structure of the matrix and in the development of adapted numerical methods that use this parametrization as input.

Nonsingular totally positive matrices admit a unique factorization as a product of bidiagonal nonnegative matrices called the bidiagonal decomposition [46, 47]. If this decomposition is known accurately, it can be used to solve some linear systems of equations as well as to compute the inverse, the eigenvalues and the singular values to high relative accuracy using the methods developed by P. Koev [58–60]. Hence, finding a method to compute the bidiagonal decomposition of a nonsingular totally positive matrix to high relative accuracy gives a parametrization to achieve high relative accuracy when solving the previously mentioned problems. Our contribution in this area has been obtaining the bidiagonal decomposition to high relative accuracy of collocation matrices of generalized Laguerre polynomials [17], of collocation matrices of Bessel polynomials [16], of classes of matrices that generalize the Pascal matrix [18] and of matrices of q -integers [20]. We have also studied the extension of some optimal properties of collocation matrices of normalized B -bases (that are totally positive matrices). In particular, we have derived optimal properties for the collocation matrices of the tensor product of normalized B -bases [19].

If we know the off-diagonal entries and the row sums of a nonsingular diagonally dominant M -matrix to high relative accuracy, then we can compute its inverse, determinant and singular values also to high relative accuracy. We have looked for new methods to achieve accurate computations with more subclasses of M -matrices. We introduced a parametrization for Nekrasov Z -matrices with positive diagonal entries that can be used to compute its inverse and determinant to high relative accuracy [79, 82]. We also studied a class of matrices called B -matrices that is closely related to M -matrices. We obtained a method to compute its determinant to high relative accuracy as well the determinants of B -Nekrasov matrices in [82]. Based on the use of the scaling matrices that we have proposed, we developed new infinity norm bounds for the inverse of a Nekrasov matrix as well as error bounds for the linear complementarity problem when the associated matrix is Nekrasov [81]. We also developed infinity norm bounds for the inverses of B_{π}^R -matrices, a class that extends B -matrices, and we applied them to derive new error bounds for the linear complementarity problem whose as-

sociated matrix is a B_{π}^R -matrix [83]. Some classes of matrices have been extended to a higher order case to develop new theory for tensors. For example, the class of B -matrices was generalized to B -tensors and gave a new simple criterion for the identification of a new class of positive definite tensors. We have proposed an extension of the class of B_{π}^R -matrices to B_{π}^R -tensors in [80], giving a new class of positive definite tensors that can be identified by using a criterion based only on the tensor entries. Finally, we characterized tridiagonal Toeplitz P -matrices and studied the cases when a bidiagonal decomposition could be obtained and used as a parametrization to achieve accurate computations [21].

Resumen

El tema principal de esta tesis doctoral es el Álgebra Lineal Numérica, con un énfasis especial en dos clases de matrices estructuradas: las matrices totalmente positivas y las M -matrices. Para algunas subclases de estas matrices, es posible desarrollar algoritmos para resolver numéricamente varios de los problemas más comunes en álgebra lineal con alta precisión relativa independientemente del número de condición de la matriz. La clave para lograr cálculos precisos está en el uso de una parametrización diferente que represente la estructura especial de la matriz y en el desarrollo de algoritmos adaptados que trabajen con dicha parametrización.

Las matrices totalmente positivas no singulares admiten una factorización única como producto de matrices bidiagonales no negativas llamada factorización bidiagonal [46, 47]. Si conocemos esta representación con alta precisión relativa, se puede utilizar para resolver ciertos sistemas de ecuaciones y para calcular la inversa, los valores propios y los valores singulares con alta precisión relativa utilizando los métodos desarrollados por P. Koev [58–60]. Por tanto, si encontramos un método para calcular la factorización bidiagonal de una matriz totalmente positiva no singular con alta precisión relativa tendremos una representación de la matriz que nos permite resolver muchos problemas con ella también con alta precisión relativa. Nuestra contribución en este campo ha sido la obtención de la factorización bidiagonal con alta precisión relativa de matrices de colocación de polinomios de Laguerre generalizados [17], de matrices de colocación de polinomios de Bessel [16], de clases de matrices que generalizan la matriz de Pascal [18] y de matrices de q -enteros [20]. También hemos estudiado la extensión de varias propiedades óptimas de las matrices de colocación de B -bases normalizadas (que en particular son matrices totalmente positivas). En particular, hemos demostrado propiedades de optimalidad de las matrices de colocación del producto tensorial de B -bases normalizadas [19].

Si conocemos las sumas de filas y las entradas extradiagonales de una M -matriz no singular diagonal dominante con alta precisión relativa, entonces podemos calcular su inversa, determinante y valores singulares también con alta precisión relativa. Hemos buscado nuevos métodos para lograr cálculos precisos con nuevas clases de M -matrices o matrices relacionadas. Hemos propuesto una parametrización para las Z -matrices de Nekrasov con entradas diagonales positivas que puede utilizarse para calcular su inversa y determinante con alta precisión relativa [79, 82]. También hemos estudiado la clase denominada B -matrices, que está muy relacionada con las M -matrices. Hemos obtenido un método para calcular los determinantes de esta clase con alta precisión relativa y otro para calcular los determinantes de las matrices de B -Nekrasov también con alta precisión relativa en [82]. Basándonos en la utili-

zación de dos matrices de escalado que hemos introducido, hemos desarrollado nuevas cotas para la norma infinito de la inversa de una matriz de Nekrasov y para el error del problema de complementariedad lineal cuando su matriz asociada es de Nekrasov [81]. También hemos obtenido nuevas cotas para la norma infinito de las inversas de B_{π}^R -matrices, una clase que extiende a las B -matrices, y las hemos utilizado para obtener nuevas cotas del error para el problema de complementariedad lineal cuya matriz asociada es una B_{π}^R -matriz [83]. Algunas clases de matrices han sido generalizadas al caso de mayor dimensión para desarrollar una teoría para tensores extendiendo la conocida para el caso matricial. Por ejemplo, la definición de la clase de las B -matrices ha sido extendida a la clase de B -tensores, dando lugar a un criterio sencillo para identificar una nueva clase de tensores definidos positivos. En esta memoria proponemos una extensión de la clase de las B_{π}^R -matrices a B_{π}^R -tensores (ver [80]), definiendo así una nueva clase de tensores definidos positivos que puede ser identificada en base a un criterio sencillo basado solo en cálculos que involucran a las entradas del tensor. Finalmente, hemos caracterizado los casos en los que las matrices de Toeplitz tridiagonales son P -matrices y hemos estudiado cuándo pueden ser representadas en términos de una factorización bidiagonal que sirve como parametrización para lograr cálculos con alta precisión relativa [21].

Contents

I	INTRODUCTION AND BACKGROUND	1
1	Introduction	3
2	Introducción	9
3	Background	15
3.1	Basic concepts and P -matrices	15
3.2	M -matrices, related matrices and Gaussian elimination	17
3.3	Totally positive matrices and bases	20
3.4	Basic concepts and definitions about tensors	24
3.5	High relative accuracy in numerical linear algebra	26
II	ARTICLES	31
Article 1:	Accurate computations with Laguerre matrices	33
Article 2:	Accurate algorithms for Bessel matrices	45
Article 3:	Accurate inverses of Nekrasov Z -matrices	63
Article 4:	Infinity norm bounds for the inverse of Nekrasov matrices using scaling matrices	79
Article 5:	B_{π}^R -tensors	91
Article 6:	Accurate bidiagonal decomposition and computations with generalized Pascal matrices	107
Article 7:	High relative accuracy with matrices of q -integers	119
Article 8:	Optimal properties of tensor product of B -bases	141

Article 9: Characterizations and accurate computations for tridiagonal Toeplitz matrices	149
Article 10: Accurate determinants of some classes of matrices	173
Article 11: Error bounds for linear complementarity problems of B_{π}^R -matrices	189
III THEMATIC UNIT AND SUMMARY OF THE ARTICLES	205
4 Totally positive matrices	207
4.1 Accurate computations with TP matrices	208
4.1.1 Accurate computations with Laguerre matrices	210
4.1.2 Accurate computations with Bessel matrices	211
4.1.3 Accurate bidiagonal decomposition and computations with general- ized Pascal matrices	214
4.1.4 High relative accuracy with matrices of q -integers	218
4.2 Optimal properties of tensor products of B -bases	221
4.3 Tridiagonal Toeplitz P -matrices	222
5 M-matrices and related problems	227
5.1 High relative accuracy for Nekrasov Z -matrices with positive diagonal entries	228
5.2 Accurate computation of the determinant of B -matrices	229
5.3 Bounds based on diagonal scaling for Nekrasov matrices	231
5.3.1 Infinity norm bounds for the inverse of Nekrasov matrices	232
5.3.2 Error bounds for the LCP of Nekrasov matrices	233
5.4 B_{π}^R -matrices and B_{π}^R -tensors	234
5.4.1 B_{π}^R -tensors	235
5.4.2 Error bounds for LCPs of B_{π}^R -matrices	237
IV CONCLUSIONS AND FUTURE WORK	241
6 Conclusions and future work	243
7 Conclusiones y trabajo futuro	249
Appendix	255
References	257
Index of abbreviations	265

Part I

INTRODUCTION AND BACKGROUND

Chapter 1

Introduction

This thesis main topic is Numerical Linear Algebra, specifically the study of numerical methods for special classes of structured matrices. One of the main problems in this area consists on the identification of important classes of matrices whose structure can be exploited to achieve computations to high relative accuracy. Achieving accurate results is a highly desirable property for any numerical method, especially if the solution can be computed to high relative accuracy independently of the conditioning of the problem. However, until now high relative accuracy has only been guaranteed for a very small number of numerical methods for a very short list of mathematical problems. In particular, in numerical methods devised for matrices with a special structure. Among the first precedents of these methods, we would like to highlight the accurate algorithms for the computation of the singular values based on the methods developed by J. Demmel et al. [31] that have been applied to matrices related to diagonal dominance [32, 87], as well as the accurate methods for the computation of inverses, eigenvalues or singular values that have been found for certain classes of nonsingular totally positive matrices (i.e., matrices whose minors are all nonnegative) [16–18, 20, 23–26, 58–60, 67–74, 76]. Finding the right parametrization for these classes of matrices has been crucial for the development of these accurate algorithms. If we intend to obtain a small error in our computations in finite precision arithmetic even when the matrix is ill-conditioned, we need to take as input a different representation or parametrization of that matrix. The bad conditioning means that a small perturbation of the matrix entries may translate into a huge error in the computed solution in other case. This alternative representation should have a natural interpretation and it should reflect the particular structure of the matrix, as it has been the case in the previous examples. Let us note that for some particular classes of matrices it is not possible to find such a representation. For instance, the class of Toeplitz matrices presents a really simple structure but it does not admit a representation that can be used to compute its determinant to high relative accuracy [30]. In conclusion, methods that assure computations to high relative accuracy have been found mainly for special classes of matrices with a strong structure related to either positivity or diagonal dominance, and in every case the use of a special parametrization that captures that special structure has been required.

Nonnegative matrices appear frequently in many applications of diverse areas, such as Physics, Chemistry, Biology, Engineering, Economics or Social sciences. Furthermore, the

well-known results about the positivity or nonnegativity of their dominant eigenvalue and its associated eigenvector, the Perron-Frobenius theorems, have been fundamental for the mathematical modeling of many real situations. In general, many classes of matrices whose structure is characterized by positivity prove to be quite useful in applications. In particular, totally positive matrices have applications in many areas, such as Computer Aided Geometric Design, Approximation Theory, Statistics, Finance or Biomathematics. In fact, the development of the theory on Total positivity, which includes the study of totally positive matrices, has over a century of history behind and its many applications can be consulted in the classical books written by Karlin [55] and by Gantmacher and Klein [37], in the survey work of Ando [3], in the book edited by Gasca and Micchelli [44] as well as in the two recent books on totally positive matrices [36, 89]. The relevance of totally positive matrices in many applications is due to their variation diminishing properties, meaning that the linear applications defined by these matrices do not increase the number of sign changes: the number of sign changes between consecutive entries of the image vector is bounded above by the number of sign changes between consecutive entries of the input vector. On the other hand, there is a different class of matrices closely related to diagonal dominance that is also found in many applications. The matrices belonging to that class are called nonsingular M -matrices and they have nonpositive off-diagonal entries and an entrywise nonnegative inverse. These matrices appear in areas such as Economics, Numerical Analysis, Linear Programming or Dynamical Systems [6]. In both cases, either for totally positive matrices or M -matrices, it has been fundamental the previous identification of the right parametrization for the development of numerical methods that assure computations to high relative accuracy.

For nonsingular totally positive matrices, the starting point for finding a good representation has been the (unique) factorization as a product of nonnegative bidiagonal matrices obtained by M. Gasca and J.M. Peña in [46, 47]. This factorization, called the bidiagonal decomposition, was then used by P. Koev [60] for building methods that compute the singular values, eigenvalues, inverses and the solution of some linear systems of equations to high relative accuracy. These accurate methods require that the bidiagonal decomposition is also known to high relative accuracy, a requirement that has been fulfilled only for certain subclasses of totally positive matrices. The parameters given by the bidiagonal decomposition can be obtained in terms of the elimination algorithm called Neville elimination [45] (see Section 3.3). Neville elimination is an elimination procedure alternative to Gaussian elimination. Neville elimination produces zeros in a matrix column by adding to each row an appropriate multiple of the previous one, instead of a multiple of the pivot row like it is done in Gaussian elimination. The parameters given by the bidiagonal decomposition of a nonsingular totally positive matrix are the multipliers of the Neville elimination of that matrix and of its transpose as well as the associated diagonal pivots and, in any case, it can be shown that they are quotients of minors of that matrix. The computation of the bidiagonal decomposition using Neville elimination usually implies many subtractions, which can translate into a significant loss in accuracy when those subtractions are of approximate numbers of the same size. Therefore, the right strategy implies looking for an expression for these parameters in terms of the original data, in a way that all the subtractions appearing are of initial data. This idea has already been used to achieve accurate computations with some subclasses of nonsingular totally positive matrices such as Vandermonde matri-

ces [33], Bernstein-Vandermonde matrices [71, 72], Pascal matrices [2], Shoemaker-Coffey matrices [22], Said-Ball-Vandermonde matrices [74], Cauchy-Vandermonde matrices [75], Jacobi-Stirling matrices [25], collocation matrices of rational bases [24], collocation matrices of the q -Bernstein polynomials basis [26], Lupaş matrices [27] or from orthogonal polynomials associated to the Marchenko-Pastur law [73], as well as with other families of totally positive matrices. In the work presented in this dissertation, we have carried out a systematic search for subclasses of nonsingular totally positive matrices whose bidiagonal decomposition can be obtained accurately and used to solve many algebraic problems to high relative accuracy. In fact, this objective has been achieved with some new classes of totally positive matrices important in applications in areas such as Combinatorics and Orthogonal Polynomials. Specifically, for Laguerre matrices (in the article [17] presented on page 33), which are collocation matrices of the generalized Laguerre polynomials (a classical family of orthogonal polynomials), for Bessel matrices (in the article [16] presented on page 45), which are collocation matrices of Bessel polynomials, also for generalized Pascal matrices (in the article [18] presented on page 107) as well as for matrices of q -integers (in the article [20] presented on page 119). We also considered applications of totally positive matrices appearing in Computer Aided Geometric Design, in particular related to the optimal properties of the collocation matrices of normalized B -bases (bases with optimal shape preserving properties [9] as well as other optimal properties [28]) and its Kronecker products, deriving new optimal properties (in the article [19] presented on page 141). Finally, we also derived characterizations and accurate methods for tridiagonal Toeplitz P -matrices (in the article [21] presented on page 149).

Other class of matrices whose structure has proven to be useful to achieve accurate computations is that of M -matrices. Let us recall that nonsingular M -matrices have positive diagonal entries, nonnegative off-diagonal entries and an entrywise nonnegative inverse. These matrices have important applications in Numerical Analysis, Linear Programming, Dynamical Systems and Economics (see [6]). For matrices related to diagonal dominance and M -matrices, rank revealing decompositions play a key role in the study and obtention of algorithms that assure computations to high relative accuracy. A rank revealing decomposition is a factorization of the form XDY^T , where D is a nonsingular diagonal matrix and both X and Y are well-conditioned matrices. If we know a rank revealing decomposition of a matrix to high relative accuracy, then it can be used to compute its singular values to high relative accuracy using an algorithm developed by J. Demmel et al. [30]. Moreover, rank revealing decompositions have been obtained for the class of diagonally dominant M -matrices in [32, 87] using as parametrization the row sums and off-diagonal entries. In this dissertation, we have looked for new classes more general than diagonally dominant M -matrices that admit computations to high relative accuracy. For that, we should first find the right parametrization for the class. A fundamental tool for the obtention of rank revealing decompositions is given by specific pivoting strategies adapted to the special structure of the considered class of matrices, since the use of Gaussian elimination with the right pivoting strategies can provide an LDU decomposition with well-conditioned matrices L and U (see [32, 87]). We have obtained a parametrization for the class of Nekrasov Z -matrices with positive diagonal entries, which extends diagonally dominant M -matrices, and we have used that parametrization to compute their inverses and determinants to high relative accuracy (in articles [79] and [82],

included at pages 63 and 173, respectively). This class of matrices has proved to be quite useful [41] in the development of error bounds for the linear complementarity problem, which is an optimization problem with many important applications. The linear complementarity problem has a unique solution when its associated matrix is a P -matrix [11], i.e., when all its principal minors are positive. Both nonsingular totally positive matrices and nonsingular M -matrices are P -matrices. We have also obtained new bounds for the infinity norm of the inverse of Nekrasov matrices, which can be used to derive new error bounds for the linear complementarity problem (in article [81], presented on page 79). A different subclass of P -matrices is given by B -matrices [85]. We have devised a method to compute the determinants of B -matrices to high relative accuracy (in the article [82], included at page 173). Moreover, we have also studied B_{π}^R -matrices [83], a subclass of P -matrices that extends B -matrices, and we developed new error bounds for them (in article [83], included at page 189). Finally, these classes of matrices related to diagonal dominance have a potential application in the theory on hypermatrices. The research interest on hypermatrices (or tensors, see [90]) has been growing a lot recently because of its applications to Big Data, which attracts a lot of well-deserved attention because of its relevance in the field of Information Technology. This fact has translated into many publications that propose extensions of classical results for matrices to the context of tensors. We have considered the development of new sufficient conditions to assure that a symmetric tensor is positive definite, which is a property highly desired in optimization problems. In fact, we have characterized a new class of positive definite tensors via the extension of B_{π}^R -matrices to tensors (in article [80], which can be consulted at page 91).

This dissertation is organized in four parts. The first part is formed by this Introduction and Chapter 3, where we introduce basic results that are fundamental for the work presented later. In particular, the chapter introduces P -matrices, M -matrices and totally positive matrices as well as totally positive bases. It also includes basic concepts about tensors and high relative accuracy. The second part includes at pages 33, 45, 63, 79, 91, 107, 119, 141, 149, 173 and 189 the articles [17], [16], [79], [81], [80], [18], [20], [19], [21], [82] and [83], respectively, that belong to the compendium of publications of this dissertation. The third part is formed by Chapter 4 and Chapter 5. These chapters are used to justify the thematic unit of the publications as well as to present the main results of the publications. Chapter 4 is devoted to the articles on totally positive matrices. This chapter starts by presenting the new results on high relative accuracy for the following subclasses of nonsingular totally positive matrices: Laguerre matrices, Bessel matrices, generalized Pascal matrices and matrices of q -integers. It also presents the optimal properties on tensor products of normalized B -bases that we have obtained. Finally, tridiagonal Toeplitz P -matrices are characterized and we show how to perform accurate computations with them. Chapter 5 is devoted to M -matrices and other classes of related matrices. We start the chapter presenting our results about the accurate computation of the inverse and determinands of Nekrasov Z -matrices with positive diagonal entries as well as the computation of the determinants of B -matrices also to high relative accuracy. After the problems on computations to high relative accuracy, we introduce new infinity norm bounds for the inverses of Nekrasov matrices as well as error bounds for the linear complementarity problem when its associated matrix is Nekrasov with positive diagonal entries. These bounds are based on the use of a diagonal scaling matrix that transforms a Nekrasov matrix into an

strictly diagonally dominant matrix. The last section summarizes our results on B_{π}^R -matrices, a class that extends that of B -matrices. We have developed new error bounds for the linear complementarity problem involving this class as well as an extension to the tensor case that provides a new class of positive definite tensors. Finally, the last part includes the conclusions of this thesis and it presents some possible future work based on the research work presented in this dissertation.

Chapter 2

Introducción

Esta tesis trata problemas de Álgebra Lineal Numérica, sobre todo del campo de estudio de métodos numéricos adaptados a clases de matrices con estructura especial, campo que muestra una intensa y creciente actividad investigadora. Uno de los problemas principales en este campo consiste en identificar clases de matrices importantes por sus aplicaciones y para las que se puedan encontrar métodos numéricos cuyo cálculo se podrá llevar a cabo con alta precisión relativa. Conseguir cálculos precisos es una propiedad muy deseable para cualquier método numérico. En este sentido, el ideal es conseguir alta precisión relativa (independientemente del condicionamiento del problema). Sin embargo, hasta ahora sólo se ha podido garantizar dicha alta precisión relativa en un número muy reducido de métodos numéricos para una lista muy corta de problemas matemáticos. En particular, en métodos aplicados a matrices con una estructura especial. Entre los primeros precedentes, destacamos los algoritmos de alta precisión relativa para el cálculo de valores singulares, basados en las técnicas desarrolladas por el profesor James Demmel y colaboradores [31], que pudieron ser aplicados a matrices relacionadas con las diagonalmente dominantes [32, 87], así como los algoritmos de alta precisión relativa para el cálculo de inversas o valores propios y singulares que se obtuvieron para ciertas matrices totalmente positivas (es decir, matrices con todos sus menores no negativos) [16–18, 20, 23–26, 58–60, 67–74, 76]. Además, para la obtención de dichos algoritmos, encontrar una adecuada parametrización de las matrices es crucial en este campo. Si queremos obtener un pequeño error en nuestros cálculos con aritmética de precisión finita a pesar de que la matriz esté muy mal condicionada, estamos obligados a tener que trabajar con una parametrización alternativa de la matriz. Esto es debido a que una ligera perturbación en las entradas de la matriz dará lugar a un gran error en los cálculos debido al mal condicionamiento. La parametrización alternativa de la matriz deberá tener una interpretación natural y será muy conveniente que se adapte a la estructura de la matriz, como ha ocurrido en los casos mencionados anteriormente. Advertimos de antemano que no toda clase de matrices estructuradas es susceptible de admitir una parametrización que dé lugar a algoritmos de alta precisión relativa. Por ejemplo, para unas matrices estructuradas tan sencillas como las de Toeplitz, no es posible encontrar una parametrización que permita calcular con alta precisión relativa sus determinantes [30]. En resumen, las principales clases de matrices para las que se han podido desarrollar métodos con alta precisión relativa han sido clases de matrices cuya

estructura tiene alguna relación con la positividad o bien con la dominancia diagonal, siendo además necesario partir de una adecuada parametrización de las mismas.

Las matrices no negativas surgen con gran frecuencia en aplicaciones en los campos más diversos, como la Física, Química, Biología, Ingeniería, Economía y Ciencias Sociales. Además, los bien conocidos resultados sobre la positividad o no negatividad de su valor propio dominante y de su vector propio asociado (teoremas de Perron-Frobenius) ha sido una herramienta clave en la modelización matemática de muchas situaciones reales. En general, las clases de matrices relacionadas con la positividad se han mostrado muy fructíferas en las aplicaciones. En particular, las matrices totalmente positivas tienen aplicaciones en muchos campos, como diseño geométrico asistido por ordenador, teoría de aproximación, estadística, finanzas o biomatemática. De hecho, la teoría de la Total Positividad, que incluye el estudio de las matrices totalmente positivas, tiene más de un siglo de antigüedad y sus muchas aplicaciones se pueden ver tanto en libros clásicos como el de Karlin [55] o el de Gantmacher y Klein [37] como en el survey de Ando [3], en el libro editado por Gasca y Micchelli [44] y en los dos libros recientes sobre matrices totalmente positivas [36, 89]. Destaquemos que el interés de las matrices totalmente positivas no singulares en muchas aplicaciones es debido a su propiedad (conocida como disminución de la variación) consistente en que, como aplicaciones lineales, disminuyen la variación de signo: el número de variaciones de signo entre las componentes consecutivas del vector imagen no supera el número de variaciones de signo entre las componentes consecutivas del vector de partida. Por otro lado, recordemos una clase de matrices relacionadas con las diagonalmente dominantes, que es la clase de las M -matrices no singulares (que son matrices con extradiagonales no positivos e inversa no negativa). Estas matrices también tienen aplicaciones muy importantes en economía, análisis numérico, programación lineal y sistemas dinámicos, entre otros campos [6]. En ambos casos, tanto para las matrices totalmente positivas como para las M -matrices, ha sido fundamental en los estudios de alta precisión relativa la obtención previa de una parametrización adecuada de las mismas.

En el caso de matrices totalmente positivas no singulares, el punto de partida para parametrizarlas adecuadamente es la factorización (única) como producto de matrices bidiagonales obtenida por M. Gasca y J.M. Peña en [46, 47]. Esta factorización bidiagonal fue usada por P. Koev [60] para obtener algoritmos precisos (para el cálculo de valores singulares, valores propios, inversas o resolución de ciertos sistemas de ecuaciones lineales) usando las entradas de esa factorización como parámetros conocidos. Ello implica conocer con alta precisión relativa dichos parámetros, lo que se ha conseguido sólo para ciertas subclases de matrices totalmente positivas. Los parámetros de la factorización bidiagonal se pueden obtener en términos de la técnica de eliminación conocida como eliminación de Neville [45] (véase Sección 3.3). La eliminación de Neville es un procedimiento de eliminación, alternativo a la eliminación gaussiana, en el que se producen ceros en cada columna restando a cada fila un múltiplo de la fila inmediatamente anterior, en vez de un múltiplo de la fila pivote como se hace en la eliminación gaussiana. Los parámetros de la mencionada factorización bidiagonal de las matrices totalmente positivas no singulares son los multiplicadores de la eliminación de Neville de la matriz, de su traspuesta y los pivotes diagonales y así, en cualquier caso, se puede demostrar que son cocientes de determinantes de la matriz. Las expresiones de los parámetros de la factorización bidiagonal obtenidas por este procedimiento usan restas, que pueden dan

lugar a una pérdida importante de cifras significativas (y por tanto de la alta precisión relativa) cuando se realizan subtracciones de cantidades de tamaño similar. Por tanto, hay que intentar expresar dichos parámetros en términos de datos iniciales de problemas que involucren dichas matrices y de modo que sólo se permitan restas en el caso de datos iniciales. Esta idea ya se había aplicado a garantizar cálculos con alta precisión relativa en subclases de matrices totalmente positivas como las de Vandermonde [33], las de Bernstein-Vandemonde [71, 72], las de Pascal [2], las de Shoemaker-Coffey [22], las de Said-Ball-Vandermonde [74], las de Cauchy-Vandermonde [75], las matrices de Jacobi-Stirling [25], las matrices de colocación de bases racionales [24], de la base de q -Bernstein polinomios [26], las matrices de Lupaş [27], o las de los polinomios ortogonales asociados a la medida de Marchenko-Pastur [73], entre otras familias de matrices. En esta Tesis, se ha realizado una búsqueda sistemática de posibles subclases de matrices totalmente positivas para las que se puedan encontrar con alta precisión relativa los parámetros de la factorización bidiagonal y, en consecuencia, se puedan resolver con alta precisión relativa los problemas algebraicos mencionados. Además, esto se ha conseguido también para nuevas subclases de matrices totalmente positivas importantes en las aplicaciones, sobre todo en los campos de combinatoria y polinomios ortogonales. Concretamente, para la clase de matrices de Laguerre (en el artículo [17], que es presentado en la página 33), que son las matrices de colocación correspondientes a los polinomios de Laguerre (familia clásica de polinomios ortogonales), así como con las matrices de Bessel (en el artículo [16], que es presentado en la página 45), que son las matrices de colocación correspondientes a los polinomios de Bessel, también con matrices de Pascal generalizadas (en el artículo [18], que es presentado en la página 107) así como con matrices de q -enteros (en el artículo [20], que es presentado en la página 119). También incluimos aplicaciones sobre matrices totalmente positivas que surgen en problemas de diseño geométrico asistido por ordenador, en particular en relación con propiedades óptimas de las matrices de colocación de las B-bases normalizadas (bases con óptimas propiedades de preservación de forma [9] y con otras propiedades óptimas [28]) y sus productos de Kronecker, obteniendo nuevas propiedades de optimalidad (en el artículo [19], que es presentado en la página 141). Finalmente, también se obtienen caracterizaciones y resultados de alta precisión relativa para las P-matrices Toeplitz tridiagonales (en el artículo [21], que es presentado en la página 149).

Otra de las clases de matrices relacionada con algoritmos de alta precisión relativa es la clase de las M-matrices. Recordemos que las M-matrices (en el caso no singular) son matrices que tienen elementos diagonales positivos, elementos extradiagonales no positivos y cuya inversa es no negativa. Constituyen una clase de matrices que ha dado lugar a importantes aplicaciones en Análisis Numérico, en programación lineal, en sistemas dinámicos y en economía (véase [6]). En el caso de matrices relacionadas con la dominancia diagonal y las M-matrices, juega un papel muy importante en la búsqueda de algoritmos de alta precisión relativa la obtención y análisis de las llamadas descomposiciones reveladoras del rango, que consisten en una factorización XDY^T donde D es diagonal no singular y X e Y son matrices bien condicionadas. Si se conoce con alta precisión relativa una descomposición reveladora del rango de una matriz, entonces se pueden obtener con alta precisión relativa sus valores singulares mediante un algoritmo que obtuvieron J. Demmel y colaboradores [30]. Además, en [32, 87] se obtuvieron descomposiciones reveladoras del rango con alta precisión relativa para las M-matrices diagonal dominantes, usando como parámetros conocidos de partida

las sumas de filas y los elementos extradiagonales. En esta memoria, se estudian clases de matrices más generales que las M-matrices diagonal dominantes para las que habrá que, en primer lugar, encontrar la parametrización adecuada para poder obtener algoritmos con alta precisión relativa. Con objeto de obtener dichas descomposiciones reveladoras del rango, un instrumento fundamental es la búsqueda de adecuadas estrategias de pivotaje para las mencionadas clases de matrices, ya que la eliminación gaussiana con adecuadas estrategias de pivotaje puede proporcionar descomposiciones LDU que son descomposiciones reveladoras del rango (véase [32, 87]). Un ejemplo de clase de matrices estudiada en la memoria para la que se ha encontrado una parametrización que garantiza cálculos algebraicos con alta precisión relativa (tanto para calcular la inversa como el determinante) y que generaliza las M-matrices diagonal dominantes es la clase de matrices de Nekrasov (en los artículos [79] y [82], presentados en las páginas 63 y 173, respectivamente). Esta clase de matrices ha sido muy útil recientemente [41] en el estudio de cotas de error en el problema de complementariedad lineal (problema de optimización con aplicaciones muy importantes). El problema de complementariedad lineal tiene solución única cuando la matriz asociada es una P-matriz [11], es decir, cuando todos sus menores principales son positivos. Tanto la matrices totalmente positivas no singulares como las M-matrices no singulares son P-matrices. Para las Z-matrices de Nekrasov también obtenemos nuevas cotas para la norma de su inversa, que a su vez permiten obtener nuevas cotas de error en el problema de complementariedad lineal (en el artículo [81], que es presentado en la 79). Otra clase de P-matrices es la de las B-matrices [85], para la que también se garantiza el cálculo del determinante con alta precisión relativa (en el artículo [82], que es presentado en la página 173). Además, en la memoria se analizan las B_{π}^R -matrices [83], que forman otra clase de P-matrices más general que la de las B-matrices y para ellas también se obtienen cotas de error en el problema de complementariedad lineal (en el artículo [83], que es presentado en la página 189). Finalmente, estas clases de matrices relacionadas con la dominancia diagonal tienen una aplicación potencial en teoría de hipermatrices. El interés por las hipermatrices (o tensores, véase [90]) ha aumentado recientemente de manera considerable por sus aplicaciones en el tratamiento computacional de “Big Data” (Datos Masivos), campo de gran interés en la actualidad en tecnologías de la información y de la comunicación. Esto ha dado lugar a un gran número de publicaciones que intentan extender al contexto de tensores técnicas y resultados que han sido útiles en el caso matricial. Uno de los problemas en que nos centramos es la búsqueda de condiciones suficientes de tensores simétricos para que sean definidos positivos, cuestión muy importante en problemas de optimización. Concretamente, abordamos estas cuestiones en la extensión de las B_{π}^R -matrices a tensores (en el artículo [80], que es presentado en la página 91).

Este trabajo está estructurado en cuatro partes del modo siguiente. La primera parte está compuesta por esta Introducción y el Capítulo 3. El Capítulo 3 introduce conceptos y resultados básicos de la Tesis. Concretamente, se introducen las P-matrices, M-matrices y matrices totalmente positivas, así como las bases totalmente positivas. También conceptos básicos de tensores y la alta precisión relativa. En la segunda parte, presentamos en las páginas 33, 45, 63, 79, 91, 107, 119, 141, 149, 173 y 189 los artículos [17], [16], [79], [81], [80], [18], [20], [19], [21], [82] y [83] respectivamente, que pertenecen al compendio de publicaciones de esta tesis. La tercera parte está compuesta por el Capítulo 4 y el Capítulo 5. El objetivo de estos capítulos es justificar la unidad temática de las publicaciones y presentar

los resultados principales de estos artículos. En el Capítulo 4 comenzamos mostrando los resultados que permiten los cálculos con alta precisión relativa para las siguientes subclases de matrices totalmente positivas: matrices de Laguerre, matrices de Bessel, matrices de Pascal generalizadas y matrices de q -enteros. También se muestran las propiedades óptimas de los productos tensoriales de B-bases normalizadas. Finalmente, se caracterizan las P-matrices Toeplitz tridiagonales y se garantizan para ellas cálculos con alta precisión relativa. El Capítulo 5 está dedicado a M-matrices y otras clases de matrices relacionadas, considerando tanto los problemas de la obtención de algoritmos con alta precisión relativa como otros problemas relacionados con dichas clases de matrices. Comenzamos considerando la alta precisión relativa para la inversa y los determinantes de las Z-matrices Nekrasov con entradas diagonales positivas y para los determinantes de las B-matrices. A continuación, se estudian cotas para matrices de Nekrasov tanto para la norma de su inversa como para el error de los problemas de complementariedad lineal correspondientes. Dichas cotas se basan en el uso de una matriz de escalado que transforma las matrices en diagonal dominantes. La última sección considera B_{π}^R -matrices, que generalizan las B-matrices. Se estudian cotas para el error de los problemas de complementariedad lineal correspondientes y también su extensión a tensores y problemas relacionados. Finalmente, la parte cuarta incluye las conclusiones de la Tesis así como posible trabajo futuro relacionado con la investigación realizada en la Tesis.

Chapter 3

Background

This first chapter provides the foundation for the publications that compose the core of this thesis. It presents a short survey on the classes of structured matrices object of study and the tools used to achieve accurate computations. It is organized in five sections. The first section introduces basic notation and the class of P -matrices, which includes most of our examples. The second section is devoted to M -matrices, H -matrices and Gaussian elimination. The third section presents totally positive matrices, Neville elimination and the bidiagonal decomposition. The fourth section introduces tensors (or hypermatrices) and presents some of the basic concepts used on the study of the extension of matrices to a higher order case. Finally, we introduce the concepts of finite precision arithmetic, roundoff error, high relative accuracy and we identify the right tools that grant accurate results for the classes of diagonally dominant M -matrices and nonsingular totally positive matrices.

3.1 Basic concepts and P -matrices

Let us start by introducing some notation that will be used through this dissertation. In general, we will assume that matrices are square unless stated otherwise. For example, $A := (a_{ij})_{1 \leq i, j \leq n}$ will denote an $n \times n$ matrix, where a_{ij} is the entry in the place (i, j) of A . The first important case of structured matrices is given by diagonal matrices. We say that $D := (d_{ij})_{1 \leq i, j \leq n}$ is diagonal if $d_{ij} = 0$ whenever $i \neq j$. In this case, it is possible to represent D in terms of its diagonal entries, so we will use the notation $D := \text{diag}(d_1, \dots, d_n)$, where $d_i := d_{ii}$ for $i = 1, \dots, n$.

Let us denote by $Q_{k,n}$ the set of strictly increasing sequences of k integers chosen from $\{1, \dots, n\}$. Let $\alpha = (\alpha_1, \dots, \alpha_k)$ and $\beta = (\beta_1, \dots, \beta_k)$ be two sequences of $Q_{k,n}$. Then $A[\alpha|\beta]$ denotes the $k \times k$ submatrix of A formed using the rows numbered by $\alpha_1, \dots, \alpha_k$ and the columns numbered by β_1, \dots, β_k . Whenever $\alpha = \beta$, the submatrix $A[\alpha|\alpha]$ is called a principal submatrix and it is denoted by $A[\alpha]$, and $\det A[1, \dots, k]$ is called a leading principal minor of A . For each $\alpha \in Q_{k,n}$, the dispersion number $d(\alpha)$ is defined by

$$d(\alpha) := \alpha_k - \alpha_1 - (k - 1). \quad (3.1)$$

So, α consists of consecutive integers if and only if $d(\alpha) = 0$.

Let us denote by $E_i(x)$, with $i = 2, \dots, n$, the $n \times n$ lower elementary bidiagonal matrix whose $(i, i - 1)$ entry is x :

$$E_i(x) = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & x & 1 & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix}. \quad (3.2)$$

In particular, $E_i(x)$ can be identified by its 2×2 principal submatrix using the rows and columns with indices $i - 1$ and i . The matrix $E_i^T(x) := (E_i(x))^T$ is called *upper elementary bidiagonal matrix*.

As we anticipated earlier, the matrices studied in this dissertation are closely related to the class of P -matrices.

Definition 3.1. A square matrix is a P -matrix if all its principal minors are positive.

The next proposition characterizes a P -matrix in terms of the positivity of the real eigenvalues of its principal submatrices.

Proposition 3.2. (cf. 2.5.6.5 in p. 120 of [53]) An $n \times n$ matrix A is a P -matrix if and only if every real eigenvalue of every principal submatrix of A is positive.

The class of P -matrices admits different characterizations. Some of their characterizations directly relate them to their applications. For example, we can characterize P -matrices in terms of a common problem arising in linear programming: the *linear complementarity problem* (LCP). Given an $n \times n$ real matrix A and a vector $q \in \mathbb{R}^n$, the linear complementarity problem $\text{LCP}(A, q)$ consists of finding, if possible, vectors $x \in \mathbb{R}^n$ satisfying

$$Ax + q \geq 0, \quad x \geq 0, \quad x^T(Ax + q) = 0, \quad (3.3)$$

where the inequalities are entrywise. As the following theorem states, the existence of a unique solution of (3.3) characterizes P -matrices (p. 275 of [6]).

Theorem 3.3. The matrix $M = (m_{ij})_{1 \leq i, j \leq n}$ is a P -matrix if and only if the linear complementarity problem $\text{LCP}(M, q)$ (3.3) has a unique solution x^* for every $q \in \mathbb{R}^n$.

Let us also recall that an $n \times n$ real matrix A is called a Q -matrix if $\text{LCP}(A, q)$ has a solution for any $q \in \mathbb{R}^n$ (see p. 276 of [6]).

The problem of characterizing the class of P -matrices and the development of practical criteria for identifying this class has attracted a lot of attention. The problem of recognizing whether a given matrix is a P -matrix is called the P -problem. It has been shown that the P -problem is CO-NP-complete [12] and that, in general, the P -problem seems inevitably of exponential time complexity [95]. However, the complexity of recognizing some important subclasses of P -matrices, such as nonsingular M -matrices [86], nonsingular totally positive matrices [45] or B -matrices [85], has polynomial complexity.

M-matrices and totally positive matrices arise in many applications and, joint with some related classes of matrices, are the main topic of this dissertation. The next sections will introduce these classes of matrices and the properties and techniques that will be key to achieving accurate computations with them.

3.2 *M*-matrices, related matrices and Gaussian elimination

We say that a real matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ is a *Z-matrix* if all its off-diagonal entries are nonpositive, i.e., $a_{ij} \leq 0$ for all (i, j) such that $i \neq j$.

Definition 3.4. A *Z-matrix* $A = (a_{ij})_{1 \leq i, j \leq n}$ is called an *M-matrix* if it can be expressed in the form: $A = sI - B$, with $B \geq 0$ and $s \geq \rho(B)$ (where $\rho(B)$ is the spectral radius of B).

If $s > \rho(B)$, then A is a nonsingular *M-matrix*. It is possible to achieve accurate computations with *M-matrices* when they are also diagonally dominant. In fact, the condition of diagonal dominance is closely related to this class of matrices and can be used to characterize it (see Theorem 3.7). Let us first introduce the definition of diagonally dominant matrices.

Definition 3.5. We say that $A = (a_{ij})_{1 \leq i, j \leq n}$ is a (*row*) *diagonally dominant* (DD) matrix if

$$|a_{ii}| \geq \sum_{j \neq i} |a_{ij}| \quad \text{for all } i = 1, \dots, n. \quad (3.4)$$

If (3.4) holds strictly for all $i = 1, \dots, n$, then we say that A is *strictly diagonally dominant* (SDD).

As an immediate consequence of the definition, strictly diagonally dominant matrices are nonsingular. The following result gives a useful relationship between *M-matrices* and nonsingular *M-matrices*.

Lemma 3.6. (cf. Lemma 4.1 in chapter 6 of [6]) Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a *Z-matrix*. Then A is an *M-matrix* if and only if $A + \varepsilon I$ is a nonsingular *M-matrix* for all scalars $\varepsilon > 0$.

Nonsingular *M-matrices* admit a lot of different characterizations (Theorem (2.3) of chapter 6 of [6] presents 50 different characterizations). Nonsingular *M-matrices* are well-known in many applications [6] because their inverses are entrywise nonnegative, i.e., $A^{-1} \geq 0$.

Theorem 3.7. Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a *Z-matrix*. Then the following conditions are equivalent:

- i) A is a nonsingular *M-matrix*.
- ii) Every real eigenvalue of A is positive.
- iii) Every principal minor of A is positive.
- iv) Every leading principal minor of A is positive.

- v) A is nonsingular and A^{-1} is nonnegative ($A^{-1} \geq 0$).
- vi) The diagonal entries of A are positive and there exists a diagonal matrix D such that AD is a strictly diagonally dominant matrix.
- vii) $A = LU$, where L is a lower triangular matrix, U is an upper triangular matrix and the all diagonal entries of L and U are positive.
- viii) A is nonsingular and $A + D$ is nonsingular for every diagonal matrix D with positive diagonal entries.

Let us observe that, since i) implies iii) in Theorem 3.7, nonsingular M -matrices are also P -matrices. In order to achieve accurate computations with DD M -matrices, we will use an adapted version of the well-known method of *Gaussian elimination*. Given a nonsingular matrix $A = (a_{ij})_{1 \leq i, j \leq n}$, Gaussian elimination is a method used to produce zeros below the main diagonal of A . This procedure consists of $n - 1$ steps leading to the following sequence of matrices:

$$A = A^{(1)} \rightarrow \tilde{A}^{(1)} \rightarrow A^{(2)} \rightarrow \tilde{A}^{(2)} \rightarrow \dots \rightarrow A^{(n)} = \tilde{A}^{(n)} = DU, \quad (3.5)$$

where $A^{(k)}$ has zero entries under the main diagonal in its first $k - 1$ columns and DU is an upper triangular matrix. We obtain $\tilde{A}^{(k)}$ from $A^{(k)}$ by reordering the rows and/or columns using a pivoting strategy. A *pivoting strategy* in Gaussian elimination consists on a reordering of the rows and/or columns of A at every step to choose the pivot entry that will be used to produce zeros on the next step. In (3.5), the pivoting strategy is applied to transform $A^{(k)}$ into $\tilde{A}^{(k)}$. Two well-known pivoting strategies are partial pivoting and complete pivoting. *Partial pivoting* consists on a reordering of rows at every step looking for the entry with the largest absolute value in the column $A^{(k)}[k, \dots, n|k]$ where we want to produce zeros at the next step. *Complete pivoting* consists on a reordering of rows and columns at every step, looking for the entry with the largest absolute value in the whole submatrix $A^{(k)}[k, \dots, n]$. For all pivoting strategies, it is needed to choose a nonzero pivot $\tilde{a}_{kk}^{(k)}$. Pivoting strategies that apply the same permutation to the rows and columns at every step are called *symmetric* pivoting strategies.

The entry $\tilde{a}_{kk}^{(k)}$ in $\tilde{A}^{(k)}$ is the pivot chosen by the pivoting strategy and it will be used to produce zeros below the main diagonal in the k -th step. For that purpose, we subtract multiples of the k -th row to the rows beneath it. Hence, the resulting matrix $A^{(k+1)} = (a_{ij}^{(k+1)})_{1 \leq i, j \leq n}$ is given by

$$a_{ij}^{(k+1)} = \begin{cases} \tilde{a}_{ij}^{(k)}, & \text{if } 1 \leq i \leq k, \\ \tilde{a}_{ij}^{(k)} - \frac{\tilde{a}_{ik}^{(k)}}{\tilde{a}_{kk}^{(k)}} \tilde{a}_{kj}^{(k)}, & \text{if } k < i \leq n. \end{cases}$$

We can also compute the inverse of A using the well-known method of Gauss-Jordan. We can describe Gauss-Jordan (without pivoting) through the following sequence of steps

$$A = A^{(1)} \rightarrow A^{(2)} \rightarrow \dots \rightarrow A^{(n)} = DU \rightarrow A^{(n+1)} \rightarrow \dots \rightarrow A^{(2n-1)} = D \rightarrow A^{(2n)} = I,$$

where the first $n - 1$ steps consist of using Gaussian elimination without pivoting to produce zeros below the main diagonal of A . Then on the next $n - 1$ steps we apply Gaussian elimination without pivoting to produce zeros over the diagonal of A (this procedure can be seen as applying Gaussian elimination to $(A^{(n)})^T$). Then we finish multiplying $A^{(2n-1)}$ by a diagonal matrix so that we obtain the identity matrix. If we apply the same elementary operations to the identity matrix (i.e., to $B = B^{(1)} = I$), when we obtain the result $A^{(2n)} = I$ we also compute $B^{(2n)} = A^{-1}$. Analogously to Gaussian elimination, Gauss-Jordan can also be carried out using a row pivoting strategy.

Given a complex matrix $A = (a_{ij})_{1 \leq i, j \leq n}$, its *comparison matrix* $\mathcal{M}(A) = (\tilde{a}_{ij})_{1 \leq i, j \leq n}$ satisfies that $\tilde{a}_{ii} := |a_{ii}|$ and $\tilde{a}_{ij} := -|a_{ij}|$ for all $j \neq i$ and $i, j = 1, \dots, n$. Let us notice that a comparison matrix is always a *Z*-matrix. The next definition introduces a class that contains *M*-matrices.

Definition 3.8. We say that a complex matrix A is an *H*-matrix if its comparison matrix $\mathcal{M}(A) = (\tilde{a}_{ij})_{1 \leq i, j \leq n}$ is a nonsingular *M*-matrix.

A wider class of *H*-matrices can be defined without requiring the nonsingularity of its comparison matrix. For a study on a more general definition of *H*-matrices, see [8]. Nonsingular *M*-matrices A can be characterized in terms of the existence of a scaling diagonal matrix such as the product AD is a strictly diagonally dominant matrix (see *vi*) in Theorem 3.7). In fact, this property characterizes the class of *H*-matrices (see p. 124 of [53]) as it can be seen in the following theorem.

Theorem 3.9. $A = (a_{ij})_{1 \leq i, j \leq n}$ is an *H*-matrix if and only if there exists a diagonal matrix D such that AD is a strictly diagonally dominant matrix.

The definition of diagonal dominance does not depend on the sign structure of a matrix. Hence, finding a diagonal matrix D as described in Theorem 3.9 for an *M*-matrix M means finding a scaling matrix for any *H*-matrix A satisfying that $\mathcal{M}(A) = M$. This connection helped us developing error bounds for the LCP and infinity norm bounds for the inverses of Nekrasov matrices in [81] using as a starting point a diagonal scaling matrix obtained for Nekrasov *Z*-matrices with positive diagonal entries in [79]. *Nekrasov Z*-matrices with positive diagonal entries are a subclass of *M*-matrices that contains SDD *M*-matrices. Let $N := \{1, \dots, n\}$. Given a complex matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ with $a_{ii} \neq 0$ for all $i \in N$, we define

$$h_1(A) := \sum_{j \neq 1} |a_{1j}|, \quad h_i(A) := \sum_{j=1}^{i-1} |a_{ij}| \frac{h_j(A)}{|a_{jj}|} + \sum_{j=i+1}^n |a_{ij}|, \quad i = 2, \dots, n. \quad (3.6)$$

We say that A is a *Nekrasov matrix* if $|a_{ii}| > h_i(A)$ for all $i \in N$ (see [13–15, 94]). Nekrasov matrices are nonsingular *H*-matrices. In particular, a Nekrasov *Z*-matrix with positive diagonal entries is a nonsingular *M*-matrix.

Another subclass of *P*-matrices closely related to *M*-matrices is that of *B*-matrices. The condition that defines *B*-matrices was studied in [51], where it was shown that this class of matrices has positive determinant. In [85], the class was named *B*-matrices and it was shown that it is a subclass of *P*-matrices (Corollary 2.6 of [85]). The class was first studied in the development of localization criteria for the eigenvalues of a real matrix. Since then,

the class of B -matrices has been used in more applications such as the development of error bounds for the LCP [39] and the development of simple criteria for identifying subclasses of P -matrices. Because of that, the class of B -matrices has also been extended to characterize wider subclasses of P -matrices [78] and to the higher order case due to the interest of finding easy recognition conditions for subclasses of P -tensors based on their entries. Let us now recall the definition of B -matrix [85].

Definition 3.10. A square real matrix $A := (a_{ij})_{1 \leq i, j \leq n}$ with positive row sums is a B -matrix if all its off-diagonal elements are bounded above by the corresponding row means, i.e., for all $i = 1, \dots, n$,

$$\sum_{j=1}^n a_{ij} > 0, \quad \frac{1}{n} \left(\sum_{k=1}^n a_{ik} \right) > a_{ij} \quad \forall j \neq i. \quad (3.7)$$

B -matrices admit a decomposition that relates them to SDD Z -matrices (and so, to SDD M -matrices). Given a real matrix $B = (b_{ij})_{1 \leq i, j \leq n}$, we define for each $i = 1, \dots, n$, $r_i^+ := \max_{j \neq i} \{0, b_{ij}\}$. Then B can be decomposed in the form

$$B = B^+ + C, \quad (3.8)$$

$$B^+ = \begin{pmatrix} b_{11} - r_1^+ & \dots & b_{1n} - r_1^+ \\ \vdots & & \vdots \\ b_{n1} - r_n^+ & \dots & b_{nn} - r_n^+ \end{pmatrix}, \quad C = \begin{pmatrix} r_1^+ & \dots & r_1^+ \\ \vdots & & \vdots \\ r_n^+ & \dots & r_n^+ \end{pmatrix}. \quad (3.9)$$

This decomposition gives the following characterization of B -matrices.

Proposition 3.11. *Let M be a real matrix and let us write $M = B^+ + C$ following (3.8) and (3.9). Then M is a B -matrix if and only if the matrix B^+ is an SDD Z -matrix.*

The decomposition defined by (3.8) and (3.9) gives a good starting point for finding extensions of the class of B -matrices. For example, we say that B is a B -Nekrasov matrix if the matrix B^+ given by (3.8) and (3.9) is a Nekrasov Z -matrix with positive diagonal entries (see [43]) or that B is a MB -matrix if B^+ is a nonsingular M -matrix [65]. Both B -Nekrasov matrices and MB -matrices are also P -matrices. And, B -matrices are contained in B -Nekrasov matrices while B -Nekrasov matrices form a subclass of MB -matrices.

3.3 Totally positive matrices and bases

The second class of matrices that we have studied is known as *totally positive* matrices [89] or also *totally nonnegative* matrices [36].

Definition 3.12. We say that a matrix is *totally positive* (TP) if all its minors are nonnegative and that it is *strictly totally positive* (STP) if all its minors are positive.

Nonsingular TP matrices are in particular P -matrices (cf. Corollary 3.8 of [3]). Checking the definition of a TP matrix or STP matrix would require the computation of many determinants. However, because of the strong structure of these classes, it is far easier to check total positivity and strict total positivity than checking that a matrix is a P -matrix. For instance, the elimination procedure called Neville elimination (3.11) gives us a method to check if a $n \times n$ matrix is TP (or STP) with $\mathcal{O}(n^3)$ elementary operations (see [45]). Moreover, there is also a simple sufficient condition based on the matrix entries to assure total positivity. Given a positive matrix $A = (a_{ij})_{1 \leq i, j \leq n}$, the following condition is sufficient for its total positivity (see [56] or section 2.6 of [89]):

$$a_{ij}a_{i+1, j+1} \geq 4 \cos^2 \left(\frac{\pi}{n+1} \right) a_{i, j+1}a_{i+1, j}, \quad (3.10)$$

with $i, j = 1, \dots, n-1$. If all these inequalities are strict, then A is STP.

In the previous section, we have commented that Gaussian elimination is an important tool to achieve accurate computations with M -matrices. In this section, we introduce an alternative procedure to Gaussian elimination, called Neville elimination (NE), which is going to be fundamental to our study of TP matrices. As we have just stated, Neville elimination (NE) is an elimination procedure used to produce zeros below the main diagonal of a matrix. NE makes zeros in a column of a matrix by adding to each row an appropriate multiple of the previous one (see [45]). Given a nonsingular matrix $A = (a_{ij})_{1 \leq i, j \leq n}$, the NE procedure consists of $n-1$ steps, leading to a sequence of matrices as follows:

$$A = A^{(1)} \rightarrow \tilde{A}^{(1)} \rightarrow A^{(2)} \rightarrow \tilde{A}^{(2)} \rightarrow \dots \rightarrow A^{(n)} = \tilde{A}^{(n)} = U, \quad (3.11)$$

with U an upper triangular matrix.

In (3.11), $\tilde{A}^{(t)}$ is obtained from the matrix $A^{(t)}$ by moving to the bottom the rows with a zero entry in column t below the main diagonal, if necessary. The matrix $A^{(t+1)}$, $t = 1, \dots, n-1$, is obtained from $\tilde{A}^{(t)}$ by computing

$$a_{ij}^{(t+1)} = \begin{cases} \tilde{a}_{ij}^{(t)} - \frac{\tilde{a}_{it}^{(t)}}{\tilde{a}_{i-1,t}^{(t)}} \tilde{a}_{i-1,j}^{(t)}, & \text{if } t \leq j \leq n, t+1 \leq i \leq n \text{ and } \tilde{a}_{i-1,t}^{(t)} \neq 0, \\ \tilde{a}_{ij}^{(t)}, & \text{otherwise,} \end{cases} \quad (3.12)$$

for all $t \in \{1, \dots, n-1\}$. The entry

$$p_{ij} := \tilde{a}_{ij}^{(j)}, \quad 1 \leq j \leq i \leq n, \quad (3.13)$$

is the (i, j) pivot of the NE of A , and the pivots p_{ii} are called *diagonal pivots*. If all the pivots are nonzero, then we can use the following expression to compute them (Lemma 2.6 of [45]):

$$p_{i1} = a_{i1}, \quad 1 \leq i \leq n, \\ p_{ij} = \frac{\det A[i-j+1, \dots, i|1, \dots, j]}{\det A[i-j+1, \dots, i-1|1, \dots, j-1]}, \quad 1 \leq j \leq i \leq n. \quad (3.14)$$

Finally, the number m_{ij} defined as

$$m_{ij} = \begin{cases} \frac{\tilde{a}_{ij}^{(j)}}{\tilde{a}_{i-1,j}^{(j)}} = \frac{p_{ij}}{p_{i-1,j}}, & \text{if } \tilde{a}_{i-1,j}^{(j)} \neq 0, \\ 0, & \text{if } \tilde{a}_{i-1,j}^{(j)} = 0, \end{cases} \quad (3.15)$$

is called the (i, j) multiplier of NE of A , where $1 \leq j < i \leq n$.

NE is a method with a lot of useful properties when applied to TP matrices. If A is a nonsingular TP matrix, then no rows exchanges are needed when applying NE and so, in this case, $A^{(t)} = \tilde{A}^{(t)}$ for all t . In fact, in Corollary 5.5 of [45] the following characterization of nonsingular TP matrices was provided based on NE.

Theorem 3.13. *Let A be a nonsingular matrix. Then A is TP if and only if there are no row exchanges in the Neville elimination of A and U^T and the pivots of both Neville elimination procedures are nonnegative.*

So, by Theorem 3.13 Neville elimination characterizes nonsingular TP matrices. Moreover, the pivots and multipliers associated to this elimination procedure gives us a different representation of this class of matrices: the bidiagonal decomposition. In [47], it was seen that nonsingular TP matrices can be expressed as a unique bidiagonal decomposition.

Theorem 3.14. *(cf. Theorem 4.2 of [47]). Let A be a nonsingular $n \times n$ TP matrix. Then A admits a decomposition of the form*

$$A = F_{n-1} \cdots F_1 D G_1 \cdots G_{n-1}, \quad (3.16)$$

where F_i and G_i , $i \in \{1, \dots, n-1\}$, are the lower and upper triangular nonnegative bidiagonal matrices given by

$$F_i = \begin{pmatrix} 1 & & & & & & & \\ 0 & 1 & & & & & & \\ & & \ddots & & & & & \\ & & & \ddots & & & & \\ & & & & 0 & & & \\ & & & & & 1 & & \\ & & & & & & m_{i+1,1} & \\ & & & & & & & \ddots & \\ & & & & & & & & \ddots & \\ & & & & & & & & & m_{n,n-i} & \\ & & & & & & & & & & 1 \end{pmatrix}, \quad G_i^T = \begin{pmatrix} 1 & & & & & & & \\ 0 & 1 & & & & & & \\ & & \ddots & & & & & \\ & & & \ddots & & & & \\ & & & & 0 & & & \\ & & & & & \tilde{m}_{i+1,1} & & \\ & & & & & & 1 & \\ & & & & & & & \ddots & \\ & & & & & & & & \ddots & \\ & & & & & & & & & \tilde{m}_{n,n-i} & \\ & & & & & & & & & & 1 \end{pmatrix}, \quad (3.17)$$

and D a diagonal matrix $\text{diag}(p_{11}, \dots, p_{nn})$ with positive diagonal entries. If, in addition, the entries m_{ij} , \tilde{m}_{ij} satisfy

$$m_{ij} = 0 \Rightarrow m_{hj} = 0 \quad \forall h > i$$

and

$$\tilde{m}_{ij} = 0 \Rightarrow m_{ik} = 0 \quad \forall k > j,$$

then the decomposition given by (3.16) and (3.17) is unique.

By Theorems 4.1 and 4.2 of [47] we also know that, for $1 \leq j < i \leq n$, m_{ij} and p_{ii} in the bidiagonal decomposition given by (3.16) with (3.17) are the multipliers and the diagonal

pivots when applying the NE to A and, using the arguments of p. 116 of [47], \tilde{m}_{ij} are the multipliers when applying the NE to A^T .

The bidiagonal decomposition gives a different representation of a TP matrix. If we look at (3.16) and (3.17), we see that for an $n \times n$ matrix we have a representation in terms of n^2 parameters (the entries p_{ii} , m_{ij} and \tilde{m}_{ij}), which plays a key role to derive accurate computations with these matrices. In [60], Plamen Koev introduced the following abbreviated matrix notation for the bidiagonal decomposition of a matrix:

$$(\mathcal{BD}(A))_{ij} = \begin{cases} m_{ij}, & \text{if } i > j, \\ \tilde{m}_{ji}, & \text{if } i < j, \\ p_{ii}, & \text{if } i = j. \end{cases} \quad (3.18)$$

If A is a TP matrix, then A^T is also TP. Transposing formula (3.16) of Theorem 3.14 we obtain the unique bidiagonal decomposition of A^T :

$$A^T = G_{n-1}^T \cdots G_1^T D F_1^T \cdots F_{n-1}^T,$$

where F_i and G_i , $i \in \{1, \dots, n-1\}$, are the lower and upper triangular nonnegative bidiagonal matrices given in formula (3.17). Then it can be checked that

$$\mathcal{BD}(A^T) = \mathcal{BD}(A)^T. \quad (3.19)$$

Knowing the bidiagonal decomposition of nonsingular totally positive matrices accurately will allow us to solve many algebraic problems with them with high relative accuracy [59, 60]. We will elaborate on this topic in Section 3.5 and it will be one of the main ideas used in Chapter 4 to achieve accurate computations.

Given a system of functions $u = (u_0, \dots, u_n)$ defined on an interval $I \subseteq \mathbb{R}$, the collocation matrix of u at $t_0 < \dots < t_n$ in I is given by

$$M \begin{pmatrix} u_0, & \dots, & u_n \\ t_0, & \dots, & t_n \end{pmatrix} := \begin{pmatrix} u_0(t_0) & \cdots & u_n(t_0) \\ \vdots & & \vdots \\ u_0(t_n) & \cdots & u_n(t_n) \end{pmatrix} \quad (3.20)$$

We say that the system of functions u is *totally positive* if all its collocation matrices are totally positive. If $\sum_{i=0}^n u_i(t) = 1$ for all $t \in I$, then we say that the system is *normalized*. Normalized totally positive (NTP) systems play an important role in Computer Aided Geometric Design (CAGD) because of their shape preserving properties [88].

A totally positive system of linearly independent functions u defined on $I \subseteq \mathbb{R}$ is said to be a B -basis if all totally positive bases v of the space generated by u satisfy that

$$(v_0, \dots, v_n) = (u_0, \dots, u_n)A, \quad \text{with } A \text{ a nonsingular TP matrix} \quad (3.21)$$

Among all NTP bases of a space, the basis with optimal shape preserving properties is the *normalized B-basis* [9]. Hence, collocation matrices of normalized B -bases are an important example of TP matrices. Some examples of well-known B -bases are:

- In the space of polynomials of degree less than or equal to n on the interval $[0, 1]$, the normalized B -basis is the *Bernstein basis* $(b_i^n)_{0 \leq i \leq n}$. This basis is formed by the Bernstein polynomials of degree n defined as

$$b_i^n(t) := \binom{n}{i} (1-t)^{n-i} t^i, \quad i = 0, 1, \dots, n. \quad (3.22)$$

- In the space of polynomials of degree less than or equal to n on the interval $[0, \infty)$, the B -basis is given by the monomial basis $(x^i)_{0 \leq i \leq n}$. This space has no normalized totally positive basis. The collocation matrices of the monomial basis are also known as *Vandermonde matrices*. The Vandermonde matrix is defined to be the matrix $V(x_1, \dots, x_n) := (x_i^{j-1})_{1 \leq i, j \leq n}$ and all its minors are positive as long as it is defined on increasingly ordered positive nodes $0 < x_1 < x_2 < \dots < x_n$.
- Let us consider a sequence $(w_i)_{0 \leq i \leq n}$ of positive weights. Then the system of functions (r_0^n, \dots, r_n^n) defined on the interval $[0, 1]$ by

$$r_i^n(t) := \frac{w_i b_i^n(t)}{\sum_{j=0}^n w_j b_j^n(t)}, \quad i = 0, \dots, n,$$

is called the rational Bernstein basis and it is the normalized B -basis of the corresponding spanned space of functions. Both the Bernstein basis $(b_i^n)_{0 \leq i \leq n}$ and the rational Bernstein basis can also be defined on any closed interval $[a, b]$.

3.4 Basic concepts and definitions about tensors

We have considered the extension of some of the studied classes of matrices to the higher dimensional case. The concept of *tensor* or *hypermatrix* is introduced as a higher order generalization of matrices to develop an extension of matrix theory to be used with multi-indexed datasets. We are going to introduce the concept of tensor and give some basic definitions of important classes of tensors. Some of these definitions are extensions of the definitions introduced in the previous sections for matrices.

A tensor (or hypermatrix) $\mathcal{A} = (a_{i_1 \dots i_m})$ is a multi-array of entries $a_{i_1 \dots i_m}$ where $i_j = 1, \dots, n_j$ for $j = 1, \dots, m$. We will consider real m th order n -dimensional tensors, that is, the case where $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m, n]}$ is a multi-array of real entries $a_{i_1 \dots i_m} \in \mathbb{R}$ with $i_k \in N := \{1, \dots, n\}$ for $k = 1, \dots, m$. We say that \mathcal{A} is a *symmetric* tensor if its entries are invariant under any permutation of its indices. Let us consider the set of entries $a_{ii_2 \dots i_m}$ for $i, i_2, \dots, i_m \in N$ as the i -th row of \mathcal{A} . Then, the i -th row sum of \mathcal{A} is given by

$$R_i(\mathcal{A}) := \sum_{i_2, \dots, i_m=1}^n a_{ii_2 \dots i_m}.$$

A tensor \mathcal{A} is called *diagonally dominant* if

$$|a_{i \dots i}| \geq \sum_{i_2, \dots, i_m \neq (i, \dots, i)}^n |a_{ii_2 \dots i_m}|, \quad i \in N. \quad (3.23)$$

Let us notice that this definition is an extension of diagonally dominant matrix given by (3.4) in Definition 3.5. If (3.23) holds strictly, then \mathcal{A} is called *strictly diagonally dominant*. We say that $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ is a *B-tensor* (*B₀-tensor*) if

$$R_i(\mathcal{A}) > 0 (\geq 0), \quad i \in N, \quad (3.24)$$

and

$$\frac{R_i(\mathcal{A})}{n^{m-1}} > a_{ij_2 \dots j_m} (\geq a_{ij_2 \dots j_m}), \quad \forall (j_2, \dots, j_m) \neq (i, \dots, i). \quad (3.25)$$

The definition of *B-tensor* (see p. 201 [90]) also extends the definition of *B-matrix* given in Definition 3.10. We say that a tensor is *nonnegative* if all its entries are nonnegative, and that it is a *Z-tensor* if all its off-diagonal entries are nonpositive. Let us also define the *identity tensor* \mathcal{I} , whose entries are ones on the main diagonal (i.e., entries such that $i_1 = \dots = i_m$) and zeros elsewhere. A tensor $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ is called an (a *strong*) *M-tensor* if there exists a nonnegative tensor $\mathcal{B} = (b_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ and a positive scalar $s \geq \rho(\mathcal{B})$ ($> \rho(\mathcal{B})$) such that $\mathcal{A} = s\mathcal{I} - \mathcal{B}$, where $\rho(\mathcal{B})$ is the *spectral radius* of \mathcal{B} (see page 15 of [90]). (Strictly) diagonally dominant *Z-tensors* are also (strong) *M-tensors* (as it is the case in the 2-dimensional case with SDD *Z-matrices* and nonsingular *M-matrices*).

A tensor \mathcal{A} is called *positive semidefinite* (*definite*) if for each (nonzero) $x \in \mathbb{R}^n$

$$\mathcal{A}x^m \geq 0 (> 0),$$

where $\mathcal{A}x^m = \sum_{i_1, \dots, i_m=1}^n a_{i_1 i_2 \dots i_m} x_{i_1} \cdots x_{i_m}$. Notice that there are not any nontrivial positive semidefinite tensors when m is odd. Let us recall that, given an m -th order tensor $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ and $x \in \mathbb{R}^n$, then $\mathcal{A}x^{m-1} \in \mathbb{R}^n$ is given by

$$(\mathcal{A}x^{m-1})_i := \sum_{i_2, \dots, i_m=1}^n a_{ii_2 \dots i_m} x_{i_2} \cdots x_{i_m}, \quad \text{for each } i = 1, \dots, n.$$

Definition 3.15. (see [34] or page 192 of [90]) A tensor $\mathcal{A} \in \mathbb{R}^{[m,n]}$ is called a *P-tensor* if for each nonzero $x \in \mathbb{R}^n$ there exists an index $i \in N$ such that

$$x_i^{m-1} (\mathcal{A}x^{m-1})_i > 0. \quad (3.26)$$

A tensor $\mathcal{A} \in \mathbb{R}^{[m,n]}$ is called a *P₀-tensor* if for each nonzero $x \in \mathbb{R}^n$ there exists some index $i \in N$ such that

$$x_i \neq 0 \text{ and } x_i^{m-1} (\mathcal{A}x^{m-1})_i \geq 0. \quad (3.27)$$

In [93] it was shown that, in the even order case, a symmetric tensor is positive definite (semidefinite) if and only if it is a *P-tensor* (*P₀-tensor*). The following result shows the close relationship between *M-tensors* and positive definiteness.

Proposition 3.16. (Theorem 4.1 of [99] and Lemma 3 of [64]) Let $\mathcal{A} \in \mathbb{R}^{[m,n]}$ be a symmetric *Z-tensor* and let m be even. Then

1. \mathcal{A} is positive definite if and only if \mathcal{A} is a strong M -tensor.
2. \mathcal{A} is positive semidefinite if and only if \mathcal{A} is an M -tensor.

The next section will introduce the basic concepts on error analysis for numerical methods and will show some of the right tools for achieving accurate computations with nonsingular totally positive matrices and nonsingular DD M -matrices.

3.5 High relative accuracy in numerical linear algebra

When studying a numerical algorithm, we should take into account the effect of errors in our computations. Following the classical reference [50], we can identify three main sources of errors in numerical computations: rounding, truncation and data uncertainty. In our work, we will focus on the effect of rounding errors and data uncertainty. Rounding error is an unavoidable consequence of working with a finite precision arithmetic. Its effect should be taken into account because a bad method could magnify these errors and result in inaccurate numerical solutions. For studying the effect of errors, this first question raises: How can we measure them? Let us suppose that we want to compute an approximation \hat{x} to a real number x . The first definition, the *absolute error*, measures the difference between \hat{x} and x .

Definition 3.17. The absolute error of \hat{x} is given by $E_{abs} := |x - \hat{x}|$.

The absolute error is a straightforward definition, but it presents a small “problem” for our analysis: it is scale dependent. Because of that, we prefer using the *relative error*.

Definition 3.18. The relative error of \hat{x} , with $x \neq 0$, is given by $E_{rel} := \frac{|x - \hat{x}|}{|x|}$.

For the case of a vector x and its approximation \hat{x} , we define the absolute and relative errors in terms of a vector norm $\|\cdot\|$.

Definition 3.19. The absolute error of \hat{x} is defined as $E_{abs} := \|x - \hat{x}\|$ and its relative error, whenever $x \neq 0$, is given by $E_{rel}(\hat{x}) := \frac{\|x - \hat{x}\|}{\|x\|}$.

If we are interested in the error of the smaller entries of a vector \hat{x} the relative error might not be informative, so we will also consider the *componentwise relative error*.

Definition 3.20. The componentwise relative error of \hat{x} , with $x_i \neq 0$, is given by $E_{comp} := \max_i \frac{|x_i - \hat{x}_i|}{|x_i|}$.

The absolute and relative errors of \hat{x} are called *forward errors*. Since we usually do not know the error obtained when we use a numerical method, we perform error analysis by developing upper bounds for this quantity. There is a different strategy, called *backward error analysis*, for studying how good is a computed solution. Analyzing the backward error means that we look for a perturbation of the original problem that has \hat{x} as its exact solution and then we measure the distance between these two problems. This idea relates error analysis with perturbation theory. The backward and forward errors are related by the conditioning of the

problem, that is, the sensitivity of the problem to perturbations of the data. In general, when backward error, forward error and the condition number of a problem are well defined, we have the useful rule of thumb:

$$\text{forward error} \lesssim \text{condition number} \times \text{backward error}.$$

The condition number is problem related. In numerical linear algebra, it is widespread the use of the normwise condition number associated to the problem of finding a solution for the linear system of equations $Ax = b$, where A is a nonsingular square matrix (see section 2.2 of [29]).

Definition 3.21. The condition number of a nonsingular matrix A is defined to be $\kappa(A) := \|A\| \cdot \|A^{-1}\|$, where $\|\cdot\|$ is a matrix norm.

Some of the most common cases are $\kappa_\infty(A)$ (using the infinity norm $\|\cdot\|_\infty$) and $\kappa_2(A)$ (using the 2-norm $\|\cdot\|_2$). This condition number depends only on A , and if it is too large it might impede getting a good forward error bound.

We have mentioned that one source of error comes from working on a finite precision arithmetic. Given a subset F of the real numbers ($F \subset \mathbb{R}$), we say that F is a *floating point number system* if its elements are of the form:

$$y = \pm m \times \beta^{e-t}, \quad (3.28)$$

where the number m is a real number called *mantissa* and it satisfies that $0 \leq m \leq \beta^{t-1}$. The number system F is defined by four integers: the base β , the precision t and the exponent range $e_{\min} \leq e \leq e_{\max}$.

Given a real number x that lies in the range of F , we can approximate it (by *rounding*) to the closest number in F with a relative error no larger than $u = \frac{1}{2}\beta^{1-t}$, which is called the *unit roundoff*. For the rounding error analysis, we will assume the following model for the floating point number system F . Given $x, y \in F$:

$$fl(x \odot y) = (x \odot y)(1 + \delta), \quad |\delta| < u, \quad \odot = +, -, *, /, \quad (3.29)$$

where $fl(\cdot)$ means the computed value of the expression. In this context, we say that we have computed a solution to high relative accuracy (HRA) if its forward relative error satisfies the following relationship

$$\text{forward relative error} \leq Ku, \quad \text{for some constant } K. \quad (3.30)$$

In double precision, the unit roundoff is of the order of 10^{-16} . We will see that the relative errors of the solutions computed with the high relative methods studied in the following chapters usually stay very close to this quantity.

While high relative accuracy is a highly desirable property for a numerical method, it is not possible to achieve it for every problem. For instance, we cannot assure HRA for the simple problem of evaluating the expression $x + y + z$ (see [30]). Another example with a negative answer comes from the evaluation of the determinant of a class of structured matrices: *Toeplitz matrices*. An $n \times n$ Toeplitz matrix B is of the form

$$B = \begin{pmatrix} a_0 & a_1 & \cdots & a_{n-2} & a_{n-1} \\ a_{-1} & a_0 & \ddots & & a_{n-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ a_{-n+2} & & \ddots & \ddots & a_1 \\ a_{-n+1} & a_{-n+2} & \cdots & a_{-1} & a_0 \end{pmatrix}. \quad (3.31)$$

Toeplitz matrices are characterized by $2n - 1$ parameters: the different elements that are repeated on every entry of each diagonal of B . But even with this simple structure, we cannot assure high relative accuracy for the computation of their determinants when the size n grows arbitrarily (see [30]).

However, in other problems the question of whether we can achieve computations to high relative accuracy is positive. In fact, there is a sufficient condition to assure the high relative accuracy of an algorithm, the condition of *no inaccurate cancellations* NIC (see [30]): the algorithm only uses multiplications, divisions, sums of real numbers of the same sign and subtractions of initial data. So, an algorithm that avoids subtractions (with the exception of subtractions of initial data) can be carried out to high relative accuracy. And an algorithm that avoids all subtractions it is called *subtraction-free* (SF) and it also satisfies the condition NIC. Hence, SF algorithms assure high relative accuracy.

In the examples that we are going to introduce, the accurate results are obtained thanks to the choice of the right representation of the problem. For example, we can expect high relative accuracy with the computation of the singular value decomposition if we know a good decomposition of the original matrix.

Given an $n \times n$ matrix A , let X , Y and D be three $n \times n$ matrices. We say that $A = XDY^T$ is a *rank revealing decomposition* (RRD) if X and Y are well-conditioned and D is a nonsingular diagonal matrix.

The interest of obtaining a rank revealing decomposition of a matrix is that it can be used to compute its singular value decomposition efficiently and with high relative accuracy following [31].

For diagonally dominant M -matrices, it is possible to compute an LDU decomposition that serves as RRD, as well as their determinants and their inverses with high relative accuracy from a special parametrization. This parametrization is given by the row sums and the off-diagonal entries of the diagonally dominant M -matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ (see [1, 32, 87]). We call these parameters DD-parameters:

$$\begin{cases} a_{ij}, & i \neq j, \\ s_i := \sum_{j=1}^n a_{ij}, & i \in N. \end{cases} \quad (3.32)$$

Using the parametrization (3.32) as input, it is possible to adapt Gaussian elimination to compute the LDU decomposition with high relative accuracy [32, 87]. Moreover, if we use an adequate pivoting strategy, we obtain a LDU decomposition with well-conditioned matrices L and U that serves as a RRD. The known pivoting strategies for this purpose are symmetric, meaning that they exchange rows and columns with the same indices at every step of Gaussian elimination. Two examples of symmetric pivoting strategies are symmetric com-

plete pivoting, which was used in [32], and maximal absolute diagonal dominance (m.a.d.d.) pivoting [87].

For nonsingular totally positive matrices, the bidiagonal decomposition (Theorem 3.14) can be used as a parametrization to achieve accurate computations. In [59, 60], Plamen Koev devised algorithms to solve many algebraic problems with nonsingular TP matrices to high relative accuracy using the bidiagonal decomposition as input. He implemented these algorithms and they are available in the library TNTool to be used in Matlab and Octave. The library can be downloaded from Koev's personal webpage [58], and it also includes subsequent contributions of more authors. Some of the functions from the library that have been key to achieving high relative accuracy in our work are the following:

- `TNEigenvalues`: Computes the eigenvalues of A to HRA from $\mathcal{BD}(A)$.
- `TNSingularValues`: Computes the singular values of A to HRA from $\mathcal{BD}(A)$.
- `TNInverseExpand`: Computes the explicit inverse A^{-1} to HRA from $\mathcal{BD}(A)$. This function was contributed by Ana Marco and José Javier Martínez [77].
- `TNSolve`: Computes the solution to the linear system of equations $Ax = b$ and assures the HRA whenever b has an alternating sign pattern. It takes as input $\mathcal{BD}(A)$ and b .
- `TNProduct`: Computes $\mathcal{BD}(AB)$, the bidiagonal decomposition of the product of two nonsingular TP matrices A and B , from $\mathcal{BD}(A)$ and $\mathcal{BD}(B)$ to HRA.
- `TNVandBD`: Computes the bidiagonal decomposition of the Vandermonde matrix on the points $t_1 < \dots < t_n$ to HRA. It requires the nodes $\{t_i\}_{1 \leq i \leq n}$ as input.

Thanks to these algorithms, we can achieve accurate computations with nonsingular TP matrices if we know their bidiagonal decomposition accurately. This approach can present a huge difference in the accuracy of the results with respect to the common methods, especially because TP matrices are often ill-conditioned in the traditional sense. For instance, the symmetric Pascal matrix gives an example of this phenomenon. The $n \times n$ Pascal matrix P_L is the matrix whose entry (i, j) is given by the combinatorial number $\binom{i+j-2}{i-1}$. The Pascal matrix is a known example of ill-conditioned matrix, and computing the eigenvalues and singular values with usual methods can lead to inaccurate results (see [2] for experiments and discussion). But this ill-conditioned matrix admits a really simple bidiagonal decomposition. Using the compact notation, we have that $\mathcal{BD}(P_L) = (1)_{1 \leq i, j \leq n}$, i.e., all the multipliers and diagonal pivots associated to the NE of P_L are 1's. And, as we would expect, working with this simple representation of P_L gives really good results [2] despite the bad conditioning of the matrix. In the next chapter, we will also observe this difference in accuracy with the matrices studied in this dissertation.

Part II
ARTICLES

Article 1

- [17] J. Delgado, H. Orera and J. M. Peña. Accurate computations with Laguerre matrices. Numer. Linear Algebra Appl. 26 (2019), e2217, 10 pp.

RESEARCH ARTICLE

Accurate computations with Laguerre matrices

Jorge Delgado¹  | Héctor Orera² | Juan Manuel Peña²

¹Departamento de Matemática Aplicada, Escuela Universitaria Politécnica de Teruel, Universidad de Zaragoza, Teruel, Spain

²Departamento de Matemática Aplicada, Facultad de Ciencias, Universidad de Zaragoza, Zaragoza, Spain

Correspondence

Jorge Delgado, Departamento de Matemática Aplicada, Escuela Universitaria Politécnica de Teruel, Universidad de Zaragoza, 44003 Teruel, Spain.

Email: jorgedel@unizar.es

Present Address

Departamento de Matemática Aplicada, Escuela Universitaria Politécnica de Teruel, Universidad de Zaragoza, 44003 Teruel, Spain.

Funding information

MINECO/FEDER, Grant/Award Number: MTM2015-65433-P; Gobierno de Aragón; Fondo Social Europeo

Summary

This paper provides an accurate method to obtain the bidiagonal factorization of collocation matrices of generalized Laguerre polynomials and of Lah matrices, which in turn can be used to compute with high relative accuracy the eigenvalues, singular values, and inverses of these matrices. Numerical examples are included.

KEYWORDS

generalized Laguerre polynomials, high relative accuracy, Laguerre matrices, Lah matrices, total positivity

1 | INTRODUCTION

Laguerre polynomials form a classical family of orthogonal polynomials (cf. the work of Beals et al.¹) and present many applications. For instance, they are used for Gaussian quadrature to numerically compute integrals. The larger family of generalized Laguerre polynomials (see Section 3) presents important applications in quantum mechanics (see the work of Koornwinder et al.²). This paper deals with the accurate computation when using collocation matrices of generalized Laguerre polynomials. The matrices considered in this paper are totally positive (TP), that is, all their minors are nonnegative. Nonsingular TP matrices have a bidiagonal factorization (see Section 2), which can be used as a parameterization to perform algebraic algorithms with high relative accuracy (HRA). In fact, if we know this bidiagonal factorization of a nonsingular TP matrix with HRA, then we can apply the algorithms presented by Koev³⁻⁵ to compute its inverse, all its eigenvalues and singular values, or the solution of some linear systems associated to the matrix with HRA. This paper performs the previous task for collocation matrices of generalized Laguerre polynomials (also called Laguerre matrices). In Section 3, we also perform this task for Lah matrices, formed by the unsigned Lah numbers. Lah matrices are closely related with some Laguerre matrices.

The layout of this paper is as follows. Section 2 presents auxiliary results and basic notations related with the bidiagonal factorization. Section 3 shows the construction with HRA of the bidiagonal factorization of Laguerre and Lah matrices.

Section 4 includes illustrative numerical examples confirming the theoretical results for the computation of eigenvalues, singular values, inverses, and the solution of linear systems.

2 | AUXILIARY RESULTS

Neville elimination is an alternative procedure to Gaussian elimination. Neville elimination produces zeros in a column of a matrix by adding to each row an appropriate multiple of the previous one (see the work of Gasca et al.⁶). Given a nonsingular matrix $A = (a_{ij})_{1 \leq i, j \leq n}$, the Neville elimination procedure has $n - 1$ steps, leading to a sequence of matrices as follows:

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

with U as an upper triangular matrix.

On the one hand, $\tilde{A}^{(t)}$ is obtained from the matrix $A^{(t)}$ by moving to the bottom the rows with a zero entry in column t below the main diagonal, if necessary. The matrix $A^{(t+1)}$ comes from $\tilde{A}^{(t)}$ by

$$a_{ij}^{(t+1)} = \begin{cases} \tilde{a}_{ij}^{(t)} - \frac{\tilde{a}_{it}^{(t)}}{\tilde{a}_{i-1,t}^{(t)}} \tilde{a}_{i-1,j}^{(t)}, & \text{if } t \leq j < i \leq n \text{ and } a_{i-1,t}^{(t)} \neq 0, \\ \tilde{a}_{ij}^{(t)}, & \text{otherwise,} \end{cases} \tag{2}$$

for all $t \in \{1, \dots, n - 1\}$.

The entry

$$p_{ij} := \tilde{a}_{ij}^{(j)}, \quad 1 \leq j \leq i \leq n \tag{3}$$

is the (i, j) pivot of the Neville elimination of A , and the pivots p_{ii} are called *diagonal pivots*. The number

$$m_{ij} = \begin{cases} \frac{\tilde{a}_{ij}^{(j)}}{\tilde{a}_{i-1,j}^{(j)}} = \frac{p_{ij}}{p_{i-1,j}}, & \text{if } \tilde{a}_{i-1,j}^{(j)} \neq 0, \\ 0, & \text{if } \tilde{a}_{i-1,j}^{(j)} = 0, \end{cases}$$

is called the (i, j) multiplier of Neville elimination of A , where $1 \leq j < i \leq n$.

Neville elimination is a very useful procedure when working with TP matrices. A matrix is TP if all its minors are nonnegative and it is *strictly TP* (STP) if they are positive (see the work of Ando⁷).

If A is an order n nonsingular TP matrix, then no rows exchanges are needed when applying Neville elimination (see Corollary 5.5 in the work of Gasca et al.⁶). Therefore, in this case, $A^{(t)} = \tilde{A}^{(t)}$ for all t .

In the work of Gasca et al.⁸, it was shown that nonsingular TP matrices satisfy a unique bidiagonal decomposition. Let us first recall the mentioned result.

Theorem 1. (cf. Theorem 4.1 in the work of Gasca et al.⁸)

Let A be a nonsingular $n \times n$ TP matrix. Then, A admits a decomposition of the form

$$A = F_{n-1} \dots F_1 D G_1 \dots G_{n-1}, \tag{4}$$

where F_i and G_i , $i \in \{1, \dots, n - 1\}$, are the lower and upper triangular nonnegative bidiagonal matrices given by

$$F_i = \begin{pmatrix} 1 & & & & & \\ 0 & 1 & & & & \\ & & \ddots & & & \\ & & & 0 & & \\ & & & & 1 & \\ & & & & m_{i+1,1} & 1 \\ & & & & & \ddots \\ & & & & & & m_{n,n-i} & 1 \end{pmatrix}, \quad G_i^T = \begin{pmatrix} 1 & & & & & \\ 0 & 1 & & & & \\ & & \ddots & & & \\ & & & 0 & & \\ & & & & 1 & \\ & & & & \tilde{m}_{i+1,1} & 1 \\ & & & & & \ddots \\ & & & & & & \tilde{m}_{n,n-i} & 1 \end{pmatrix}, \tag{5}$$

and D is a diagonal matrix $\text{diag}(p_{11}, \dots, p_{nn})$ with positive diagonal entries. If, in addition, the entries m_{ij}, \tilde{m}_{ij} satisfy

$$m_{ij} = 0 \Rightarrow m_{hj} \quad \forall h > i$$

and

$$\tilde{m}_{ij} = 0 \Rightarrow m_{ik} \quad \forall k > j,$$

then the decomposition (4) is unique.

In Theorem 4.1 in the work of Gasca et al.⁸, it was also shown that m_{ij} and p_{ii} in the bidiagonal decomposition given by (4) with (5) are the multipliers and the diagonal pivots when applying the Neville elimination to A and \tilde{m}_{ij} are the multipliers when applying the Neville elimination to A^T .

Koev⁴ introduced a compact matrix notation $BD(A)$ for the bidiagonal decomposition (4) defined by

$$(BD(A))_{ij} = \begin{cases} m_{ij}, & \text{if } i > j, \\ \tilde{m}_{ji}, & \text{if } i < j, \\ p_{ii}, & \text{if } i = j. \end{cases} \quad (6)$$

An algorithm can be performed with HRA if it does not include subtractions (except of the initial data), that is, if it only includes products, divisions, sums of numbers of the same sign, and subtractions of the initial data (cf. the work of Koev^{4,9}). In particular, a subtraction-free algorithm provides results with HRA. In the work of Koev⁴, assuming that the parameters of $BD(A)$ are known with HRA, Koev presented algorithms for computing the eigenvalues of the matrix A , the singular values of the matrix A , the inverse of the matrix A , and the solution of linear systems of equations $Ax = b$, where b has a chessboard pattern of alternating signs to HRA.

3 | ACCURATE COMPUTATIONS WITH COLLOCATION MATRICES OF GENERALIZED LAGUERRE POLYNOMIALS

Let us recall that, for $\alpha > -1$, the generalized Laguerre polynomials are given by

$$L_n^{(\alpha)}(t) = \sum_{k=0}^n (-1)^k \binom{n+\alpha}{n-k} \frac{t^k}{k!}, \quad n = 0, 1, 2, \dots, \quad (7)$$

and that they are orthogonal polynomials on $[0, \infty)$ with respect to the weight function $x^\alpha e^{-x}$.

Given a real number x and a positive integer k , let us denote the corresponding *falling factorial* by

$$x^{(k)} := x(x-1)(x-2) \cdots (x-k+1).$$

Let us also denote $x^{(0)} := 1$. Let $M := (L_{j-1}^{(\alpha)}(t_{i-1}))_{1 \leq i, j \leq n+1}$ be the collocation matrix of the generalized Laguerre polynomials at $(0 >) t_0 > t_1 > \dots > t_n$, let P_U be the $(n+1) \times (n+1)$ upper triangular Pascal matrix with $\binom{j-1}{i-1}$ as its (i, j) -entry for $j \geq i$, and let S_α and J be the $(n+1) \times (n+1)$ diagonal matrices defined as follows:

$$S_\alpha := \text{diag}((\alpha + i)^i)_{0 \leq i \leq n}, \quad J := \text{diag}((-1)^i)_{0 \leq i \leq n}. \quad (8)$$

The following result assures that, given the parameters $(0 >) t_0 > t_1 > \dots > t_n$, many algebraic computations with these collocation matrices M can be performed with HRA, as well as the strict total positivity and a particular factorization of these matrices.

Theorem 2. Let $M := (L_{j-1}^{(\alpha)}(t_{i-1}))_{1 \leq i, j \leq n+1}$ for $(0 >) t_0 > t_1 > \dots > t_n$ with $\alpha > -1$, let P_U be the $(n+1) \times (n+1)$ upper triangular Pascal matrix, let S_α and J be the $(n+1) \times (n+1)$ diagonal matrices given by (8), and let $V := (t_{i-1}^{j-1})_{1 \leq i, j \leq n+1}$. Then, $M = VJS_\alpha^{-1}P_US_0^{-1}S_\alpha$ is an STP matrix, and given the parametrization t_i ($0 \leq i \leq n$), the following computations can be performed with HRA: all the eigenvalues, all the singular values, the inverse of M , and the solution of the linear systems $Mx = b$, where $b = (b_0, \dots, b_n)^T$ has alternating signs.

Proof. Let $A = (a_{ij})_{1 \leq i, j \leq n+1}$ be the matrix of change of basis between the basis of the generalized Laguerre polynomials and the monomial basis:

$$\left(L_0^{(\alpha)}(t), L_1^{(\alpha)}(t), \dots, L_n^{(\alpha)}(t) \right) = (1, t, \dots, t^n)A. \quad (9)$$

Observe that $a_{ij} = 0$ for $j < i$ and that, for $j \geq i$,

$$\begin{aligned} a_{ij} &= \binom{j-1+\alpha}{j-i} \frac{(-1)^{i-1}}{(i-1)!} \\ &= \frac{(j-1+\alpha) \cdots (i+\alpha)(-1)^{i-1} (j-1)!}{(j-i)!(i-1)! (j-1)!} \end{aligned}$$

and so

$$\begin{aligned} a_{ij} &= (-1)^{i-1} \binom{j-1}{i-1} \frac{(j-1+\alpha) \cdots (i+\alpha)}{(j-1)!} \\ &= \binom{j-1}{i-1} \frac{(-1)^{i-1} (j-1+\alpha) \cdots (\alpha+1)}{(j-1)!(i-1+\alpha)^{i-1}}. \end{aligned}$$

Hence, we can derive $A = JS_\alpha^{-1}P_US_0^{-1}S_\alpha$. Then, we can deduce from (9) that

$$M = VJS_\alpha^{-1}P_US_0^{-1}S_\alpha. \quad (10)$$

We have that $VJ = ((-t_{i-1})^{j-1})_{1 \leq i, j \leq n+1}$, and because $0 < -t_0 < -t_1 < \dots < -t_n$, VJ is a Vandermonde matrix with strictly increasing positive nodes and so it is STP (see the work of Gantmacher et al.^{10(p111)} and Fallat et al.^{11(p12)}). It is well known (see the work of Fallat et al.^{11(p52)}) that the upper triangular Pascal matrix is (nonsingular) TP and so $S_\alpha^{-1}P_US_0^{-1}S_\alpha$ is also nonsingular TP because $S_\alpha^{-1}, S_0^{-1}, S_\alpha$ are positive diagonal matrices. Then, we can write (10) as

$$M = BC, \quad B := VJ, \quad C := S_\alpha^{-1}P_US_0^{-1}S_\alpha, \quad (11)$$

and so, by Theorem 3.1 in the work of Ando⁷, M is STP because it is a product of an STP matrix and a nonsingular TP matrix.

If we have a Vandermonde matrix with strictly increasing positive nodes, we can construct its bidiagonal factorization with HRA (see section 3 of the work of Koev³). Therefore, we can obtain with HRA the $BD(VJ)$ from the parameters $(0 <) -t_0 < -t_1 < \dots < -t_n$. For P_U ,

$$BD(P_U) = \begin{pmatrix} 1 & \cdots & \cdots & 1 \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{pmatrix},$$

(see the work of Alonso et al.¹²) that is,

$$P_U = \bar{G}_1 \cdots \bar{G}_n \quad (12)$$

with \bar{G}_k ($1 \leq k \leq n$) as the bidiagonal upper triangular matrix defined as

$$\bar{G}_k = \begin{pmatrix} 1 & \bar{g}_1^{(k)} & & \\ & \ddots & & \\ & & \ddots & \bar{g}_n^{(k)} \\ & & & 1 \end{pmatrix},$$

where $\bar{g}_i^{(k)} = 1$ if $i \geq k$ and $\bar{g}_i^{(k)} = 0$ if $i < k$. We want to obtain the bidiagonal factorization of C :

$$C = DG_1 \cdots G_n, \quad (13)$$

where D is a diagonal matrix and each G_k ($1 \leq k \leq n$) is a bidiagonal upper triangular matrix with unit diagonal. Let us denote by $g_i^{(k)}$ the $(i, i+1)$ entry of G_k for each $i = 1 \dots, n$. By (11), $C = S_\alpha^{-1}P_U\bar{D}$, where $\bar{D} = \text{diag}(d_0, d_1, \dots, d_n) := S_0^{-1}S_\alpha$. Then, observe that, for each $1 \leq k \leq n$, $\bar{G}_k\bar{D} = \bar{D}G_k$, and so, for $i < k$, $g_i^{(k)} = 0$ and for $i \geq k$,

$$g_i^{(k)} = \frac{d_{i+1}}{d_i} = \frac{(i+\alpha)^i(i-1)!}{(i-1+\alpha)^{i-1}i!} = \frac{i+\alpha}{i}.$$

Then, taking into account (12) and that $S_\alpha^{-1}\bar{D} = S_\alpha^{-1}S_0^{-1}S_\alpha = S_0^{-1}$, we can obtain the bidiagonal factorization of C :

$$C = S_0^{-1}G_1 \cdots G_n$$

(observe using (13) that, by the uniqueness of the bidiagonal factorization, $D = S_0^{-1}$).

Then, following Section 5.2 of the work of Koev⁴, we can construct from (11) $BD(M)$ with HRA, through the subtraction-free Algorithm 5.1 in the work of Koev⁴, because we know with HRA the bidiagonal factorization of both factors B and C of M .

Finally, the construction of $BD(M)$ with HRA guarantees that the algebraic computations mentioned in the statement of this theorem can be performed with HRA (see Section 2 of this paper or section 3 of the work of Koev⁴). \square

Applying the previous result to the case $\alpha = 0$ provides the result for classical Laguerre polynomials. If we extend (7) to the case $\alpha = -1$, we can derive (in Theorem 3) an analogous result to Theorem 2 for the particular set of polynomials:

$$L_0^{(-1)}(t) = 1, \quad L_n^{(-1)}(t) = \sum_{k=1}^n (-1)^k \binom{n-1}{n-k} \frac{t^k}{k!}, \quad n = 1, 2, \dots \tag{14}$$

The interest of these polynomials arises from the close relationship between their coefficients and the unsigned Lah numbers (cf. the work of Boyadzhiev et al.¹³), which will be described as follows:

$$L_n^{(-1)}(t) = \frac{1}{n!} \sum_{k=1}^n (-1)^k L(n, k) t^k \text{ for } n \geq 1 \text{ with } L(n, k) := \binom{n-1}{k-1} \frac{n!}{k!}, k \leq n. \tag{15}$$

The unsigned Lah numbers $L(n, k)$ are included as the sequence A105278 in the On-line Encyclopedia of Integer Sequences (OEIS). The Lah numbers were introduced by Ivo Lah in 1955 (cf. the work of Lah¹⁴) and arise in applications such as combinatorics and analysis (see the work of Riordan^{15(pp44-45)}).

Before introducing Theorem 3, it is convenient to define the matrix P_U^* because it will play the role that P_U played in Theorem 2. The $(n + 1) \times (n + 1)$ matrix P_U^* is obtained from an $n \times n$ upper triangular Pascal matrix P_U by adding $(1, 0, \dots, 0)$ as a first row and column.

$$P_U^* = \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & P_U \end{array} \right). \tag{16}$$

Observe that P_U^* is a nonsingular TP matrix because P_U is nonsingular TP.

Theorem 3. Let $M = (L_{j-1}^{(-1)}(t_{i-1}))_{1 \leq i, j \leq n+1}$ for $(0 >) t_0 > t_1 > \dots > t_n$, let P_U^* be the $(n + 1) \times (n + 1)$ upper triangular matrix given by (16), let S_0 and J be the $(n + 1) \times (n + 1)$ diagonal matrices given by (8), and let $V := (t_{i-1}^{j-1})_{1 \leq i, j \leq n+1}$. Then, $M = VJS_0^{-1}P_U^*$; it is an STP matrix, and given the parametrization t_i ($0 \leq i \leq n$), the following computations can be performed with HRA: all the eigenvalues, all the singular values, the inverse of M , and the solution of the linear systems $Mx = b$, where $b = (b_0, \dots, b_n)^T$ has alternating signs.

Proof. Let $A = (a_{ij})_{1 \leq i, j \leq n+1}$ be the matrix of change of basis given by (9) when $\alpha = -1$. Observe that $a_{11} = 1$, $a_{1j} = 0$ for $j = 2, \dots, n + 1$, $a_{ij} = 0$ for $j < i$ and that, for $j \geq i \geq 2$, $a_{ij} = \frac{(-1)^{j-1}}{(i-1)!} \binom{j-2}{i-2}$. Then,

$$A = JS_0^{-1}P_U^* \quad \text{and} \quad M = VJS_0^{-1}P_U^*. \tag{17}$$

Rearranging the factors of M , we obtain the factorization as follows:

$$M := BC, \text{ where } B := VJ \text{ and } C := S_0^{-1}P_U^*.$$

Following the reasoning given in the proof of Theorem 2, we can deduce that M is STP because the Vandermonde matrix $B = VJ = ((-t_{i-1})^{j-1})_{1 \leq i, j \leq n+1}$ is STP and the matrix C is a nonsingular TP matrix because it is the product of a positive diagonal matrix and a nonsingular TP matrix. Again, by algorithm 5.1 in the work Koev⁴, we can obtain $BD(M)$ if we know both $BD(B)$ and $BD(C)$ to HRA. The bidiagonal factorization of a TP Vandermonde matrix can be obtained with HRA (section 3 of the work of Koev³), and so we only need to find $BD(C)$ with HRA. From (3), it is straightforward to deduce that

$$BD(P_U^*) = \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & \dots & \dots & 1 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}, \tag{18}$$

and so, we have that

$$BD(C) = BD(S_0^{-1}P_U^*) = \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ 0 & 1! & 1 & \dots & 1 \\ \vdots & 0 & 2! & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 0 & 0 & \dots & 0 & n! \end{pmatrix}.$$

Then, we can construct $BD(M)$ with HRA and perform all the previously mentioned algebraic computations with HRA. □

The Lah matrix Λ is the matrix formed by the unsigned Lah numbers (cf. the work of Martinjak et al.¹⁶). This matrix can be written as $\Lambda = JAS_0$, where A is the matrix of change of basis between the basis of the generalized Laguerre polynomials with $\alpha = -1$ and the monomial basis given by (15). The following result also shows that many computations with Lah matrices can be performed with HRA.

Proposition 1. *Let Λ be the Lah matrix, let P_U^* be the $(n + 1) \times (n + 1)$ upper triangular matrix given by (16), and let S_0 be the $(n + 1) \times (n + 1)$ diagonal matrix given by (8). Then, $\Lambda = S_0^{-1}P_U^*S_0$ is a TP matrix,*

$$BD(\Lambda) = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & 1 & 2 & 3 & 4 & \dots & n \\ \vdots & 0 & 1 & 3 & 4 & \dots & n \\ \vdots & \vdots & \ddots & 1 & 4 & \dots & n \\ \vdots & \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & & \ddots & 1 & n \\ 0 & 0 & \dots & \dots & \dots & 0 & 1 \end{pmatrix}$$

and the following computations can be performed with HRA: all the eigenvalues, all the singular values, the inverse of Λ , and the solution of the linear systems $\Lambda x = b$, where $b = (b_0, \dots, b_n)^T$ has alternating signs.

Proof. By (15), $\Lambda = S_0^{-1}P_U^*S_0$. Because P_U^* is nonsingular TP, we conclude that Λ is also nonsingular TP. From (18), we have that $P_U^* = \bar{G}_1 \cdots \bar{G}_{n-1}$, where \bar{G}_k ($1 \leq k \leq n - 1$) is the bidiagonal upper triangular matrix

$$\bar{G}_k = \begin{pmatrix} 1 & \bar{g}_1^{(k)} & & & \\ & \ddots & & & \\ & & \ddots & & \\ & & & \bar{g}_n^{(k)} & \\ & & & & 1 \end{pmatrix},$$

with $\bar{g}_i^{(k)} = 1$ if $i \geq k + 1$ and $\bar{g}_i^{(k)} = 0$ if $i < k + 1$. We want to obtain the bidiagonal factorization of Λ :

$$\Lambda = DG_1 \cdots G_{n-1}, \tag{19}$$

where D is a diagonal matrix and each G_k ($1 \leq k \leq n - 1$) is a bidiagonal upper triangular matrix with unit diagonal. Let us denote by $g_i^{(k)}$ the $(i, i + 1)$ entry of G_k for each $i = 1, \dots, n - 1$. Then, observe that, for each $1 \leq k \leq n - 1$, $\bar{G}_k S_0 = S_0 G_k$, and so, for $i < k + 1$, $g_i^{(k)} = 0$ and, for $i \geq k + 1$,

$$g_i^{(k)} = \frac{d_{i+1}}{d_i} = \frac{i!}{(i-1)!} = i.$$

By the uniqueness of the bidiagonal factorization, we derive $D = S_0^{-1}S_0 = I_{(n+1) \times (n+1)}$ and $\Lambda = G_1 \cdots G_{n-1}$. Because we obtained $BD(\Lambda)$ with HRA, we can perform all the algebraic computations included in the statement of this proposition with HRA. □

Let us recall that, in the work of Martinjak et al.¹⁶, it was already proved that the submatrix obtained from Λ by removing its first row and column is TP.

In the next section, we shall illustrate the accurate computations in the case of the classical Laguerre polynomials (of (7) with $\alpha = 0$). The corresponding collocation matrices will be called Laguerre matrices.

4 | NUMERICAL TESTS

Assuming that the parameterization $BD(A)$ of a square TP matrix A is known with HRA, Koev⁴ devised algorithms to compute the inverse, the eigenvalues, and the singular values of A and the solution of linear systems of equations $Ax = b$, where b has a chessboard pattern of alternating signs. Koev implemented these algorithms in order to be used with MATLAB and Octave in the software library *TNTool* available in the work of Koev⁵. The corresponding functions are `TNInverseExpand`, `TNEigenvalues`, `TNSingularValues`, and `TNSolve`, respectively. These three functions require as input argument the data determining the bidiagonal decomposition (4) of A , $BD(A)$ given by (6), to HRA. `TNSolve` also requires a second argument, the vector b of the linear system $Ax = b$ to be solved.

In the library *TNTool*, Koev also provided the function `TNProduct(B1, B2)`, which, given the bidiagonal decompositions $B1$ and $B2$ to HRA of two TP matrices F and G , provided the bidiagonal decomposition of the TP matrix FG to HRA. We can observe in the factorization $M = VJS_0^{-1}P_U$ of Theorem 2 for $\alpha = 0$ that M can be expressed as the product of three TP matrices: the Pascal matrix P_U , S_0^{-1} , and the TP Vandermonde matrix VJ . In the work of Alonso et al.¹², the bidiagonal factorization to HRA of Pascal matrix P_U was shown. S_0^{-1} is a diagonal TP matrix so its bidiagonal decomposition is itself and $BD(S_0^{-1}) = S_0^{-1}$. Taking into account the form of its diagonal entries, it can be obtained to HRA. Finally, VJ is a TP Vandermonde matrix with node sequence $-\mathbf{t} = (-t_i)_{i=0}^n$, and by using `TNVandBD(-t)` of library *TNTool*, $BD(VJ)$ to HRA can be obtained. Taking into account these facts, the pseudocode providing $BD(M)$ to HRA can be seen in Algorithm 1.

Algorithm 1 Computation of the bidiagonal decomposition of M to HRA

Require: $\mathbf{t} = (t_i)_{i=0}^n$ such that $0 > t_0 > t_1 > \dots > t_n$

Ensure: B bidiagonal decomposition of M to HRA

$B1 = TNV$ and $BD(-\mathbf{t})$

$B2 = \text{diag}(1/0!, 1/1!, \dots, 1/n!)$

for $i = 0 : n$ **do**

for $j = 0 : i - 1$ **do**

$B(i, j) = 0$

end for

for $j = i : n$ **do**

$B(i, j) = 1$

end for

end for

$B = TN \text{ Product}(B1, B2)$

$B = TN \text{ Product}(B, B3)$

We have implemented the previous algorithm to be used in MATLAB and Octave in a function `TNBDLaguerre`.

The bidiagonal decompositions with HRA of Laguerre matrices obtained with `TNBDLaguerre` can be used with `TNInverseExpand`, `TNEigenValues`, `TNSingularValues`, and `TNSolve` in order to obtain accurate solutions for the above mentioned algebraic problems. Now, we include some numerical experiments illustrating high accuracy.

Let us consider the Laguerre matrices M_n of order $n + 1$ given by the collocation matrices of the classical Laguerre polynomials $(L_0^{(0)}(x), \dots, L_n^{(0)}(x))$ at the nodes $(-i - 1)_{0 \leq i \leq n}$, that is,

$$M_n = \left(L_{j-1}^{(0)}(t_{i-1}) \right)_{1 \leq i, j \leq n+1}, \quad (20)$$

for $n = 1, 2, \dots, 49$.

First, we have computed in MATLAB by using `TNBDLaguerre` the bidiagonal decomposition of the matrices M_n to HRA. Then, we have used that bidiagonal decomposition of M_n for computing their eigenvalues and their singular values with `TNEigenValues` and `TNSingularValues`, respectively. In the case of eigenvalues, we also compute their

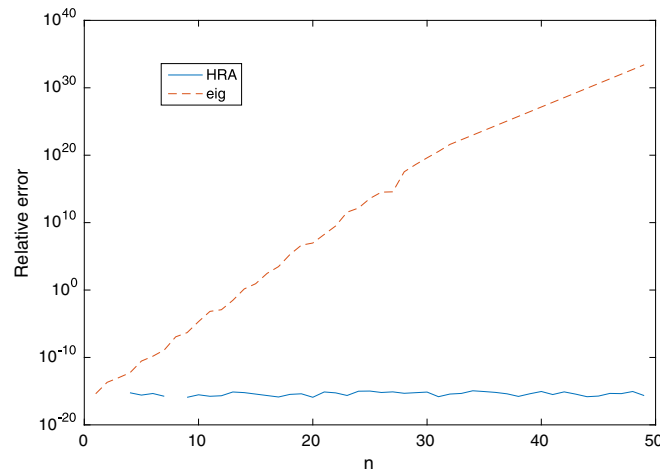


FIGURE 1 Relative errors for the lowest eigenvalue of Laguerre matrices. HRA = high relative accuracy

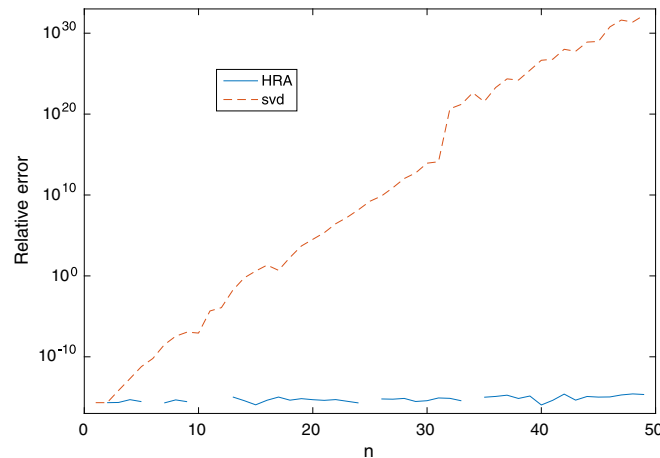


FIGURE 2 Relative errors for the lowest singular value of Laguerre matrices. HRA = high relative accuracy

approximations with the MATLAB function `eig`. We have also computed the eigenvalues of M_n by using Mathematica with a 100 digits precision. Then, we compute the relative errors corresponding to the approximations of the eigenvalues obtained with both methods `eig` and `TNEigenValues` with `TNBDLaguerre`, considering the eigenvalues provided by Mathematica as exact. We have observed that the approximations of all the eigenvalues obtained with `TNBDLaguerre` are very accurate, whereas the approximations of the lower eigenvalues obtained with command `eig` are not very accurate. In particular, the lower the eigenvalue is, the more inaccurate the approximation obtained with `eig` is. In order to illustrate this fact, Figure 1 shows the relative errors of the approximations to the lowest eigenvalue of the matrices M_1, \dots, M_{49} obtained by both `eig` and `TNEigenValues` with `TNBDLaguerre`. We can observe in the figure that our method provides very accurate results in contrast to the poor results provided by `eig`.

For the case of singular values, we have also computed their approximations with the MATLAB function `svd`. In order to show the accuracy of the approximations to the singular values computed in both ways, we calculate the singular values of the matrices M_n with Mathematica using a precision of 100 digits. As in the case of eigenvalues, we observed that the lower the singular value is, the more unaccurate the approximation obtained with `svd` is, whereas the approximations of all the singular values provided by the new method are very accurate. Figure 2 shows the relative errors of the approximations to the lowest singular value of the matrices M_1, \dots, M_{49} obtained by both `svd` and `TNSingularValues` with `TNBDLaguerre`. We can observe in the figure that the HRA algorithm outperforms `svd`.

We have also computed with MATLAB approximations to M_i^{-1} , $i = 1, \dots, 49$, with `inv` and `TNInverseExpand` using the bidiagonal decomposition given by `TNBDLaguerre`. With Mathematica, we have computed the inverse of these

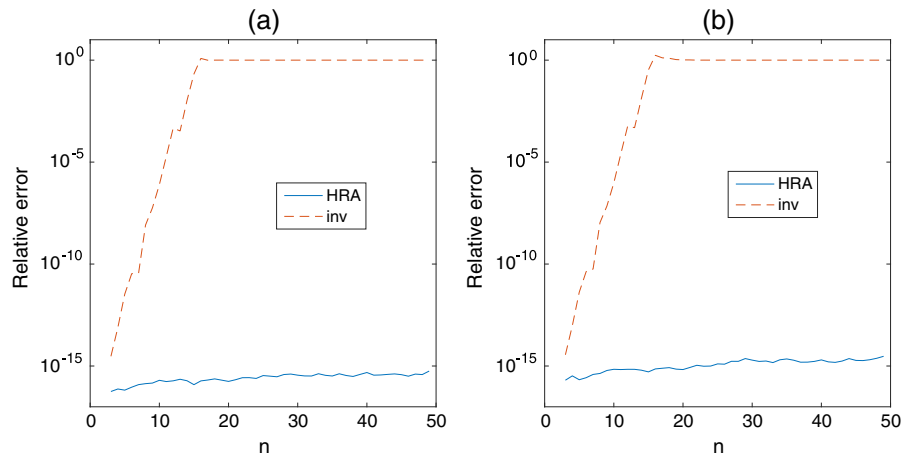


FIGURE 3 Relative errors for M_i^{-1} , $i = 1, \dots, 49$. (a) Mean relative error. (b) Maximum relative error. HRA = high relative accuracy

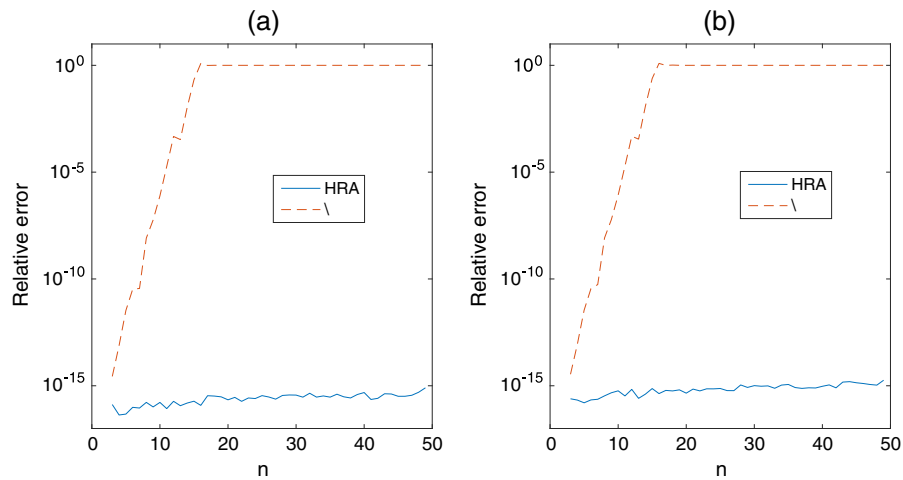


FIGURE 4 Relative errors for the systems $M_i x = b_i$, $i = 1, \dots, 49$. (a) Mean relative error. (b) Maximum relative error. HRA = high relative accuracy

Laguerre matrices with exact arithmetic. Then, we have computed the corresponding componentwise relative errors. Finally, we have obtained the mean and maximum componentwise relative error. Figure 3a shows the mean relative error and Figure 3b shows the maximum relative error. We can also observe in this case that the results obtained with `TNInverseExpand` are much more accurate than the ones obtained with `inv`.

Now, we consider the linear systems $M_i x = b_i$, $i = 1, \dots, 49$, where M_i is the Laguerre matrix of order $i + 1$ previously defined and $b_i \in \mathbb{R}^{i+1}$ has the absolute value of its entries randomly generated as integers in the interval $[1, 1000]$, but with alternating signs. We have computed approximations to the solution x of the linear system with MATLAB, the first one using `TNSolve` and the bidiagonal decomposition of the Laguerre matrices A obtained with `TNBDLaguerre`, and the second one using the MATLAB command `A\b`. By using Mathematica with exact arithmetic, we have computed the exact solution of the systems, and then, we have computed the componentwise relative errors for the two approximations obtained with MATLAB. Then, we have obtained the mean and maximum componentwise relative error. Figure 4a shows the mean relative error and Figure 4b shows the maximum relative error. Again, the results obtained with HRA algorithms are very accurate in contrast to the results obtained with the usual MATLAB command.

Finally, we consider the linear systems $M_i x = \tilde{b}_i$, $i = 1, \dots, 49$, where now $\tilde{b}_i \in \mathbb{R}^{i+1}$ has its entries randomly generated as integers in the interval $[-1000, 1000]$ and so it has not a chessboard pattern of alternating signs. Hence, HRA is lost when `TNSolve` is used. In spite of this, Figure 5a,b shows that, even in this case, our algorithm outperforms the usual MATLAB command `A\b`.

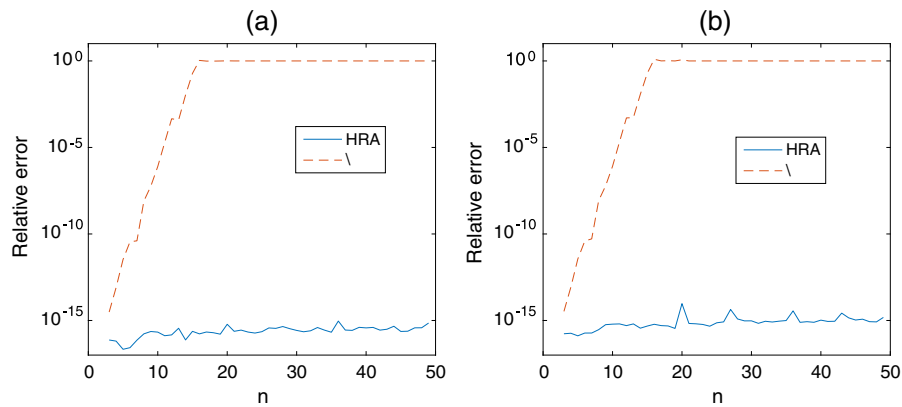


FIGURE 5 Relative errors for the systems $M_i x = \tilde{b}_i$, $i = 1, \dots, 49$. (a) Mean relative error. (b) Maximum relative error. HRA = high relative accuracy

ACKNOWLEDGEMENTS

This work was partially supported through the Spanish research grant MTM2015-65433-P (MINECO/FEDER), by Gobierno de Aragón, and by Fondo Social Europeo.

ORCID

Jorge Delgado  <http://orcid.org/0000-0003-2156-9856>

REFERENCES

1. Beals R, Wong R. Special functions and orthogonal polynomials. Cambridge, UK: Cambridge University Press; 2016.
2. Koornwinder TH, Wong RSC, Koekoek R, Swarttouw RF. Orthogonal polynomials. In: Olver FWJ, Lozier DM, Boisvert RF, Clark CW, editors. NIST handbook of mathematical functions. Cambridge, UK: Cambridge University Press; 2010.
3. Koev P. Accurate eigenvalues and SVDs of totally nonnegative matrices. *SIAM J Matrix Anal Appl.* 2005;27:1–23.
4. Koev P. Accurate computations with totally nonnegative matrices. *SIAM J Matrix Anal Appl.* 2007;29:731–751.
5. Koev P. <http://www.math.sjsu.edu/~koev/software/TNTool.html>. [Accessed 28 September 2018].
6. Gasca M, Peña JM. Total positivity and Neville elimination. *Linear Algebra Appl.* 1992;165:25–44.
7. Ando T. Totally positive matrices. *Linear Algebra Appl.* 1987;90:165–219.
8. Gasca M, Peña JM. On factorizations of totally positive matrices. In: Gasca M, Micchelli CA, editors. Total positivity and its applications. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1996; p. 109–130.
9. Demmel J, Koev P. The accurate and efficient solution of a totally positive generalized Vandermonde linear system. *SIAM J Matrix Anal Appl.* 2005;27:142–152.
10. Gantmacher FR, Krein MG. Oszillationsmatrizen, oszillationskerne und kleine schwingungen mechanischer systeme. Berlin, Germany: Akademie-Verlag; 1960.
11. Fallat SM, Johnson CR. Totally nonnegative matrices. Princeton, NJ: Princeton University Press; 2011. Princeton Series in Applied Mathematics, No. 35.
12. Alonso P, Delgado J, Gallego R, Peña JM. Conditioning and accurate computations with Pascal matrices. *J Comput Appl Math.* 2013;252:21–26.
13. Boyadzhiev K. Lah numbers, Laguerre polynomials of order negative one, and the nth derivative of $\exp(1/x)$. *Acta Univ Sapientiae Matem.* 2016;8:22–31.
14. Lah I. Eine neue art von zahlen, ihre eigenschaften und anwendung in der mathematischen statistik. *Mitteilungsbl Math Statist.* 1955;7:203–212.
15. Riordan J. Introduction to combinatorial analysis. Mineola, NY; Dover Publications, Inc.; 2002.
16. Martinjak I, Škrekovski R. Lah numbers and Lindström's lemma. *Comptes Rendus Math.* 2018;356:5–7.

How to cite this article: Delgado J, Orera H, Peña JM. Accurate computations with Laguerre matrices. *Numer Linear Algebra Appl.* 2019;26:e2217. <https://doi.org/10.1002/nla.2217>

Article 2

- [16] J. Delgado, H. Orera and J. M. Peña. Accurate algorithms for Bessel matrices. *J. Sci. Comput.* 80 (2019), 1264-1278.



Accurate Algorithms for Bessel Matrices

Jorge Delgado¹ · Héctor Orera² · J. M. Peña²

Received: 23 November 2018 / Revised: 2 May 2019 / Accepted: 8 May 2019 / Published online: 17 May 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

In this paper, we prove that any collocation matrix of Bessel polynomials at positive points is strictly totally positive, that is, all its minors are positive. Moreover, an accurate method to construct the bidiagonal factorization of these matrices is obtained and used to compute with high relative accuracy the eigenvalues, singular values and inverses. Similar results for the collocation matrices for the reverse Bessel polynomials are also obtained. Numerical examples illustrating the theoretical results are included.

Keywords Bessel matrices · Totally positive matrices · High relative accuracy · Bessel polynomials · Reverse Bessel polynomials

Mathematics Subject Classification 65F05 · 65F15 · 65G50 · 33C10 · 33C45 · 15A23

1 Introduction

Bessel polynomials are ubiquitous and occur in many fields such as partial differential equations, number theory, algebra and statistics (see [11]). They form an orthogonal sequence of polynomials and are related to the modified Bessel function of the second kind (see pp. 7 and 34 of [11]). They are closely related to the reverse Bessel polynomials, which have many applications in Electrical Engineering. In particular, they play a key role in network analysis of electrical circuits (see page 145 of [11] and references therein). In Combinatorics, the coefficients of the reverse Bessel polynomials are also known as signless Bessel numbers of the first kind. The Bessel numbers have been studied from a combinatorial perspective and are closely related to the Stirling numbers [12,22]. In [16] it was shown that Bessel polynomials occur naturally in the theory of traveling spherical waves. Bessel polynomials are also very important for some problems of static potentials, signal processing and electronics. For example, the Bessel polynomials are used in Frequency Modulation (FM) synthesis and in the Bessel filter. In the case of FM synthesis, the polynomials are used to compute the band-

✉ Jorge Delgado
jorgedel@unizar.es

¹ Departamento de Matemática Aplicada, Escuela Universitaria Politécnica de Teruel, Universidad de Zaragoza, 44003 Teruel, Spain

² Departamento de Matemática Aplicada, Facultad de Ciencias, Universidad de Zaragoza, 50009 Zaragoza, Spain

width of a modulated in frequency signal. The zeros of Bessel polynomials and generalized Bessel polynomials also play a crucial role in applications in Electrical Engineering. On the accurate computations of the zeros of generalized Bessel polynomials see [20].

This paper deals with the accurate computation when using collocation matrices of Bessel polynomials and reverse Bessel polynomials. It is shown that these matrices provide new structured classes for which algebraic computations (such as the computation of the inverse, of all the eigenvalues and singular values, or the solutions of some linear systems) can be performed with high relative accuracy (HRA). Moreover, a crucial result for this purpose has been the total positivity of the considered matrices. Let us recall that a matrix is *totally positive* (*strictly totally positive*) if all its minors are nonnegative (positive) and will be denoted TP (STP). These matrices have also been called in the literature totally nonnegative (totally positive). Many applications of these matrices can be seen in [2,7,21]. For some subclasses of TP matrices a bidiagonal factorization with HRA has been obtained (cf. [1,4–6,17–19]). In [3] it was proved that the first positive zero of a Bessel function of the first kind is the half of the critical length of a cycloidal space, relating Bessel functions with the total positivity theory and computer-aided geometric design. We prove here a new surprising connection of total positivity with Bessel functions through the collocation matrices of Bessel polynomials.

The paper is organized as follows. In Sect. 2, we present some auxiliary results and basic notations related to the bidiagonal factorization of totally positive matrices as well as with high relative accuracy. In Sect. 3 we introduce the Bessel polynomials and we first prove that the matrix of change of basis between the Bessel polynomials and the monomials is TP. We also define the Bessel matrices and prove that they are STP. Finally, we provide the construction with high relative accuracy of the bidiagonal factorization of Bessel matrices, which in turn can be used to apply the algorithms presented by Koev in [15] for the algebraic computations mentioned above with high relative accuracy. A similar task for the collocation matrices of the reverse Bessel polynomials is performed in Sect. 4. Finally, Sect. 5 includes illustrative numerical examples confirming the theoretical results for the computation of eigenvalues, singular values, inverses, and the solution of linear systems with the matrices considered in this paper.

2 Auxiliary Results

Neville elimination (NE) is an alternative procedure to Gaussian elimination. NE makes zeros in a column of a matrix by adding to each row an appropriate multiple of the previous one (see [9]). Given a nonsingular matrix $A = (a_{ij})_{1 \leq i, j \leq n}$, the NE procedure consists of $n - 1$ steps, leading to a sequence of matrices as follows:

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

with U an upper triangular matrix.

On the one hand, $\tilde{A}^{(t)}$ is obtained from the matrix $A^{(t)}$ by moving to the bottom the rows with a zero entry in column t below the main diagonal, if necessary. The matrix $A^{(t+1)}$, $t = 1, \dots, n - 1$, is obtained from $\tilde{A}^{(t)}$ by computing

$$a_{ij}^{(t+1)} = \begin{cases} \tilde{a}_{ij}^{(t)} - \frac{\tilde{a}_{it}^{(t)}}{\tilde{a}_{i-1,t}^{(t)}} \tilde{a}_{i-1,j}^{(t)}, & \text{if } t \leq j < i \leq n \text{ and } \tilde{a}_{i-1,t}^{(t)} \neq 0, \\ \tilde{a}_{ij}^{(t)}, & \text{otherwise,} \end{cases} \quad (2)$$

for all $t \in \{1, \dots, n - 1\}$.

The entry

$$p_{ij} := \tilde{a}_{ij}^{(j)}, \quad 1 \leq j \leq i \leq n, \quad (3)$$

is the (i, j) pivot of the NE of A , and the pivots p_{ii} are called *diagonal pivots*. The number

$$m_{ij} = \begin{cases} \frac{\tilde{a}_{ij}^{(j)}}{\tilde{a}_{i-1,j}^{(j)}} = \frac{p_{ij}}{p_{i-1,j}}, & \text{if } \tilde{a}_{i-1,j}^{(j)} \neq 0, \\ 0, & \text{if } \tilde{a}_{i-1,j}^{(j)} = 0, \end{cases} \quad (4)$$

is called the (i, j) multiplier of NE of A , where $1 \leq j < i \leq n$.

NE is a very useful method when applied to TP matrices. If A is a nonsingular TP matrix, then no rows exchanges are needed when applying NE and so, in this case, $A^{(t)} = \tilde{A}^{(t)}$ for all t . In fact, in Theorem 5.4 of [9] the following characterization of nonsingular TP matrices was provided.

Theorem 1 *Let A be a nonsingular matrix. Then A is TP if and only if there are no row exchanges in the NE of A and U^T and the pivots of both NE are nonnegative.*

In [10] it was seen that nonsingular TP matrices can be expressed as a unique bidiagonal decomposition.

Theorem 2 (cf. Theorem 4.2 of [10]). *Let A be a nonsingular $n \times n$ TP matrix. Then A admits a decomposition of the form*

$$A = F_{n-1} \cdots F_1 D G_1 \cdots G_{n-1}, \quad (5)$$

where F_i and G_i , $i \in \{1, \dots, n-1\}$, are the lower and upper triangular nonnegative bidiagonal matrices given by

$$F_i = \begin{pmatrix} 1 & & & & & & & & & & \\ & 1 & & & & & & & & & \\ & & \ddots & & & & & & & & \\ & & & \ddots & & & & & & & \\ & & & & 0 & & & & & & \\ & & & & & 1 & & & & & \\ & & & & & & m_{i+1,1} & & & & \\ & & & & & & & 1 & & & \\ & & & & & & & & \ddots & & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & & m_{n,n-i} & \\ & & & & & & & & & & & 1 \end{pmatrix}, \quad G_i^T = \begin{pmatrix} 1 & & & & & & & & & & \\ & 1 & & & & & & & & & \\ & & \ddots & & & & & & & & \\ & & & \ddots & & & & & & & \\ & & & & 0 & & & & & & \\ & & & & & \tilde{m}_{i+1,1} & & & & & \\ & & & & & & 1 & & & & \\ & & & & & & & 1 & & & \\ & & & & & & & & \ddots & & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & & \tilde{m}_{n,n-i} & \\ & & & & & & & & & & & 1 \end{pmatrix}, \quad (6)$$

and D a diagonal matrix $\text{diag}(p_{11}, \dots, p_{nn})$ with positive diagonal entries. If, in addition, the entries m_{ij} , \tilde{m}_{ij} satisfy

$$m_{ij} = 0 \Rightarrow m_{hj} = 0 \quad \forall h > i$$

and

$$\tilde{m}_{ij} = 0 \Rightarrow \tilde{m}_{kj} = 0 \quad \forall k > i$$

then the decomposition (5) is unique.

By Theorems 4.1 and 4.2 of [10] we also know that, for $1 \leq j < i \leq n$, m_{ij} and p_{ii} in the bidiagonal decomposition given by (5) with (6) are the multipliers and the diagonal pivots when applying the NE to A and, using the arguments of p. 116 of [10], \tilde{m}_{ij} are the multipliers when applying the NE to A^T .

In [14] it was devised a concise matrix notation $\mathcal{BD}(A)$ for the bidiagonal decomposition (5) and (6) given by

$$(\mathcal{BD}(A))_{ij} = \begin{cases} m_{ij}, & \text{if } i > j, \\ \tilde{m}_{ji}, & \text{if } i < j, \\ p_{ii}, & \text{if } i = j. \end{cases} \quad (7)$$

Remark 1 If A is a TP matrix, then A^T is also TP. Transposing formula (5) of Theorem 2 we obtain the unique bidiagonal decomposition of A^T :

$$A^T = G_{n-1}^T \cdots G_1^T D F_1^T \cdots F_{n-1}^T,$$

where F_i and $G_i, i \in \{1, \dots, n-1\}$, are the lower and upper triangular nonnegative bidiagonal matrices given in formula (6). It can also be checked that

$$\mathcal{BD}(A^T) = \mathcal{BD}(A)^T.$$

An algorithm can be performed with high relative accuracy if all the included subtractions are of initial data, that is, if it only includes products, divisions, sums of numbers of the same sign and subtractions of the initial data (cf. [6,14]). In [14], given a nonsingular TP matrix A whose $\mathcal{BD}(A)$ is known with HRA, Koev presented algorithms for computing to HRA the eigenvalues of the matrix A , the singular values of the matrix A , the inverse of the matrix A and the solution of linear systems of equations $Ax = b$, where b has a pattern of alternating signs.

3 Bessel Polynomials and Matrices

Let us consider the Bessel polynomials defined by

$$B_n(x) = \sum_{k=0}^n \frac{(n+k)!}{2^k(n-k)!k!} x^k, \quad n = 0, 1, 2, \dots, \tag{8}$$

Given a real positive integer n , let us define the corresponding *semifactorial* by

$$n!! = \prod_{k=0}^{\lfloor n/2 \rfloor - 1} (n - 2k).$$

Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be the lower triangular matrix such that

$$(B_0(x), B_1(x), \dots, B_{n-1}(x))^T = A(1, x, \dots, x^{n-1})^T, \tag{9}$$

that is, the lower triangular matrix A is defined by

$$a_{ij} := \begin{cases} \frac{(i+j-2)!}{2^{j-1}(i-j)!(j-1)!} = \frac{(2j-2)!}{2^{j-1}(j-1)!} \binom{i+j-2}{i-j}, & \text{if } i \geq j, \\ 0, & \text{if } i < j. \end{cases} \tag{10}$$

Theorem 3 proves the total positivity of A , and provides $\mathcal{BD}(A)$. In addition, its proof gives the explicit form of all the entries of the matrices $A^{(k)}$ of (1) computed through the NE of A .

Theorem 3 Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be the lower triangular matrix in (9) defined by (10). Then we have that

(i) the pivots of the NE of A are given by

$$p_{ij} = \frac{1}{2^{j-1}} \frac{(i-1)!}{(i-j)!} \prod_{r=1}^{j-1} \frac{(2i-r-1)}{(i-j+r)}, \quad 1 \leq j \leq i \leq n, \tag{11}$$

and the multipliers by

$$m_{ij} = \frac{(2i-2)(2i-3)}{(2i-j-1)(2i-j-2)}, \quad 1 \leq j < i \leq n, \tag{12}$$

- (ii) A is a nonsingular TP matrix
 (iii) and the bidiagonal factorization of A is given by

$$BD(A)_{ij} = \begin{cases} \frac{(2i-2)(2i-3)}{(2i-j-1)(2i-j-2)}, & \text{if } i > j, \\ 1, & \text{if } i = j = 1, \\ (2i-3)!!, & \text{if } i = j > 1, \\ 0, & \text{if } i < j, \end{cases} \quad (13)$$

and can be computed to HRA.

Proof (i) If $A^{(k)} = (a_{ij}^{(k)})_{1 \leq i, j \leq n}$ is the matrix obtained after $k-1$ steps of the NE of A (see (1)) for $k = 2, \dots, n$, let us prove by induction on $k \in \{2, \dots, n\}$ that

$$a_{ij}^{(k)} = \frac{1}{2^{j-1}} \frac{(i+j-k-1)!}{(i-j)!(j-k)!} \prod_{r=1}^{k-1} \frac{(2i-r-1)}{(i-k+r)} \quad (14)$$

for $i \geq j$. For the case $k = 2$, let us start by computing the first step of the NE of A :

$$\begin{aligned} a_{ij}^{(2)} &= a_{ij} - \frac{a_{i1}}{a_{i-1,1}} a_{i-1,j} = a_{ij} - a_{i-1,j} \\ &= \frac{1}{2^{j-1}} \left(\frac{(i+j-2)!}{(i-j)!(j-1)!} - \frac{(i+j-3)!}{(i-j-1)!(j-1)!} \right) \\ &= \frac{(i+j-3)!}{2^{j-1}(i-j-1)!(j-1)!} \left(\frac{i+j-2}{i-j} - 1 \right) \\ &= \frac{(i+j-3)!(2j-2)}{2^{j-1}(i-j)!(j-1)!} \\ &= \frac{(i+j-3)!}{2^{j-1}(i-j)!(j-2)!} \frac{(2i-2)}{(i-1)}. \end{aligned}$$

Hence, formula (14) holds for $k = 2$. Now, let us assume that (14) holds for some $k \in \{2, \dots, n-1\}$ and let us prove that it also holds for $k+1$. Performing the k -th step of the NE we have

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - \frac{a_{ik}^{(k)}}{a_{i-1,k}^{(k)}} a_{i-1,j}^{(k)}.$$

Then, by using the induction hypothesis, we obtain

$$\begin{aligned} a_{ij}^{(k+1)} &= a_{ij}^{(k)} - \frac{a_{ik}^{(k)}}{a_{i-1,k}^{(k)}} a_{i-1,j}^{(k)} \\ &= a_{ij}^{(k)} - \frac{(2i-2)(2i-3)}{(2i-k-1)(2i-k-2)} a_{i-1,j}^{(k)} \\ &= \frac{(i+j-k-1)!}{2^{j-1}(i-j)!(j-k)!} \prod_{r=1}^{k-1} \frac{(2i-r-1)}{(i-k+r)} \\ &\quad - \frac{(2i-2)(2i-3)}{(2i-k-1)(2i-k-2)} \frac{(i+j-k-2)!}{2^{j-1}(i-j-1)!(j-k)!} \prod_{r=1}^{k-1} \frac{(2i-r-3)}{(i-k+r-1)}. \end{aligned}$$

Simplifying the previous formula, $a_{ij}^{(k+1)}$ can be written as

$$a_{ij}^{(k+1)} = \frac{(i + j - k - 2)!}{2^{j-1}(i - j - 1)!(j - k)!} \prod_{r=1}^{k-1} \frac{(2i - r - 1)}{(i - k + r)} \left(\frac{i + j - k - 1}{i - j} - \frac{i - 1}{i - k} \right).$$

From the previous expression we can deduce that

$$\begin{aligned} a_{ij}^{(k+1)} &= \frac{(i + j - k - 2)!}{2^{j-1}(i - j - 1)!(j - k)!} \prod_{r=1}^{k-1} \frac{(2i - r - 1)}{(i - k + r)} \cdot \frac{2ji - 2ki - kj - j + k^2 + k}{(i - j)(i - k)} \\ &= \frac{(i + j - k - 2)!}{2^{j-1}(i - j - 1)!(j - k)!} \prod_{r=1}^{k-1} \frac{(2i - r - 1)}{(i - k + r)} \cdot \frac{(j - k)(2i - k - 1)}{(i - j)(i - k)} \\ &= \frac{(i + j - k - 2)!}{2^{j-1}(i - j)!(j - k - 1)!} \prod_{r=1}^k \frac{(2i - r - 1)}{(i - k + r - 1)}. \end{aligned}$$

Therefore, (14) holds for $k + 1$ and the result follows.

The pivot $p_{ij} = \tilde{a}_{ij}^{(j)} = a_{ij}^{(j)}$ is given by (14) with $k = j$ and we have that, for $i > j$, $m_{ij} = \frac{p_{ij}}{p_{i-1,j}}$, obtaining formulas (11) and (12), respectively.

- (ii) The lower triangular matrix A is nonsingular since it has nonzero diagonal entries. It can be seen in the proof of (i) that the NE of A satisfies the hypotheses of Theorem 1. Since U^T is a diagonal matrix, the NE of U^T obviously satisfies the hypotheses of Theorem 1. Hence, we can conclude that A is a TP matrix.
- (iii) By (i) and taking into account that U is a diagonal matrix with diagonal entries a_{ii} ($1 \leq i \leq n$), it is straightforward to deduce that $\mathcal{BD}(A)$ is given by (13). The subtractions in this formula are of integers and, hence, they can be computed to HRA, in fact, in an exact way.

□

Let us introduce the collocation matrices of the Bessel polynomials.

Definition 1 Given a sequence of parameters $0 < t_0 < t_1 < \dots < t_{n-1}$ we call the collocation matrix of the Bessel polynomials (B_0, \dots, B_{n-1}) at that sequence,

$$M = M \begin{pmatrix} B_0, \dots, B_{n-1} \\ t_0, \dots, t_{n-1} \end{pmatrix} = (B_{j-1}(t_{i-1}))_{1 \leq i, j \leq n},$$

a *Bessel matrix*.

The following result proves that the Bessel matrices are STP and that some usual algebraic problems with these matrices can be solved to HRA.

Theorem 4 Given a sequence of parameters $0 < t_0 < t_1 < \dots < t_{n-1}$, the corresponding Bessel matrix M is an STP matrix and given the parametrization t_i ($0 \leq i \leq n - 1$), the following computations can be performed with HRA: all the eigenvalues, all the singular values, the inverse of the Bessel matrix M , and the solution of the linear systems $Mx = b$, where $b = (b_1, \dots, b_n)^T$ has alternating signs.

Proof By formula (9) we have that

$$M = VA^T,$$

where M is the Bessel matrix corresponding to the collocation matrix of the Bessel polynomials (B_0, \dots, B_{n-1}) at t_0, \dots, t_{n-1} , A is the lower triangular matrix defined by (10) and V is the Vandermonde matrix corresponding to the collocation matrix of the monomial basis of degree $n - 1$ at t_0, \dots, t_{n-1} . Since V is a Vandermonde matrix with strictly increasing positive nodes, it is STP (see page 111 of [8] and page 12 of [7]). A^T is nonsingular TP because A is nonsingular TP by Theorem 3 (ii). Then, by Theorem 3.1 of [2], the Bessel matrix M is STP because it is the product of an STP matrix and a nonsingular TP matrix.

In Section 5.2 of [14] Koev devised an algorithm (Algorithm 5.1 in the reference) that, given the bidiagonal decompositions $\mathcal{BD}(C)$ and $\mathcal{BD}(D)$ to HRA of two TP matrices C and D , provides the bidiagonal decomposition $\mathcal{BD}(CD)$ to HRA of the TP product matrix CD . Since the bidiagonal factorization $\mathcal{BD}(V)$ of the Vandermonde matrix V is known to HRA (see Section 3 of [13]) and the bidiagonal factorization $\mathcal{BD}(A^T)$ of the matrix A can be computed to HRA by Theorem 3 (iii) and Remark 1, by using Algorithm 5.1 of [14] the bidiagonal factorization $\mathcal{BD}(M) = \mathcal{BD}(VA^T)$ of the Bessel matrix M is obtained to HRA.

Finally, the construction of $\mathcal{BD}(M)$ with HRA guarantees that the algebraic computations mentioned in the statement of this theorem can be performed with HRA (see Sect. 2 of this paper or Section 3 of [14]). \square

A system of functions is STP when all its collocation matrices are STP. The following result is a straightforward consequence of the previous theorem.

Corollary 1 *The system of functions formed by the Bessel polynomials of degree less than n , $(B_0(x), B_1(x), \dots, B_{n-1}(x))$, $x \in (0, +\infty)$, is an STP system.*

4 Reverse Bessel Polynomials and Matrices

Reversing the order of the coefficients of $B_n(x)$ in (8) we can define the *reverse Bessel polynomials*:

$$B_n^r(x) = \sum_{k=0}^n \frac{(n+k)!}{2^k(n-k)!k!} x^{n-k}, \quad n = 0, 1, 2, \dots, \quad (15)$$

Let $C = (c_{ij})_{1 \leq i, j \leq n}$ be the lower triangular matrix such that

$$(B_0^r(x), B_1^r(x), \dots, B_{n-1}^r(x))^T = C(1, x, \dots, x^{n-1})^T, \quad (16)$$

that is, the lower triangular matrix C is defined by

$$c_{ij} = \begin{cases} \frac{(2i-j-1)!}{2^{i-j}(j-1)!(i-j)!}, & i \geq j, \\ 0, & i < j. \end{cases} \quad (17)$$

Theorem 5 proves the total positivity of C , and provides $\mathcal{BD}(C)$. In addition, its proof gives the explicit form of all the entries of the matrices $C^{(k)}$ computed through the NE of C .

Theorem 5 *Let $C = (c_{ij})_{1 \leq i, j \leq n}$ be the lower triangular matrix in (16) defined by (17). Then, we have that*

(i) *the pivots of the NE of C are given by*

$$\begin{aligned} p_{ij} &= \frac{(2i-2j)!}{2^{i-j}(i-j)!} & 1 \leq j \leq i \leq n & \quad \text{if } j \text{ is odd,} \\ p_{ij} &= 0 & 1 \leq j < i \leq n, \quad p_{jj} = 1 & \quad 1 \leq j \leq n \quad \text{if } j \text{ is even,} \end{aligned} \quad (18)$$

and the multipliers by

$$\begin{aligned} m_{ij} &= 2i - 1 - 2j \quad 1 \leq j < i \leq n \quad \text{if } j \text{ is odd,} \\ m_{ij} &= 0, \quad 1 \leq j < i \leq n \quad \text{if } j \text{ is even,} \end{aligned} \quad (19)$$

- (ii) C is a nonsingular TP matrix
 (iii) and the bidiagonal factorization of C is given by

$$\mathcal{BD}(C)_{ij} = \begin{cases} 2i - 2j - 1, & \text{if } i > j \text{ with } j \text{ odd,} \\ 1, & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases} \quad (20)$$

and can be computed to HRA.

Proof (i) Let us perform the first step of the NE of C :

$$c_{ij}^{(2)} = c_{ij} - \frac{c_{i1}}{c_{i-1,1}} c_{i-1,j} = c_{ij} - (2i - 3)c_{i-1,j}, \quad i > j \geq 1.$$

By using (17) in the previous expression and simplifying we have that

$$\begin{aligned} c_{ij}^{(2)} &= \frac{(2i - j - 1)!}{2^{i-j}(j-1)!(i-j)!} - \frac{(2i-3)(2i-j-3)!}{2^{i-j-1}(j-1)!(i-j-1)!} \\ &= \frac{(2i-j-3)!}{2^{i-j-1}(j-1)!(i-j-1)!} \left[\frac{(2i-j-2)(2i-j-1)}{2(i-j)} - 2i + 3 \right] \\ &= \frac{(2i-j-3)!}{2^{i-j}(j-1)!(i-j)!} (j^2 - 3j + 2) \\ &= \frac{(2i-j-3)!}{2^{i-j}(j-1)!(i-j)!} (j-2)(j-1). \end{aligned}$$

From the previous formula we deduce that

$$c_{ij}^{(2)} = \begin{cases} \frac{(2i-j-3)!}{2^{i-j}(j-3)!(i-j)!}, & \text{for } i > j \geq 3, \\ 0, & \text{for } i > j \leq 2, \\ c_{ij}^{(1)}, & \text{in otherwise.} \end{cases} \quad (21)$$

Since $c_{i2}^{(2)} = 0$ for $i = 3, \dots, n$, we have that $C^{(3)} = C^{(2)}$. So, taking into account this fact, formula (21) and formula (17), we can observe that $c_{ij}^{(3)} = c_{i-2,j-2}$ and so, performing two steps of the NE of C gives as a result a leading principal submatrix of C . In particular, C satisfies that $C^{(2)}[3, \dots, n] = C^{(3)}[3, \dots, n] = C[1, \dots, n-2]$. Hence, and by formulas (3) and (4), we deduce formulas (18) and (19).

- (ii) The lower triangular matrix C is nonsingular since it has nonzero diagonal entries. It can be seen in the proof of (i) that the NE of C satisfies the hypotheses of Theorem 1. Since the upper triangular matrix obtained after the process (1) of the NE of C is a diagonal matrix, the NE of its transpose obviously satisfies the hypotheses of Theorem 1. Hence, we can conclude that C is a TP matrix.
- (iii) By (i) it is straightforward to deduce that $\mathcal{BD}(C)$ is given by (20). The subtractions in this formula are of integers and, hence, they can be computed to HRA, in fact, in an exact way.

□

Definition 2 Given a sequence of parameters $0 < t_0 < t_1 < \dots < t_{n-1}$ we call the collocation matrix of the reverse Bessel polynomials $(B_0^r, \dots, B_{n-1}^r)$ at that sequence,

$$M_r = M \begin{pmatrix} B_0^r, \dots, B_{n-1}^r \\ t_0, \dots, t_{n-1} \end{pmatrix} = (B_{j-1}^r(t_{i-1}))_{1 \leq i, j \leq n}$$

a *reverse Bessel matrix*.

The following result proves that the reverse Bessel matrices are STP and that some usual algebraic problems with these matrices can be solved to HRA.

Theorem 6 *Given a sequence of parameters $0 < t_0 < t_1 < \dots < t_{n-1}$, the corresponding reverse Bessel matrix M_r is an STP matrix and given the parametrization t_i ($0 \leq i \leq n-1$), the following computations can be performed with HRA: all the eigenvalues, all the singular values, the inverse of the reverse Bessel matrix M_r , and the solution of the linear systems $M_r x = b$, where $b = (b_1, \dots, b_n)^T$ has alternating signs.*

Proof The results can be proved in an analogous way to those of Theorem 4. □

The following result is a straightforward consequence of the previous theorem.

Corollary 2 *The system of functions formed by the reverse Bessel polynomials of degree less than n , $(B_0^r(x), B_1^r(x), \dots, B_{n-1}^r(x))$, $x \in (0, +\infty)$, is an STP system.*

5 Numerical Experiments

In [14], assuming that the parameterization $\mathcal{BD}(A)$ of an square TP matrix A is known with HRA, Plamen Koev presented algorithms to compute $\mathcal{BD}(A^{-1})$, the eigenvalues and the singular values of A , and the solution of linear systems of equations $Ax = b$ where b has an alternating pattern of signs to HRA. Koev also implemented these algorithms in order to be used with Matlab and Octave in the software library *TNTool* available in [15]. The corresponding functions are `TNInverse`, `TNEigenvalues`, `TNSingularValues` and `TNSolve`, respectively. The functions require as input argument the data determining the bidiagonal decomposition (5) of A , $\mathcal{BD}(A)$ given by (7), to HRA. `TNSolve` also requires a second argument, the vector b of the linear system $Ax = b$ to be solved. In addition, recently a function `TNInverseExpand` was added to that library, contributed by Ana Marco and José-Javier Martínez. This function, given $\mathcal{BD}(A)$ to HRA, returns A^{-1} to HRA.

The library *TNTool* also provides the function `TNProduct(B1, B2)`, which, given the bidiagonal decompositions $B1$ and $B2$ to HRA of two TP matrices F and G , provides the bidiagonal decomposition of the TP matrix FG to HRA. We can observe in the factorization $M = VA^T$ in the proof of Theorem 4 that M can be expressed as the product of two TP matrices: the TP Vandermonde matrix V and the TP matrix A^T defined by (9) and (10). Taking into account Remark 1, the bidiagonal factorization of A^T to HRA can be obtained from Theorem 3(iii). Since V is a TP Vandermonde matrix, $\mathcal{BD}(V)$ is obtained to HRA by using `TNVandBD` of library *TNTool*. Taking into account these facts, the pseudocode providing $\mathcal{BD}(M)$ to HRA can be seen in Algorithm 1.

We have implemented the previous algorithm to be used in Matlab and Octave in a function `TNBDBessel`. The bidiagonal decompositions with HRA obtained with this function can be used with `TNInverseExpand`, `TNEigenValues`, `TNSingularValues` and `TNSolve` in order to obtain accurate solutions for the above mentioned algebraic problems.

Algorithm 1 Computation of the bidiagonal decomposition of M to HRA**Require:** $\mathbf{t} = (t_i)_{i=0}^{n-1}$ such that $0 < t_0 < t_1 < \dots < t_{n-1}$ **Ensure:** B bidiagonal decomposition of M to HRA $B1 = TNV$ and $BD(\mathbf{t})$ $B2(1, 1) = 1$ **for** $i = 2 : n - 1$ **do** **for** $j = 1 : i - 1$ **do**

$$B2(i, j) = \frac{(2i-2)(2i-3)}{(2i-j-1)(2i-j-2)}$$

end for

$$B2(i, i) = (2i - 3)!!$$

for $j = i + 1 : n - 1$ **do**

$$B2(i, j) = 0$$

end for**end for** $B = TNProduct(B1, B2^T)$

Now we include some numerical tests illustrating the high accuracy of the new methods in contrast to the accuracy of the usual methods.

First we have considered the Bessel matrix of order 20, M_{20} , corresponding to the collocation matrix of the Bessel polynomials of degree at most 19 at points $1, 2, \dots, 20$. We have computed with MATLAB the bidiagonal decomposition of M_{20} to HRA with the function `TNBDBessel`. Then, using this bidiagonal decomposition, we have computed approximations to its eigenvalues and its singular values with `TNEigenValues` and `TNSingularValues`, respectively. We have also computed approximations to the eigenvalues and singular values with the MATLAB functions `eig` and `svd`, respectively. Then we have also computed the eigenvalues and singular values of M_{20} with a precision of 100 digits using Mathematica. Taking as exact the eigenvalues and the singular values obtained with Mathematica, we have computed the relative errors for the approximation to the eigenvalues (resp. singular values) obtained by both `eig` (resp., `svd`) and `TNEigenValues` (resp., `TNSingularValues`).

Since a Bessel matrix is STP, by Theorem 6.2 of [2] all its eigenvalues are real, distinct and positive. Taking into account this fact, the eigenvalues of M_{20} have been ordered as $\lambda_1 > \lambda_2 > \dots > \lambda_{20} > 0$. The relative errors for the approximations to these eigenvalues of M_{20} can be seen in Table 1. We can observe in this table that the approximations obtained by using the bidiagonal decomposition are very accurate. In contrast, the approximations obtained with `eig` MATLAB function are only accurate for the larger eigenvalues of M_{20} . In fact, the approximations to the eigenvalues of M_{20} obtained with `eig` are not even positive for the smaller ones.

The 20 real and positive singular values of M_{20} have also been ordered as $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{20} > 0$. The relative errors for the approximations to these singular values of M_{20} can be seen in Table 2. As in the case of the eigenvalues, the approximation obtained for all the singular values using the bidiagonal decomposition are very accurate, but only the approximations obtained for the larger singular values by using `svd` are accurate.

We have also computed approximations to the inverse $(M_{20})^{-1}$ with both `inv` and `TNInverseExpand` using the bidiagonal decomposition provided by `TNBDBessel`. We have also obtained with Mathematica the inverse using exact arithmetic. Then, we have computed the componentwise relative errors for both approximations. In Table 3 the mean and maximum componentwise relative errors are shown. It can be seen that the inverse obtained with the HRA methods is very accurate in contrast to the poor approximation obtained with `inv`.

Table 1 Relative errors for the eigenvalues of the Bessel matrix M_{20}

i	λ_i	Rel. errors with HRA	Rel. errors with eig
1	4.5222e+46	4.4851e-16	1.1213e-16
2	1.2183e+42	2.5402e-16	8.0016e-15
3	7.7264e+37	2.4448e-16	3.3983e-14
4	8.7322e+33	0	2.1003e-11
5	1.5801e+30	7.1256e-16	5.3330e-10
\vdots	\vdots	\vdots	\vdots
17	1.1529e+00	0	2.4583e+10
18	1.3072e-01	2.1233e-16	2.2085e+13
19	6.1386e-03	4.2389e-16	1.8425e+22
20	1.2006e-04	3.3864e-16	4.2167e+28

Table 2 Relative errors for the singular values of the Bessel matrix M_{20}

i	σ_i	Rel. errors with HRA	Rel. errors with svd
1	4.8763e+46	2.0797e-15	4.1594e-16
2	1.5204e+42	6.1065e-16	4.0710e-16
3	1.1076e+38	3.4108e-16	9.1922e-14
4	1.4266e+34	8.0818e-16	8.3458e-11
5	2.9165e+30	3.8604e-16	1.2661e-03
\vdots	\vdots	\vdots	\vdots
17	1.0795e+00	4.1139e-16	5.0876e+11
18	1.5106e-02	2.1818e-15	3.4860e+12
19	9.1285e-05	2.0785e-15	1.2718e+13
20	1.6258e-07	3.2563e-16	1.9230e+06

Table 3 Relative errors for the inverse of the Bessel matrix M_{20}

	Rel. errors with HRA	Rel. errors with inv
Mean	1.8498e-16	2.1364e-01
Maximum	8.4304e-16	3.0878e-01

We have considered two systems of linear equations

$$M_{20}x = b^1 \quad \text{and} \quad M_{20}x = b^2,$$

where the entries of b^2 are randomly generated as integers in the interval $[1, 1000]$ and the i -th entry of b^1 is given by $b_i^1 = (-1)^{i+1}b_i^2$ for $i = 1, \dots, 20$. So, the independent vector of the system $M_{20}x = b^1$ has an alternating pattern of signs and the linear system can be solved with HRA by Theorem 4. Approximations \hat{x} to the solutions x of both linear systems have been obtained with MATLAB, the first one using `TNSolve` and the bidiagonal decomposition of the Bessel matrix obtained with `TNBDBessel`, and the second one using the usual MATLAB command `A\b`. By using Mathematica with exact arithmetic, the exact

Table 4 Relative errors for the solution of the linear system $M_{20}x = b^1$

i	$\left \frac{\widehat{x}_i - x_i}{x_i} \right $ with HRA	$\left \frac{\widehat{x}_i - x_i}{x_i} \right $ with $A \setminus b$
1	5.6243e−16	1.8543e−01
2	1.8069e−16	1.8643e−01
3	0	1.8822e−01
4	1.2831e−16	1.9058e−01
5	2.2120e−16	1.9336e−01
⋮	⋮	⋮
17	0	2.3342e−01
18	1.4910e−16	2.3627e−01
19	2.3844e−16	2.3897e−01
20	0	2.4153e−01

Table 5 Relative errors for the solution of the linear system $M_{20}x = b^2$

i	$\left \frac{\widehat{x}_i - x_i}{x_i} \right $ with TNSolve	$\left \frac{\widehat{x}_i - x_i}{x_i} \right $ with $A \setminus b$
1	1.6814e−16	8.9852e−01
2	1.2521e−16	8.3698e−01
3	1.7413e−16	7.4661e−01
4	2.7288e−16	6.5449e−01
5	2.0323e−16	5.7258e−01
⋮	⋮	⋮
17	1.8808e−16	2.1570e−01
18	1.6663e−16	2.0699e−01
19	2.5411e−16	1.9933e−01
20	0	1.9257e−01

solution of the systems have been computed, and then, the componentwise relative errors for the two approximations obtained with MATLAB have been computed. Table 4 shows the componentwise relative errors corresponding to the system $M_{20}x = b^1$. It can be observed that the approximation to the solution provided by TNSolve is very accurate. This fact can be expected since the independent vector b^1 of the system has an alternating pattern of signs and then it is known that TNSolve provides the solution to HRA (see [14]). On the other hand, it can also be observed in the table that the accuracy of the approximation provided by $A \setminus b$ is very poor.

Table 5 shows the componentwise relative errors corresponding to the system $M_{20}x = b^2$. In this case, since the independent vector b^2 has not an alternating pattern of signs, it is not guaranteed to obtain an approximation to HRA by using TNSolve. However, it can be observed that in this case the approximation to the solution provided by TNSolve is also very accurate in contrast to the poor accuracy of the approximation provided by $A \setminus b$.

It can be observed that the smaller an eigenvalue (resp., singular value) is, the larger the relative error corresponding to the usual methods is. So, now let us consider the Bessel matrices M_n of order n , for $n = 2, \dots, 15$ given by the collocation matrices of the Bessel polynomials $(B_0(x), \dots, B_{n-1}(x))$ at the points $1, \dots, n$, that is, $M_n = (B_{j-1}(i))_{1 \leq i, j \leq n}$. In

the same way that in the previous examples we have computed the eigenvalues, the singular values and the inverses of these matrices both with the usual MATLAB functions and to HRA by using `TNBDessel`. Then we have computed the relative errors for the approximation to the smallest eigenvalue and the smallest singular value of each matrix, and the componentwise relative error for the approximations to the inverses.

The relative errors for the smallest eigenvalues and the smallest singular values of the Bessel matrices M_n , $n = 2, \dots, 15$, can be seen in Fig. 1a, b, respectively.

The mean and the maximum componentwise relative errors corresponding to the approximation of the inverses $(M_n)^{-1}$ can be seen in Fig. 2a, b, respectively.

In an analogous way to the Bessel matrix we can derive an algorithm to obtain the bidiagonal decomposition of a reverse Bessel matrix to HRA. So, the pseudocode providing $BD(M_r)$ to HRA can be seen in Algorithm 2.

Algorithm 2 Computation of the bidiagonal decomposition of M_r to HRA

Require: $\mathbf{t} = (t_i)_{i=0}^{n-1}$ such that $0 < t_0 < t_1 < \dots < t_{n-1}$

Ensure: B bidiagonal decomposition of M_r to HRA

$B1 = TNV$ and $BD(\mathbf{t})$

for $i = 1 : n - 1$ **do**

for $j = 1 : i - 1$ **do**

if j is odd **then**

$B2(i, j) = 2i - 2j - 1$

end if

end for

$B2(i, i) = 1$

for $j = i + 1 : n - 1$ **do**

$B2(i, j) = 0$

end for

end for

$B = TNProduct(B1, B2^T)$

For the reverse Bessel matrices we have carried out the same numerical tests as for the Bessel matrices and we have deduced exactly the same conclusions. For the sake of brevity, for the reverse Bessel matrices only the relative errors for the smallest eigenvalue and singular

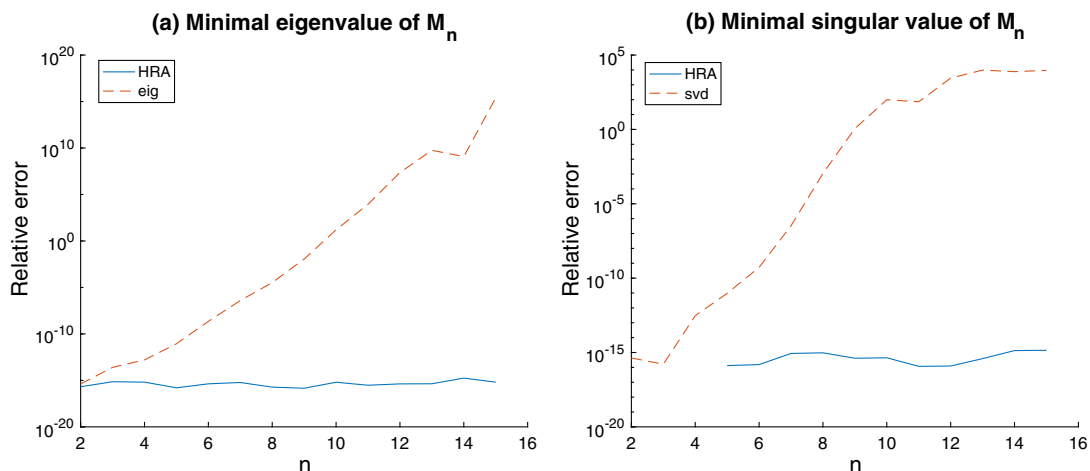


Fig. 1 Relative error for the minimal eigenvalue and singular value of M_n

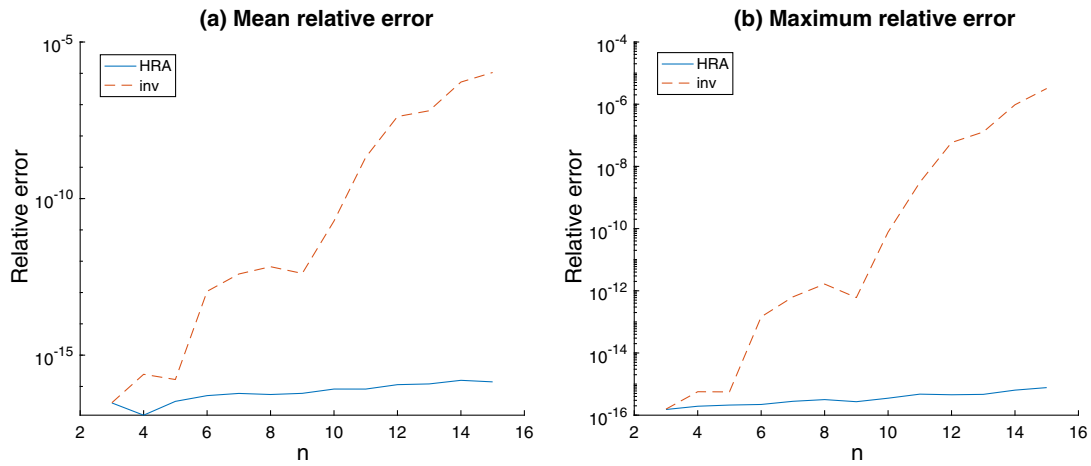


Fig. 2 Relative errors for $(M_n)^{-1}$

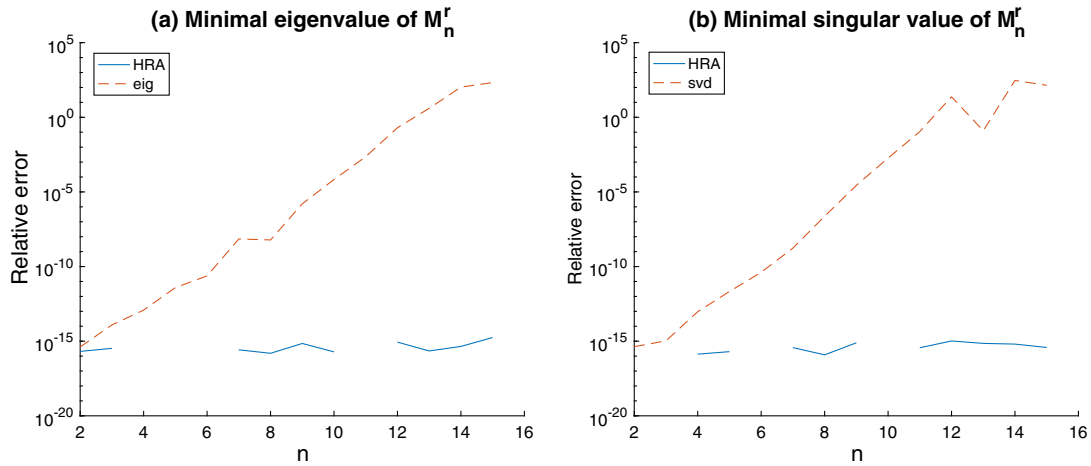


Fig. 3 Relative error for the minimal eigenvalue and singular value of M_n^r

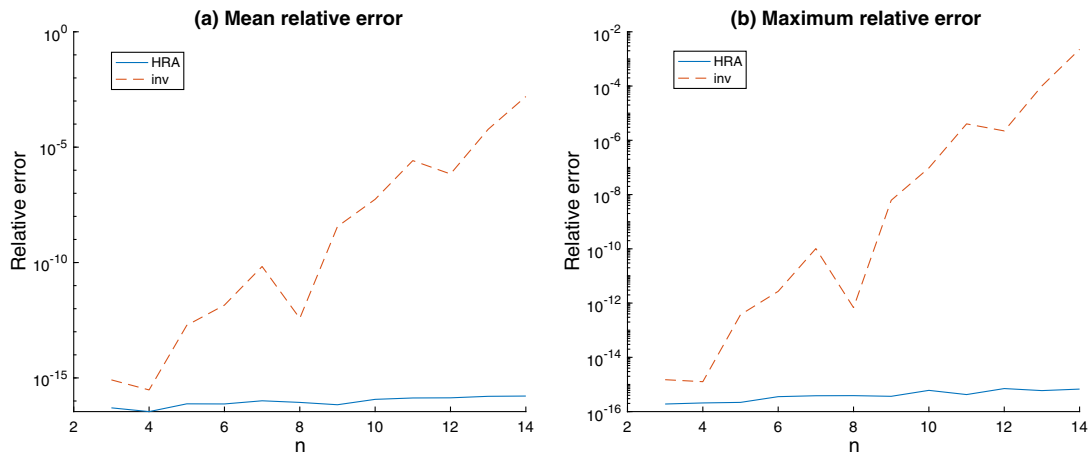


Fig. 4 Relative errors for $(M_n^r)^{-1}$

value, and the componentwise mean and maximum relative error for the inverses of the reverse Bessel matrices $M'_n = (B'_{j-1}(i))_{1 \leq i, j \leq n}$, $n = 2, \dots, 15$, are shown in Figs. 3 and 4, respectively.

Acknowledgements This work was partially supported through the Spanish research grant PGC2018-096321-B-I00 (MCIU/AEI), by Gobierno de Aragón (E41_17R) and by Feder 2014-2020 “Construyendo Europa desde Aragón”.

References

1. Alonso, P., Delgado, J., Gallego, R., Peña, J.M.: Conditioning and accurate computations with Pascal matrices. *J. Comput. Appl. Math.* **252**, 21–26 (2013)
2. Ando, T.: Totally positive matrices. *Linear Algebra Appl.* **90**, 165–219 (1987)
3. Carnicer, J.M., Mainar, E., Peña, J.M.: Critical lengths of cycloidal spaces are zeros of Bessel functions. *Calcolo* **54**, 1521–1531 (2017)
4. Delgado, J., Peña, J.M.: Fast and accurate algorithms for Jacobi–Stirling matrices. *Appl. Math. Comput.* **236**, 253–259 (2014)
5. Delgado, J., Peña, J.M.: Accurate computations with collocation matrices of q-Bernstein polynomials. *SIAM J. Matrix Anal. Appl.* **36**, 880–893 (2015)
6. Demmel, J., Koev, P.: The accurate and efficient solution of a totally positive generalized Vandermonde linear system. *SIAM J. Matrix Anal. Appl.* **27**, 142–152 (2005)
7. Fallat, S.M., Johnson, C.R.: *Totally Nonnegative Matrices*. Princeton Series in Applied Mathematics, vol. 35. Princeton University Press, Princeton (2011)
8. Gantmacher, F.R., Krein, M.G.: *Oszillationsmatrizen, oszillationskerne und kleine schwingungen mechanischer systeme*. Akademie, Berlin (1960)
9. Gasca, M., Peña, J.M.: Total positivity and Neville elimination. *Linear Algebra Appl.* **165**, 25–44 (1992)
10. Gasca, M., Peña, J.M.: On factorizations of totally positive matrices. In: Gasca, M., Micchelli, C.A. (eds.) *Total Positivity and Its Applications*, pp. 109–130. Kluwer Academic Publishers, Dordrecht (1996)
11. Grosswald, E.: *Bessel Polynomials*. Springer, New York (1978)
12. Han, H., Seo, S.: Combinatorial proofs of inverse relations and log-concavity for Bessel numbers. *Eur. J. Combin.* **29**, 1544–1554 (2008)
13. Koev, P.: Accurate eigenvalues and SVDs of totally nonnegative matrices. *SIAM J. Matrix Anal. Appl.* **27**, 1–23 (2005)
14. Koev, P.: Accurate computations with totally nonnegative matrices. *SIAM J. Matrix Anal. Appl.* **29**, 731–751 (2007)
15. Koev, P.: <http://www.math.sjsu.edu/~koev/software/TNTool.html>. Accessed November 12th (2018)
16. Krall, H.L., Frink, O.: A new class of orthogonal polynomials: the Bessel polynomials. *Trans. Am. Math. Soc.* **65**, 100–115 (1949)
17. Marco, A., Martínez, J.-J.: A fast and accurate algorithm for solving Bernstein–Vandermonde linear systems. *Linear Algebra Appl.* **422**, 616–628 (2007)
18. Marco, A., Martínez, J.-J.: Accurate computations with Said–Ball–Vandermonde matrices. *Linear Algebra Appl.* **432**, 2894–2908 (2010)
19. Marco, A., Martínez, J.-J.: Accurate computations with totally positive Bernstein–Vandermonde matrices. *Electron. J. Linear Algebra* **26**, 357–380 (2013)
20. Pasquini, L.: Accurate computation of the zeros of the generalized Bessel polynomials. *Numer. Math.* **86**, 507–538 (2000)
21. Pinkus, A.: *Totally Positive Matrices*. Tracts in Mathematics, vol. 181. Cambridge University Press, Cambridge (2010)
22. Yang, S.L., Qiao, Z.K.: The Bessel numbers and Bessel matrices. *J. Math. Res. Expo.* **31**, 627–636 (2011)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Article 3

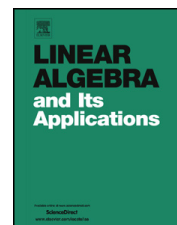
- [79] H. Orera and J. M. Peña. Accurate inverses of Nekrasov Z-matrices. *Linear Algebra Appl.* 574 (2019), 46-59.



Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa



Accurate inverses of Nekrasov Z -matrices[☆]

H. Orera^{*}, J.M. Peña

Departamento de Matemática Aplicada/IUMA, Universidad de Zaragoza, Spain



ARTICLE INFO

Article history:

Received 14 November 2017
Accepted 22 March 2019
Available online 26 March 2019
Submitted by A. Frommer

MSC:

15B48
15A09
15B35
65F05

Keywords:

Accuracy
Inverse
 H -matrices
Strictly diagonally dominant
matrices
Nekrasov matrices

ABSTRACT

We present a parametrization of a Nekrasov Z -matrix that allows us to compute its inverse with high relative accuracy. Numerical examples illustrating the accuracy of the method are included.

© 2019 Elsevier Inc. All rights reserved.

1. Introduction

Recent research in Numerical Linear Algebra has shown that, for some classes of structured matrices, some algebraic computations can be performed to high relative

[☆] This research has been partially supported by MTM2015-65433-P (MINECO/FEDER) Spanish Research Grant and by Gobierno de Aragón.

^{*} Corresponding author.

E-mail addresses: hectororera@unizar.es (H. Orera), jmpena@unizar.es (J.M. Peña).

accuracy (HRA), independently of the size of the classical condition number. These classes of matrices are defined by special sign or other structure. It is well-known (cf. p. 52 of [7]) that, if an algorithm is subtraction-free, its output can be computed to HRA. For these classes of matrices, knowing an adequate parametrization has been a crucial start point for the construction of the corresponding accurate algorithms, being many of them subtraction-free. In contrast to these classes of matrices, for other structured classes of matrices it is not possible to construct such HRA algorithms (cf. [6]).

In this paper, we present a parametrization for Nekrasov Z -matrices, which allows us to construct a subtraction-free (and so, HRA) efficient algorithm to compute their inverses.

Let us now recall some basic definitions on classes of matrices used in this paper. A real matrix A is a Z -matrix if all its off-diagonal entries are nonpositive. A Z -matrix A is a nonsingular M -matrix if its inverse is nonnegative. Given a complex matrix $A = (a_{ij})_{1 \leq i, j \leq n}$, its comparison matrix $\mathcal{M}(A) = (\tilde{a}_{ij})_{1 \leq i, j \leq n}$ has entries $\tilde{a}_{ii} := |a_{ii}|$ and $\tilde{a}_{ij} := -|a_{ij}|$ for all $j \neq i$ and $i, j = 1, \dots, n$. We say that a complex matrix is a nonsingular H -matrix if its comparison matrix is a nonsingular M -matrix. This concept corresponds with the concept of H -matrix of invertible class given in [4]. A matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ is SDD (strictly diagonally dominant by rows) if $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$ for all $i = 1, \dots, n$, and A is DD (diagonally dominant by rows) if $|a_{ii}| \geq \sum_{j \neq i} |a_{ij}|$ for all $i = 1, \dots, n$. It is well-known that an SDD matrix is nonsingular and that a square matrix A is a nonsingular H -matrix if and only if there exists a diagonal matrix W with positive diagonal entries such that AW is SDD . Nekrasov matrices (see [14]) are defined in Section 2 and form another subclass of H -matrices that includes SDD matrices. Some recent applications of Nekrasov matrices can be seen in [5], [10], [11] or [12].

Let us present the layout of the paper. Section 2 presents the parametrization of Nekrasov Z -matrices, some auxiliary results and the construction of the subtraction-free algorithms for the inverse of a Nekrasov Z -matrix in a particular case. The algorithm for a general Nekrasov Z -matrix A is constructed in Section 3. Section 4 includes some algorithms used in our method and presents numerical examples showing its accuracy. Our method also allows us to compute the solution of a linear system $Ax = b$ with $b \geq 0$ to HRA. The numerical examples also show great accuracy of our method even when b does not satisfy this requirement.

The following notations will be also used in this paper. A matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ (resp., a vector $v = (v_1, \dots, v_n)^T$) is nonnegative if $a_{ij} \geq 0$ for all i, j (resp., $v_i \geq 0$ for all i), and we write $A \geq 0$ (resp., $v \geq 0$).

2. Parametrization of Nekrasov matrices and HRA

Let us start by defining the concept of a Nekrasov matrix (see [5,14]). For this purpose, let us define recursively for a complex matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ with $a_{ii} \neq 0$, for all $i = 1, \dots, n$,

$$h_1(A) := \sum_{j \neq 1} |a_{1j}|, \quad h_i(A) := \sum_{j=1}^{i-1} |a_{ij}| \frac{h_j(A)}{|a_{jj}|} + \sum_{j=i+1}^n |a_{ij}|, \quad i = 2, \dots, n. \quad (1)$$

Let $N := \{1, \dots, n\}$. We say that A is a *Nekrasov matrix* if $|a_{ii}| > h_i(A)$ for all $i \in N$. A Nekrasov matrix is a nonsingular H -matrix [14]. Therefore, a Nekrasov Z -matrix with positive diagonal entries is a nonsingular M -matrix.

Remark 2.1. Let us recall that DD M -matrices admit some algebraic computations with high relative accuracy (HRA). A key tool is the use of an adequate parametrization of these matrices, which was provided by the off-diagonal entries and the row sums (cf. [1], [8], [13], [2]). We shall call these n^2 parameters for an $n \times n$ DD M -matrix A as *DD-parameters*. If these DD-parameters are known with HRA, then some algebraic computations of A can be performed with HRA as it is shown in the previous references.

In this paper we also study computations with HRA for the class of Nekrasov Z -matrices. Here, a good choice of parameters will also be crucial. The parameters that we shall use for an $n \times n$ Nekrasov Z -matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ with positive diagonal are the following n^2 parameters, which will be called *N -parameters*:

$$\begin{cases} a_{ij}, & i \neq j \\ \Delta_j(A) := a_{jj} - h_j(A), & j \in N \end{cases} \quad (2)$$

We can characterize an $n \times n$ Nekrasov Z -matrix with positive diagonal through the n^2 signs of the parameters given in (2). In fact, A is a Nekrasov Z -matrix with positive diagonal if and only if the first $n^2 - n$ parameters (corresponding to the off-diagonal entries, a_{ij} with $i \neq j$) are nonpositive and the last n parameters ($\Delta_j(A)$ for all $j \in N$) are positive.

Since a Nekrasov matrix is a nonsingular H -matrix, there exists a positive diagonal matrix W such that AW is *SDD*. The following lemma shows that the very simple diagonal matrix

$$S = \begin{pmatrix} \frac{h_1(A)}{a_{11}} & & & \\ & \frac{h_2(A)}{a_{22}} & & \\ & & \ddots & \\ & & & \frac{h_n(A)}{a_{nn}} \end{pmatrix} \quad (3)$$

holds that AS satisfies the weaker property of being *DD*.

Lemma 2.2. *Let A be a Nekrasov Z -matrix with positive diagonal and let S be the matrix given by (3). Then the matrix AS is a DD Z -matrix.*

Proof. Observe that $\frac{h_i(A)}{a_{ii}} \geq 0$ for $i \in N$, and so, $S \geq 0$. Then $B := AS$ preserves the signs of A , and the elements of $B = (B_{ij})_{1 \leq i, j \leq n}$ are:

$$B_{ij} = \begin{cases} a_{ij} \frac{h_i(A)}{a_{jj}}, & \text{if } i \neq j, \\ h_i(A), & \text{if } i = j. \end{cases}$$

Since A is a Z -matrix, B is also a Z -matrix. It remains to prove that B is also DD. Since A is a Nekrasov matrix, $h_j(A) < a_{jj}$ for all $j \in N$. For each $i \in N$,

$$h_i(A) = \sum_{j=1}^{i-1} |a_{ij}| \frac{h_j(A)}{a_{jj}} + \sum_{j=i+1}^n |a_{ij}| \geq \sum_{j=1}^{i-1} |a_{ij}| \frac{h_j(A)}{a_{jj}} + \sum_{j=i+1}^n |a_{ij}| \frac{h_j(A)}{a_{jj}}$$

and so B is DD. \square

For a Nekrasov Z -matrix A and the diagonal matrix S given by (3), the following result shows that if we know the n^2 N-parameters in (2) of A , then we can compute the n^2 DD-parameters of the DD M -matrix AS with HRA. This fact will allow us to take advantage of properties of DD M -matrices to obtain algorithms with HRA for Nekrasov Z -matrices.

Theorem 2.3. *Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a Nekrasov Z -matrix with positive diagonal entries and let S be the matrix given by (3). Given the n^2 N-parameters (2), we can compute the row sums and the off-diagonal entries of AS (its DD-parameters) by a subtraction-free algorithm (and so, with HRA), with at most $\frac{3n(n-1)}{2}$ additions, $2n(n-1)$ multiplications and $2n-1$ quotients.*

Proof. Observe that by (2),

$$a_{jj} = \Delta_j(A) + h_j(A), \quad j \in N. \tag{4}$$

Let us start by computing $h_1(A), a_{11}, h_2(A), a_{22}, \dots, h_n(A), a_{nn}$ using the formulas (4) and (1). We carry out n sums computing the diagonal entries by (4), n quotients in order to obtain $\frac{h_j(A)}{a_{jj}}$ when needed (and we store them) and $\frac{(n-1)n}{2}$ products and $n(n-2)$ sums to calculate $h_j(A)$ for all $j \in N$ using (1). Then we obtain the off-diagonal entries of $AS, a_{ij} \frac{h_j(A)}{a_{jj}}$, which requires $n(n-1)$ products. Finally, we compute the row sums of AS . The row sum of the i th row is:

$$\sum_{j=1}^{i-1} a_{ij} \frac{h_j(A)}{a_{jj}} + h_i(A) + \sum_{j=i+1}^n a_{ij} \frac{h_j(A)}{a_{jj}},$$

which can be expressed in the following form using (1), (2) and the sign pattern of a Z -matrix:

$$\sum_{j=i+1}^n (-a_{ij}) \left(1 - \frac{h_j(A)}{a_{jj}} \right) = \sum_{j=i+1}^n |a_{ij}| \frac{a_{jj} - h_j(A)}{a_{jj}} = \sum_{j=i+1}^n |a_{ij}| \frac{\Delta_j(A)}{a_{jj}}. \tag{5}$$

Computing the row sums requires $n - 1$ quotients of the form $\frac{\Delta_j(A)}{a_{jj}}$ for $j = 2, \dots, n$, $\frac{n(n-1)}{2}$ sums and $\frac{n(n-1)}{2}$ products. The total number of required operations is at most $\frac{3n(n-1)}{2}$ additions, $2n(n - 1)$ multiplications and $2n - 1$ quotients. We do not perform any subtraction in this procedure and so it is subtraction-free. \square

Let us introduce some basic notations related with Gaussian and Gauss–Jordan elimination. Gaussian elimination without pivoting for a nonsingular $n \times n$ matrix A consists of a procedure of at most $n - 1$ steps resulting in the following sequence of matrices:

$$A =: A^{(1)} \longrightarrow A^{(2)} \longrightarrow \dots \longrightarrow A^{(n)}, \quad (6)$$

where $A^{(t)}$ has zeros below its main diagonal in the first $(t - 1)$ columns and $A^{(n)}$ is an upper triangular matrix. To obtain $A^{(t+1)}$ from $A^{(t)}$ we produce zeros in column t below the pivot element $a_{tt}^{(t)}$ by subtracting adequate multiples of row t from the rows beneath it. The same transformation can be performed with the matrix $(A \mid B^{(1)})$, where $B^{(1)} := I$ is the identity matrix,

$$(A \mid I) =: \left(A^{(1)} \mid B^{(1)} \right) \longrightarrow \left(A^{(2)} \mid B^{(2)} \right) \longrightarrow \dots \longrightarrow \left(A^{(n)} \mid B^{(n)} \right). \quad (7)$$

Now we proceed analogously, starting from the last row and producing zeros above the main diagonal of $A^{(k)}$ ($n \leq k \leq 2n - 1$) to obtain the sequence:

$$\left(A^{(n)} \mid B^{(n)} \right) \longrightarrow \dots \longrightarrow \left(A^{(2n-1)} \mid B^{(2n-1)} \right) \longrightarrow \left(A^{(2n)} \mid B^{(2n)} \right) =: (I \mid A^{-1}). \quad (8)$$

In this case, $A^{(t)} = (a_{ij}^{(t)})_{1 \leq i, j \leq n}$, $t = n + 1, \dots, 2n - 1$, has zeros above its main diagonal in the last $(t - n)$ columns. To obtain $A^{(t+1)}$ from $A^{(t)}$, $t = n, \dots, 2n - 1$, we produce zeros in column $2n - t$ above the pivot element $a_{2n-t, 2n-t}^{(t)}$ by subtracting multiples of row $2n - t$ from the rows above it. Finally, $A^{(2n)} = I$ is obtained from $A^{(2n-1)}$ by dividing each row of $A^{(2n-1)}$ by its diagonal entries. This well-known method is called Gauss–Jordan elimination.

Let $Q_{k,n}$ be the set of increasing sequences of k positive integers in N . Given $\alpha, \beta \in Q_{k,n}$, we denote by $A[\alpha|\beta]$ the $k \times k$ submatrix of A containing rows numbered by α and columns numbered by β . If $\alpha = \beta$, then we have the principal submatrix $A[\alpha] := A[\alpha|\alpha]$. The complement α^C is the increasingly rearranged $N \setminus \alpha$.

For DD M -matrices, algorithms with HRA starting from their DD-parametrization were presented in [8] and [13]. In both papers, Gaussian elimination is used, but with a different pivoting strategy in each of them. In order to obtain the inverse with HRA, a pivoting strategy is not necessary, as the following result shows.

Proposition 2.4. *Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a DD nonsingular Z -matrix with positive diagonal entries. If we know the row sums and the off-diagonal entries of A (i.e., its DD-parameters), then we can compute A^{-1} with a subtraction-free algorithm (and so, with HRA) performing $\mathcal{O}(n^3)$ elementary operations.*

Proof. By the hypotheses, A is nonsingular and $A + D$ is SDD (and so, nonsingular) for any positive diagonal matrix D . Then, by using property (C_{10}) of Theorem 2.2 of chapter 6 of [3] we deduce that A is a nonsingular M -matrix. In order to obtain A^{-1} with HRA, we are going to use Gauss–Jordan elimination without pivoting. We form the augmented matrix $\tilde{M} := (A|I|s)$, which coincides with the first matrix in (7) but with a last column with the vector s formed by the row sums of A : $s = (s_1, \dots, s_n)^T$ and, for $i = 1, \dots, n$, $s_i := \sum_{j=1}^n a_{ij}$. Then, we apply the elementary operations of the Gaussian elimination of A to the whole matrix \tilde{M} . We start by computing the first pivot, a_{11} , by adding s_1 (≥ 0) and the sum of the absolute values of the first row off-diagonal entries: $a_{11} = s_1 + \sum_{j \neq 1} |a_{1j}|$. Then we produce zeros in the first column of A by adding positive multiples of the first row and, with the exception of the diagonal entries of $A^{(2)}[2, \dots, n]$, every entry of $\tilde{M}^{(2)} = (A^{(2)} | B^{(2)} | s^{(2)})$ is computed with HRA. Nevertheless, we can obtain analogously the first diagonal entries of $A^{(2)}[2, \dots, n], \dots, A^{(n-1)}[n-1, n]$ with HRA when they are needed as pivots at the corresponding steps of the Gaussian elimination of A , and $a_{nn}^{(n)}$ after finishing the elimination procedure. In order to start the second iteration, it only remains to obtain $a_{22}^{(2)}$ with HRA.

Since $A^{(2)}[2, \dots, n]$ is the Schur complement of an M -matrix it is also an M -matrix (see [9]). The vector of row sums is obtained as $Ae = s$, where $e = (1, \dots, 1)^T$. Observe that $s = s^{(1)} \geq 0$ and the way of constructing $\tilde{M}^{(2)}$ from \tilde{M} imply that $s^{(2)} \geq 0$. Besides, e will be also the solution of the linear system $A^{(2)}x = s^{(2)}$, which implies by the sign pattern of $A^{(2)}$ that the components of $s^{(2)}$ coincide again with the row sums of $A^{(2)}$. So, $a_{22}^{(2)}$ can be computed with HRA by adding $s_2^{(2)}$ (≥ 0) and the absolute values of the off-diagonal entries of the second row of $A^{(2)}$. Now we continue the Gaussian elimination and make zeros in the second column below $a_{22}^{(2)}$. We repeat this procedure until when we obtain the upper triangular matrix $U := A^{(n)}$ with HRA. Then $A^{(n)}$ preserves the Z -matrix sign pattern. In this process, the identity matrix becomes the lower triangular matrix $B^{(n)}$, with ones on the diagonal and nonnegative entries below it.

Now, we continue the elimination procedure of $A^{(n)}$ starting with the last row and producing zeros above the main diagonal of $A^{(k)}$ ($n \leq k \leq 2n-1$), as described in (8), and we apply it to the whole matrix $(A^{(n)} | B^{(n)})$. The sign pattern of $(A^{(n)} | B^{(n)})$ allows us to carry out this elimination process without subtractions, and so, with HRA.

The computational cost is given by the cost of Gauss–Jordan elimination (and so of $\mathcal{O}(n^3)$ elementary operations) in addition to the elementary operations to compute the pivots $a_{11}, a_{22}^{(2)}, \dots, a_{nn}^{(n)}$ and to update the vectors $s, s^{(2)}, \dots, s^{(n)}$ (of $\mathcal{O}(n^2)$ elementary operations in both cases). \square

Remark 2.5. By the characterization (I_{28}) of Theorem (2.3) of chapter 6 of [3], a Z -matrix A is a nonsingular M -matrix if and only if there exists a vector z with positive entries such that $s := Az$ has positive entries. Then the same proof of Proposition 2.4 can be used to prove that, if we know the $n^2 + n$ parameters of A given by its $n^2 - n$ off-diagonal entries, the n entries of $z := (z_1, \dots, z_n)^T$ and

the n entries of $Az = s (= (s_1, \dots, s_n)^T)$, then we can compute A^{-1} with HRA. The analogous proof to that of Proposition 2.4 will use now the augmented matrix $\tilde{M} := (A|I|s)$, where $s = Az$, z will play the role of $(1, \dots, 1)^T$ and the expression of a_{11} will be now $a_{11} = \left(s_1 + \sum_{j \neq 1} |a_{1j}|z_j\right)/z_1$. Besides, z will be again the solution of the linear systems $A^{(k)}x = s^{(k)}$ for $k = 2, \dots, n$. The result can be stated as follows: “If $A = (a_{ij})$ is a nonsingular M -matrix and we know its off-diagonal entries as well as $z > 0$ such that $s := Az > 0$, then we can compute A^{-1} with a subtraction-free algorithm (and so with HRA) performing $\mathcal{O}(n^3)$ elementary operations”.

The following result is a consequence of Theorem 2.3 and Proposition 2.4 and guarantees the construction of the inverse of Nekrasov Z -matrices A in the particular case that $h_i(A) \neq 0$ for all i .

Corollary 2.6. *Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a Nekrasov Z -matrix with positive diagonal entries such that $h_i(A) \neq 0$ for $i = 1, \dots, n$ (see (1)). If we know its n^2 N -parameters (2) then we can compute A^{-1} with a subtraction-free algorithm (and so, with HRA) performing $\mathcal{O}(n^3)$ elementary operations.*

Proof. Let S be the matrix given by (3), which can be obtained with HRA and $\mathcal{O}(n^2)$ elementary operations, without performing any subtraction. Then $B := AS$ is a nonsingular diagonally dominant M -matrix and by Theorem 2.3 we can compute its DD-parameters (i.e., off-diagonal entries and row sums) with HRA. With these DD-parameters we can compute B^{-1} with HRA by the procedure described in Proposition 2.4.

Since $B = AS$, we conclude that $A^{-1} = SB^{-1}$ and so each entry of the inverse of A can be computed by multiplying the corresponding entry of B^{-1} by the corresponding diagonal entry of S . This step can be computed with n^2 elementary operations, without performing any subtraction. \square

Remark 2.7. The accurate inverse A^{-1} obtained in Corollary 2.6 (and also for a general Nekrasov Z -matrix with positive diagonal entries, obtained in the following section) can be used to compute with HRA the solution of a linear system $Ax = b$ with $b \geq 0$ by the direct computation $x = A^{-1}b$, since the constructed matrix with HRA $A^{-1} \geq 0$ and so subtractions are not performed. In Section 4, our numerical experiments also show that the solution of the linear system $Ax = b$ for any b , computed by this procedure, is also accurately computed.

3. Accurate inverses in the general case

We show in this section that the condition $h_i(A) \neq 0$ for $i = 1, \dots, n$ can be suppressed in Corollary 2.6. In order to prove this fact, it is crucial to study first the distribution of the zero entries of a Nekrasov matrix that satisfies $h_i(A) = 0$ for some $i \in N$.

Lemma 3.1. *Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a Nekrasov matrix, and let $J = \{i_1, \dots, i_k\} \subseteq N$ ($i_1 \leq i_2 \leq \dots \leq i_k$) be the ordered set of indices such that $h_{i_j}(A) = 0$. Then at least $n - j$ off-diagonal elements of the row i_j are zero for each $j = 1, \dots, k$.*

Proof. Assuming that $J \neq \emptyset$, we start by considering the row i_1 :

$$h_{i_1}(A) = \sum_{k=1}^{i_1-1} |a_{i_1 k}| \frac{h_k(A)}{|a_{kk}|} + \sum_{k=i_1+1}^n |a_{i_1 k}| = 0. \tag{9}$$

Since $h_k(A) \neq 0$ for $k < i_1$, we deduce from (9) that $a_{i_1 k} = 0$ when $k \neq i_1$, that is, all the off-diagonal entries of the i_1 th row are zero. Now we consider the row $i_j \in J$ with $j > 1$:

$$\sum_{k=1}^{i_j-1} |a_{i_j k}| \frac{h_k(A)}{|a_{kk}|} + \sum_{k=i_j+1}^n |a_{i_j k}| = \sum_{k=1, k \notin J}^{i_j-1} |a_{i_j k}| \frac{h_k(A)}{|a_{kk}|} + \sum_{k=i_j+1}^n |a_{i_j k}| = 0.$$

In this case, we have that $a_{i_j k} = 0$ whenever $k \notin \{i_1, \dots, i_j\}$. So there are at least $n - j$ zero entries corresponding to the columns with index $k \notin \{i_1, \dots, i_j\}$. \square

By the previous result, observe that the first row of a Nekrasov matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ that satisfies $h_i(A) = 0$ has exactly $n - 1$ zero entries.

Theorem 3.2. *Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a Nekrasov Z -matrix with positive diagonal entries. If we know its n^2 N -parameters (2), then we can compute A^{-1} with HRA performing a subtraction-free algorithm of $\mathcal{O}(n^3)$ elementary operations.*

Proof. We start by computing $h_1(A), a_{11}, \dots, h_n(A), a_{nn}$ by (1) and (2) from the N -parameters of A without subtractions. This computation requires $\mathcal{O}(n^2)$ elementary operations. Let us define the ordered set $I \subseteq N$ given by the increasing sequence of indices such that $h_i(A) \neq 0$. If $I = N$ we can apply Theorem 2.6. So, from now on we consider the case $I \neq N$.

Let S be the diagonal matrix given by (3). We define the submatrices $\hat{A} := A[I]$ and $B := (AS)[I]$. Observe that B is DD because it is a principal submatrix of AS . It is possible to compute its inverse without performing subtractions and with $\mathcal{O}(n^3)$ elementary operations. In order to prove it, we first obtain an adequate parametrization of B with a subtraction-free algorithm. In this case, the required parameters are its off-diagonal elements, $a_{ij} \frac{h_j(A)}{a_{jj}}$, and its row sums (i.e., its DD-parameters), which can be written by the choice of I , formulae (1), (2) and the sign pattern of a Z -matrix in the following form, as in (5):

$$\sum_{j \in I, j \neq i} a_{ij} \frac{h_j(A)}{a_{jj}} + h_i(A) = \sum_{j=1}^{i-1} a_{ij} \frac{h_j(A)}{a_{jj}} + h_i(A) + \sum_{j=i+1}^n a_{ij} \frac{h_j(A)}{a_{jj}} = \sum_{j=i+1}^n |a_{ij}| \frac{\Delta_j(A)}{a_{jj}}.$$

So, the mentioned DD-parametrization of B can be obtained from (2) by a subtraction-free procedure and $\mathcal{O}(n^2)$ elementary operations. With these parameters we can apply Theorem 2.6 in order to obtain the inverse of the diagonally dominant M -matrix $B = (AS)[I]$ with a subtraction-free algorithm and $\mathcal{O}(|I|^3)$ elementary operations. Then it is straightforward to compute accurately (and with $\mathcal{O}(|I|^2)$ elementary operations) $\hat{A}^{-1} = S[I]B^{-1}$.

The $|I| \times |I|$ matrices \hat{A} and \hat{A}^{-1} allow us to define the following procedure, key to obtain A^{-1} with HRA. It consists of $n - |I|$ major steps resulting in a sequence of matrices as follows:

$$\hat{A} := \hat{A}^{(1)} \rightarrow \hat{A}^{(2)} \rightarrow \dots \rightarrow \hat{A}^{(n-|I|+1)} = A. \quad (10)$$

For each $p \in \{2, \dots, n - |I| + 1\}$, we obtain the matrix $\hat{A}^{(p)}$ by adding to $\hat{A}^{(p-1)}$ the row and column of A corresponding to the biggest index $i \in N$ that was not already involved in it. We form the new matrix keeping the row/column ordering of A , and then we construct the inverse $(\hat{A}^{(p)})^{-1}$ using the information provided by $(\hat{A}^{(p-1)})^{-1}$. The last step will give us A^{-1} . To carry out the first step we start by choosing the biggest $k \in I^c$. Then we form the $(n - |I| + 1) \times (n - |I| + 1)$ matrix $\hat{A}^{(2)}$ adding the corresponding entries of the k th row and column of A to \hat{A} in the corresponding place. In order to obtain the inverse of this new matrix from $C = \hat{A}^{-1}$ we use Lemma 3.1, which states that the new row added to \hat{A} has at least $|I|$ zeros that will appear as off-diagonal elements. The new row has only one element in $\hat{A}^{(2)}$ different from zero, a_{kk} . In this case the entries of $C^{(2)} := (\hat{A}^{(2)})^{-1}$ are the following:

$$c_{ij}^{(2)} = \begin{cases} c_{ij}, & i, j \in I, \\ \frac{1}{a_{kk}}, & i = j = k, \\ 0, & i = k, j \in I, \\ c_{ik}^{(2)}, & i \in I, j = k. \end{cases}$$

We need to check this fact and find the expression of the entries $c_{ik}^{(2)}$. We consider the product $\hat{A}^{(2)}C^{(2)}$, which has to be the identity matrix of order $|I| + 1$. Let us start with the case when both $i, j \in I$. Since the inverse of \hat{A} is C , the performed operation to obtain the element (i, j) of the product is:

$$\sum_{s \in I} a_{is}c_{sj} + a_{ik} \cdot 0 = \begin{cases} 0, & i \neq j, \\ 1, & i = j. \end{cases}$$

Now, if $i = k, j \in I$, we have

$$\sum_{s \in I} a_{ks}c_{sj} + a_{kk}c_{kj} = \sum_{s \in I} 0 \cdot c_{sj} + a_{kk} \cdot 0 = 0$$

If $i = j = k$, we obtain

$$\sum_{s \in I} a_{ks}c_{sk} + a_{kk}c_{kk} = \sum_{s \in I} 0 \cdot c_{sk} + \frac{a_{kk}}{a_{kk}} = 1$$

It remains the case $i \in I, j = k$, which determines the missing entries of $C^{(2)}$:

$$\sum_{s \in I} a_{is}c_{sk}^{(2)} + \frac{a_{ik}}{a_{kk}} = 0, \quad i \in I.$$

Let us define $c := (c_{ik}^{(2)})_{i \in I}$, the vector composed by the missing entries. Then, we can express the system of equations in terms of the matrix \hat{A} :

$$\hat{A}c = - (a_{ik})_{i \in I} \begin{pmatrix} 1 \\ a_{kk} \end{pmatrix}.$$

We have already computed \hat{A}^{-1} with HRA, and the right hand side is nonnegative, so we obtain c with HRA (see Theorem 2.6) by performing the product:

$$c = C (a_{ik})_{i \in I} \begin{pmatrix} -1 \\ a_{kk} \end{pmatrix} = \hat{A}^{-1} (a_{ik})_{i \in I} \begin{pmatrix} -1 \\ a_{kk} \end{pmatrix}.$$

So we obtain $C^{(2)}$. We can continue analogously. In general, after performing $p - 1$ major steps we may obtain A^{-1} and finish the procedure, or we may have to continue it adding the row and column of index k , where $k \in I^c$ is the biggest index such that the k th row was not involved in $\hat{A}^{(p-1)}$. The added row had at least $|I| + p - 1$ zeros in the original matrix, A . Now these zeros are the off-diagonal elements of the added row. We define $I^{(p)}$, the ordered set of indices of the rows from A used in $\hat{A}^{(p-1)}$. Then we perform the product $c = C^{(p-1)} (a_{ik})_{i \in I^{(p-1)}} \begin{pmatrix} -1 \\ a_{kk} \end{pmatrix}$ in order to obtain the missing entries of the matrix $C^{(p)} = (\hat{A}^{(p)})^{-1}$. After computing c , we build $C^{(p)}$:

$$c_{ij}^{(p)} = \begin{cases} c_{ij}^{(p-1)}, & i, j \in I^{(p)}, \\ \frac{1}{a_{kk}}, & i = j = k, \\ 0, & i = k, j \in I^{(p)}, \\ c, & i \in I^{(p)}, j = k. \end{cases}$$

Clearly, we can perform these calculations with HRA and with $\mathcal{O}(n^3)$ elementary operations. \square

4. Algorithms and numerical tests

In the previous section we have presented a procedure that allows us to compute the inverse of a Nekrasov Z -matrix accurately if we know its N -parameters (2) with HRA. In this section, we are going to present the algorithms to compute such inverses following Theorem 3.2 and we are going to test them with some numerical examples.

The first algorithm introduced, Algorithm 1, starts with the N-parameters of the Nekrasov Z -matrix and performs the required preparation to compute its inverse depending on the distribution of the zero entries of the matrix.

If $h_i(A) \neq 0$ for $i = 1, \dots, n$ the procedure corresponds to Theorem 2.3, and it calculates the DD-parameters of AS . Otherwise, the algorithm works with the adequate submatrix as described in Theorem 3.2. The output consists of the matrix A , where the parameters of $(AS)[I]$ are stored in the submatrix $A[I]$ (the case $I = N$ corresponds to Theorem 2.3), the ordered set of indices I and, if the cardinal $|I| > 1$, the diagonal matrix S .

Algorithm 1 nektoDD.

Input: $A = (a_{ij})(i \neq j)$, Δ ▷ The N-parameters (2)
for $i = 1 : n$ **do**
 $h_i = \sum_{j=1}^{i-1} a_{ij}k_j + \sum_{j=i+1}^n a_{ij}$
 $a_{ii} = \Delta_i + h_i$
 $k_i = h_i/a_{ii}$
end for
Build I , the set of indices such that $h_i(A) \neq 0$.
if $|I| > 1$ **then**
 for $i = I$ **do**
 $a_{ii} = \sum_{j=i+1}^n a_{ij}\Delta_j/a_{jj}$
 for $j = I \setminus \{i\}$ **do**
 $a_{ij} = a_{ij}k_j$
 end for
 end for
 Build S , the $|I| \times |I|$ diagonal matrix whose diagonal entries are k_j , $j \in I$.
else if $|I| = 1$ **then**
 $a_{II} = 1/a_{II}$
else
 $a_{nn} = 1/a_{nn}$
 $I = [n]$
end if

Once we obtain the DD-parameters of the DD M -matrix AS (or $(AS)[I]$), our goal is to compute its inverse with HRA. We can compute it using the subtraction-free implementation of Gauss–Jordan elimination without pivoting described in the proof of Proposition 2.4. For brevity, we do not include this algorithm, which can be easily derived. The inverse can be stored in A using again the submatrix $A[I]$.

If we have the case that $h_i(A) \neq 0$ for $i = 1, \dots, n$ (analogously, $|I| = n$), it only remains to perform the product $S(AS)^{-1}$, since we obtained $(AS)^{-1}$ applying Gauss–Jordan elimination. Otherwise, we need to build the inverse of AS starting with $((AS)[I])^{-1}$. Algorithm 2 performs this computation. Its input is the matrix A obtained after running Algorithm 1 and the set of indices I (we just need to perform the direct product $S[I]((AS)[I])^{-1}$ before, as done in Algorithm 3).

With Algorithm 1, Gauss–Jordan elimination adapted according to Proposition 2.4 and Algorithm 2, we can give a general method to compute the inverse of a Nekrasov Z -matrix with positive diagonal with HRA starting with its N-parameters. Algorithm 3 performs all the process.

Algorithm 2 buildnekinv.

Input: A, I ▷ $A[I]$ contains $A[I]^{-1}$
 Build the set of ordered indices $J := I^c = \{j_1, \dots, j_k\}$ such that $j_1 > j_2 > \dots > j_k$.
for $i = J$ **do**
 $a_{ii} = 1/a_{ii}$
 $A[I|i] = -A[I](A[I|i]. * a_{ii})$ ▷ $*$ means component-wise multiplication
 $I = I \cup \{i\}$ (ordered)
end for

Algorithm 3 Computation of the inverse.

Input: $A = (a_{ij})(i \neq j), \Delta$ ▷ N-parameters
 $[A, I, S] = \text{nektodd}(A = (a_{ij})(i \neq j), \Delta)$
if $|I| > 1$ **then**
 Compute $B = A[I]^{-1}$ using the adapted Gauss–Jordan elimination
 $A[I] = S * B$
end if
 $A^{-1} = \text{buildnekinv}(A)$

Table 1
 Maximum relative errors when computing A^{-1} .

Condition number	MATLAB	HRA
6.7161e+03	2.5585e-13	4.4536e-15
7.4296e+04	5.4000e-12	8.9743e-15
2.1634e+06	3.7380e-11	4.4752e-15
1.2159e+05	4.5739e-12	2.9456e-15
6.4136e+03	5.1254e-13	1.6247e-15
1.6378e+05	1.9921e-11	2.2964e-15
1.9344e+06	1.3436e-13	3.7407e-14
2.0715e+05	3.0038e-11	2.2757e-15
2.9297e+05	7.1062e-12	1.5991e-15
1.7608e+03	4.9903e-14	1.6191e-15

The numerical experiments have been carried out computing the inverses with Algorithm 3. The errors were estimated comparing the computed approximations with the exact arithmetic solutions obtained with the Symbolic Math Toolbox of MATLAB. In order to illustrate the accuracy of the method presented in this paper, the same problems are also solved using the usual MATLAB commands. In Table 1 we show the maximum relative errors obtained computing the inverse of ten 20×20 Nekrasov Z -matrices generated randomly. The column labeled MATLAB shows the error when the inverse is computed using the MATLAB command *inv*, and the column HRA shows the error when the inverse is obtained from the N-parameters using the procedure with HRA. We observe better results with our method, but the obtained difference is not large since the generated examples are not ill-conditioned.

Besides, since all off-diagonal entries are generated randomly, these first examples do not include any matrix satisfying $h_i(A) = 0$ for some $i = 1, \dots, n$. One way to obtain examples with a greater condition number consists precisely in generating matrices using this additional condition. If we impose $h_j(A) = 0$ whenever $j \in J \subseteq N$, the entries a_{ij} with $j \in J$ and $i > j$ may be arbitrarily large. By generating these entries significantly larger than the others, we obtain Nekrasov matrices that are far from being

Table 2

Maximum relative errors when computing A^{-1} , with the condition $h_i(A) = 0$ for some i .

Condition number	MATLAB	HRA
4.8463e+11	1.4620e-04	8.7273e-16
1.8512e+11	1.5955e-12	9.0349e-16
1.1334e+11	9.1933e+14	6.3183e-16
3.6138e+11	2.4059e-05	8.5640e-16
8.4356e+10	3.1290e+05	9.1776e-16
1.0278e+11	3.7958e-02	8.7454e-16
1.0960e+11	1.0160e-12	7.6520e-16
1.1049e+12	2.2165e-04	3.8750e-15
2.0787e+11	2.2370e-05	1.5643e-15
1.8109e+11	5.8134e-06	1.1298e-15

Table 3

Maximum relative errors when solving $Ax = b$ with $b = e$.

Condition number	MATLAB	HRA
1.5337e+11	5.5757e-05	9.7469e-16
5.2794e+10	9.0848e-07	3.4470e-16
8.7214e+10	1.5188e-05	6.6034e-16
1.2596e+11	2.3053e-14	1.3981e-15
6.3378e+10	2.6565e-14	5.2600e-16
7.1578e+10	4.8790e-05	3.6567e-16
4.6526e+10	1.3704e-14	5.5542e-16
7.7622e+10	5.8318e-06	5.1943e-16
4.1758e+10	4.5051e-15	5.8087e-16
7.2351e+10	1.2952e-13	5.6605e-16

diagonally dominant. For such matrices, the MATLAB command *inv* gives inaccurate inverses and the procedure with HRA introduced in Theorem 3.2 performs as expected. The results can be seen in Table 2, which contains ten examples of 20×20 Nekrasov Z -matrices.

As we mentioned earlier in Remark 2.7, computing the inverse of a Nekrasov Z -matrix A with HRA also allows us to solve with HRA the linear system $Ax = b$ with $b \geq 0$ by performing the computation $x = A^{-1}b$. In Table 3, we show the maximum relative error obtained computing the solution in ten cases considering $b = e = (1, \dots, 1)^T$. The involved matrices are 20×20 Nekrasov Z -matrices with positive diagonal, generated as in the previous case. We show the results obtained computing the solution with the MATLAB command `\` and the method with HRA, which computes the inverse from the N -parameters and performs the direct computation $x = A^{-1}b$. We observe the great accuracy of our method, in contrast to MATLAB.

In order to assure the HRA, we required $b \geq 0$. However, we may obtain accurate solutions even without this requirement. For this purpose, we generated ten 20×20 Nekrasov Z -matrices with positive diagonal entries and we solved the system $Ax = b$ with $b = (b_i)_{1 \leq i \leq n}$, $b_i = (-1)^{i+1}$. Table 4 shows the results obtained with the MATLAB command `\` and with the procedure with HRA.

Table 4Maximum relative errors when solving $Ax = b$ with $b_i = (-1)^{i+1}$.

Condition number	MATLAB	HRA
1.8080e+11	2.0521e-13	3.9922e-16
1.6297e+12	8.7030e-14	3.2741e-14
1.6561e+12	6.1643e-04	1.5069e-15
6.7289e+10	1.9126e-14	2.6656e-15
1.1951e+11	3.0361e-14	8.7663e-16
2.7320e+11	7.5251e-15	1.1204e-15
4.6654e+10	9.9933e-06	7.0215e-16
1.1328e+11	2.3099e-06	3.1267e-16
5.7753e+11	1.7024e-13	3.5437e-15
7.4226e+10	1.9813e-06	8.1572e-16

Conflict of interest statement

The authors declare that there is no conflict of interest.

Acknowledgement

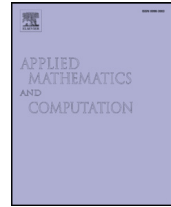
The authors thank an anonymous referee for the valuable suggestions to improve the paper.

References

- [1] A.S. Alfa, J. Xue, Q. Ye, Entrywise perturbation theory for diagonally dominant M -matrices with applications, *Numer. Math.* 90 (2002) 401–414.
- [2] A. Barreras, J.M. Peña, Accurate and efficient LDU decompositions of diagonally dominant M -matrices, *Electron. J. Linear Algebra* 24 (2014) 153–167.
- [3] A. Berman, R.J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Classics in Applied Mathematics, vol. 9, SIAM, Philadelphia, 1994.
- [4] R. Bru, C. Corral, I. Gimenez, J. Mas, Classes of general H -matrices, *Linear Algebra Appl.* 429 (2008) 2358–2366.
- [5] L. Cvetković, P-F. Dai, K. Doroslovaški, Y.T. Li, Infinity norm bounds for the inverse of Nekrasov matrices, *Appl. Math. Comput.* 219 (2013) 5020–5024.
- [6] J. Demmel, I. Dumitriu, O. Holtz, P. Koev, Accurate and efficient expression evaluation and linear algebra, *Acta Numer.* 17 (2008) 87–145.
- [7] J. Demmel, M. Gu, S. Eisenstat, I. Slapnicar, K. Veselic, Z. Drmac, Computing the singular value decomposition with high relative accuracy, *Linear Algebra Appl.* 299 (1999) 21–80.
- [8] J. Demmel, P. Koev, Accurate SVDs of weakly diagonally dominant M -matrices, *Numer. Math.* 98 (2004) 99–104.
- [9] K. Fan, Note on M -matrices, *Q. J. Math. Oxford Ser. (2)* 11 (1961) 43–49.
- [10] M. García-Esnaola, J.M. Peña, Error bounds for linear complementarity problems of Nekrasov matrices, *Numer. Algorithms* 67 (2013) 655–667.
- [11] C. Li, Q. Liu, L. Gao, Y. Li, Subdirect sums of Nekrasov matrices, *Linear Multilinear Algebra* 64 (2016) 208–218.
- [12] J. Liu, J. Zhang, L. Zhou, G. Tu, The Nekrasov diagonally dominant degree on the Schur complement of Nekrasov matrices and its applications, *Appl. Math. Comput.* 320 (2018) 251–263.
- [13] J.M. Peña, LDU decompositions with L and U well conditioned, *Electron. Trans. Numer. Anal.* 18 (2004) 198–208.
- [14] T. Szulc, Some remarks on a theorem of Gudkov, *Linear Algebra Appl.* 225 (1995) 221–235.

Article 4

- [81] H. Orera and J. M. Peña. Infinity norm bounds for the inverse of Nekrasov matrices using scaling matrices. *Appl. Math. Comput.* 358 (2019), 119-127.



Infinity norm bounds for the inverse of Nekrasov matrices using scaling matrices

H. Orera*, J.M. Peña

Departamento de Matemática Aplicada/IUMA, Universidad de Zaragoza, Zaragoza, Spain

ARTICLE INFO

MSC:
65F35
15A60
65F05
90C33

Keywords:

Infinity matrix norm
Inverse matrix
Nekrasov matrices
 H -matrices
Strictly diagonally dominant matrices
Scaling matrix

ABSTRACT

For many applications, it is convenient to have good upper bounds for the norm of the inverse of a given matrix. In this paper, we obtain such bounds when A is a Nekrasov matrix, by means of a scaling matrix transforming A into a strictly diagonally dominant matrix. Numerical examples and comparisons with other bounds are included. The scaling matrices are also used to derive new error bounds for the linear complementarity problems when the involved matrix is a Nekrasov matrix. These error bounds can improve considerably other previous bounds.

© 2019 Elsevier Inc. All rights reserved.

1. Introduction

Providing upper bounds for the infinity norm of the inverse of a matrix has many potential applications in Computational Mathematics. For instance, for bounding the condition number of the matrix, for bounding errors in linear complementarity problems (cf. [1]) or, in the class of H -matrices, for proving the convergence of matrix splitting and matrix multisplitting iteration methods for solving sparse linear systems of equations (cf. [2]).

The class of Nekrasov matrices (see [3] or Section 2) contains the class of strictly diagonally dominant matrices. Recent applications of Nekrasov matrices can be seen in [4–11]. Nekrasov matrices are H -matrices. Let us recall some related classes of matrices. A real matrix A is a nonsingular M -matrix if its inverse is nonnegative and all its off-diagonal entries are nonpositive. M -matrices form a very important class of matrices with applications to Numerical Analysis, Optimization, Economy, and Dynamic systems (cf. [12]). Given a complex matrix $A = (a_{ij})_{1 \leq i, j \leq n}$, its comparison matrix $\mathcal{M}(A) = (\tilde{a}_{ij})_{1 \leq i, j \leq n}$ has entries $\tilde{a}_{ii} := |a_{ii}|$ and $\tilde{a}_{ij} := -|a_{ij}|$ for all $j \neq i$ and $i, j = 1, \dots, n$. We say that a complex matrix is an H -matrix if its comparison matrix is a nonsingular M -matrix. About a more general definition of H -matrix, see [13]. A matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ is SDD (strictly diagonally dominant by rows) if $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$ for all $i = 1, \dots, n$.

It is well-known that an SDD matrix is nonsingular and that a square matrix A is an H -matrix if there exists a diagonal matrix S with positive diagonal entries such that AS is SDD. The role of the scaling matrix is crucial, for instance, for the problem mentioned above of the convergence of iteration methods and also for the problem of eigenvalue localization (see [11]). This paper deals with the research of such scaling matrices S for the particular case when A is a Nekrasov matrix. The scaling matrix S is applied to obtain upper bounds for the infinity norm of the inverse of a Nekrasov matrix.

* Corresponding author.

E-mail addresses: hectororera@unizar.es (H. Orera), jmpena@unizar.es (J.M. Peña).

The paper is organized as follows. Section 2 constructs scaling matrices S for Nekrasov matrices A , such that AS is SDD. Section 3 applies the scaling matrices of Section 2 to derive upper bounds of $\|A^{-1}\|_\infty$, including an algorithm to obtain the corresponding bound. Section 4 presents an improvement of the bound obtained in Section 3 and includes numerical examples, illustrating our bounds and comparing them with other previous bounds. We consider several test matrices previously considered in the literature, and we also consider some variants of these matrices. We also include a family of 3×3 matrices showing that previous bounds can be arbitrarily large, in contrast to our bounds, which are always controlled. Finally, we derive bounds for other norms. Section 5 illustrates the use of our scaling matrices to derive new error bounds for the linear complementarity problems when the involved matrix is a Nekrasov matrix. We avoid the restrictions of the bound in [1] and we present a family of matrices for which our bound is a small constant, in contrast to the bounds of [14–16], which can be arbitrarily large.

We finish this introduction with some basic notations. Let $N := \{1, \dots, n\}$. Let $Q_{k,n}$ be the set of increasing sequence of k positive integers in N . Given $\alpha, \beta \in Q_{k,n}$, we denote by $A[\alpha|\beta]$ the $k \times k$ submatrix of A containing rows numbered by α and columns numbered by β . If $\alpha = \beta$, then we have the principal submatrix $A[\alpha] := A[\alpha|\alpha]$. Finally, the diagonal matrix with diagonal entries $d_i, 1 \leq i \leq n$, will be denoted by $\text{diag}(d_i)_{i=1}^n$.

2. Scaling matrices

Let us start by defining the concept of a Nekrasov matrix (see [2–4]). For this purpose, let us define recursively, for a complex matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ with $a_{ii} \neq 0$, for all $i = 1, \dots, n$,

$$h_1(A) := \sum_{j \neq 1} |a_{1j}|, \quad h_i(A) := \sum_{j=1}^{i-1} |a_{ij}| \frac{h_j(A)}{|a_{jj}|} + \sum_{j=i+1}^n |a_{ij}|, \quad i = 2, \dots, n. \tag{1}$$

We say that A is a *Nekrasov matrix* if $|a_{ii}| > h_i(A)$ for all $i \in N$. It is well-known that a Nekrasov matrix is a nonsingular H -matrix [3]. So, there exists a positive diagonal matrix S such that AS is SDD. In particular, Nekrasov matrices can be characterized in terms of these scaling matrices (see Theorem 2.2 of [7]). Once we have found a scaling matrix, we can use it to derive infinity norm bounds for the inverse of Nekrasov matrices, which may be useful for many problems, as recalled in the Introduction. In fact, the problem of bounding the infinity norm of the inverse of a Nekrasov matrix has attracted great attention recently (see [2,5,6,8,9]).

In this section we are introducing two methods that allow us to build a scaling matrix for any given Nekrasov matrix.

Theorem 2.1. Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a Nekrasov matrix. Then the matrix $S = \begin{pmatrix} \frac{h_1(A)+\epsilon_1}{|a_{11}|} & & \\ & \ddots & \\ & & \frac{h_n(A)+\epsilon_n}{|a_{nn}|} \end{pmatrix}$,

with $\begin{cases} \epsilon_1 > 0, \\ 0 < \epsilon_i \leq |a_{ii}| - h_i(A), \quad \epsilon_i > \sum_{j=1}^{i-1} \frac{|a_{ij}|\epsilon_j}{|a_{jj}|} \quad \text{for } i = 2, \dots, n, \end{cases}$

is a positive diagonal matrix such that AS is SDD.

Proof. Let us start by proving that there exist $\epsilon_1, \dots, \epsilon_n$ satisfying the conditions stated above. We consider, as a first choice, $\epsilon_i := |a_{ii}| - h_i(A)$ for $i = 1, \dots, n$. If $\epsilon_i > \sum_{j=1}^{i-1} \frac{|a_{ij}|\epsilon_j}{|a_{jj}|}$ for all $i = 2, \dots, n$ we have finished. Otherwise, let $i > 1$ be the first index such that the inequality does not hold. Then we substitute ϵ_j by $\frac{\epsilon_j}{\hat{M}_i}$, with $j = 1, \dots, i - 1$, where \hat{M}_i is a positive number such that the inequality is satisfied. The inequalities checked at earlier steps remain true. We continue this process until the inequality holds for all $i = 2, \dots, n$.

The diagonal matrix S is positive because $h_i(A) \geq 0$ and $\epsilon_i > 0$. The entry (i, j) of AS is $a_{ij} \frac{h_j(A)+\epsilon_j}{|a_{jj}|}$. In order to prove that AS is SDD we start by checking that the condition is true for the n th row:

$$\sum_{j=1}^{n-1} |a_{nj}| \frac{h_j(A) + \epsilon_j}{|a_{jj}|} = \underbrace{\sum_{j=1}^{n-1} |a_{nj}| \frac{h_j(A)}{|a_{jj}|}}_{h_n(A)} + \sum_{j=1}^{n-1} |a_{nj}| \frac{\epsilon_j}{|a_{jj}|} < h_n(A) + \epsilon_n = |(AS)[n]|$$

The condition holds for the row $n - 1$:

$$\sum_{j=1}^{n-2} |a_{n-1,j}| \frac{h_j(A) + \epsilon_j}{|a_{jj}|} + |a_{n-1,n}| \underbrace{\frac{h_n(A) + \epsilon_n}{|a_{nn}|}}_{\leq 1} \leq h_{n-1}(A) + \sum_{j=1}^{n-2} |a_{n-1,j}| \frac{\epsilon_j}{|a_{jj}|} < h_{n-1}(A) + \epsilon_{n-1} = |(AS)[n-1]|$$

The first inequality is due to the hypothesis $\epsilon_n \leq |a_{nn}| - h_n(A)$, which implies $\frac{h_n(A) + \epsilon_n}{|a_{nn}|} \leq 1$. In general, considering the i th row for $2 \leq i < n - 1$:

$$\sum_{j=1}^{i-1} |a_{ij}| \frac{h_j(A) + \epsilon_j}{|a_{jj}|} + \sum_{j=i+1}^n |a_{ij}| \frac{h_j(A) + \epsilon_j}{|a_{jj}|} \leq h_i(A) + \sum_{j=1}^{i-1} |a_{ij}| \frac{\epsilon_j}{|a_{jj}|} < h_i(A) + \epsilon_i = |(AS)[i]|$$

and, when $i = 1$:

$$\sum_{j=2}^n |a_{1j}| \frac{h_j(A) + \epsilon_j}{|a_{jj}|} \leq h_1(A) < h_1(A) + \epsilon_1 = |(AS)[1]|.$$

The inequality for the i th row is proven using that $\epsilon_j \leq |a_{jj}| - h_j(A)$ for $j = i + 1, \dots, n$ and $\epsilon_i > \sum_{j=1}^{i-1} \frac{|a_{ij}| \epsilon_j}{|a_{jj}|}$. If $i = 1$, the last inequality is reduced to $\epsilon_1 > 0$. \square

In [Theorem 2.1](#) we introduced a diagonal matrix S that transforms any Nekrasov matrix into an SDD matrix. Its construction implied the search of the parameters ϵ_i for $i \in N$. Taking into account the existence of nonzero entries in the upper triangular part of a Nekrasov matrix, we can build a new scaling matrix S , simpler in many cases, whose product with a Nekrasov matrix is also SDD.

Theorem 2.2. *Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a Nekrasov matrix and let $k \in N$ be the first index such that there does not exist $j > k$ with $a_{kj} \neq 0$. Then the matrix $S = \text{diag}(s_i)_{i=1}^n$ with $s_i := \frac{h_i(A) + \epsilon_i}{|a_{ii}|}$ and with $\begin{cases} \epsilon_i = 0, & i = 1, \dots, k-1, \\ 0 < \epsilon_i < |a_{ii}| - h_i(A), & \epsilon_i > \sum_{j=k}^{i-1} \frac{|a_{ij}| \epsilon_j}{|a_{jj}|} \text{ for } i = k, \dots, n, \end{cases}$ is a positive diagonal matrix such that AS is SDD.*

Proof. Let us start by showing that there exist $\epsilon_1, \dots, \epsilon_n$ satisfying the conditions stated above. Since A is a Nekrasov matrix, we have that $|a_{ii}| > h_i(A)$ for $i = 1, \dots, n$. The existence of $\epsilon_1 = \dots = \epsilon_{k-1} = 0$ is trivial and, following the constructive proof of the existence of these parameters given in [Theorem 2.1](#), we can deduce the existence of $\epsilon_k, \dots, \epsilon_n$. It remains to prove that AS is an SDD matrix, which can be done analogously to the proof of [Theorem 2.1](#).

Let us first consider the i th row, when $i < k$. Since $i < k$, there exists an entry $a_{ij} \neq 0$ with $i < j$. Taking also into account that $\epsilon_j = 0$ for all $j < i$ and that $h_j(A) + \epsilon_j < |a_{jj}|$ for all $j = i + 1, \dots, n$, we deduce that:

$$\begin{aligned} & \sum_{j=1}^{i-1} |a_{ij}| \frac{h_j(A) + \epsilon_j}{|a_{jj}|} + \sum_{j=i+1}^n |a_{ij}| \frac{h_j(A) + \epsilon_j}{|a_{jj}|} \\ &= \sum_{j=1}^{i-1} |a_{ij}| \frac{h_j(A)}{|a_{jj}|} + \sum_{j=i+1}^n |a_{ij}| \frac{h_j(A) + \epsilon_j}{|a_{jj}|} \\ &< \sum_{j=1}^{i-1} |a_{ij}| \frac{h_j(A)}{|a_{jj}|} + \sum_{j=i+1}^n |a_{ij}| = h_i(A) = |(AS)[i]|. \end{aligned}$$

For the k th row we have that $a_{kj} = 0$ for every $j > k$ and so:

$$\begin{aligned} & \sum_{j=1}^{k-1} |a_{kj}| \frac{h_j(A) + \epsilon_j}{|a_{jj}|} + \sum_{j=k+1}^n |a_{kj}| \frac{h_j(A) + \epsilon_j}{|a_{jj}|} = \sum_{j=1}^{k-1} |a_{kj}| \frac{h_j(A)}{|a_{jj}|} \\ &= h_k(A) < h_k(A) + \epsilon_k = |(AS)[k]|. \end{aligned}$$

It just remains to check the i th rows, when $i > k$. Since $h_j(A) + \epsilon_j < |a_{jj}|$ for all $j > i (> k)$, we have, by the choice of ϵ_i :

$$\sum_{j=1}^{i-1} |a_{ij}| \frac{h_j(A) + \epsilon_j}{|a_{jj}|} + \sum_{j=i+1}^n |a_{ij}| \frac{h_j(A) + \epsilon_j}{|a_{jj}|} \leq h_i(A) + \sum_{j=k}^{i-1} |a_{ij}| \frac{\epsilon_j}{|a_{jj}|} < h_i(A) + \epsilon_i = |(AS)[i]|.$$

\square

The particular case $k = n$ corresponds to a diagonal matrix with $\epsilon_1 = \dots = \epsilon_{n-1} = 0$ and $\epsilon_n \in (0, |a_{nn}| - h_n(A))$. This scaling matrix was already introduced in [\[1\]](#) and it was used to derive an error bound for linear complementarity problems of Nekrasov matrices. In the following section, we shall apply the scaling matrices derived in this section to the problem of bounding the norm of the inverse of a Nekrasov matrix.

3. Bounding $\|A^{-1}\|_\infty$

With an adequate scaling matrix S (given by [Theorems 2.1](#) or [2.2](#)) we can obtain the desired bound for the inverse of a Nekrasov matrix A considering the product AS . For this purpose, we are going to use the following result introduced by Varah in [\[17\]](#):

Table 1
Computational cost of (2).

Operations	General	$k = n$	$k = 1$
additions/subtractions	$\frac{3n^2+n+2}{2} + \frac{(n-k-1)(n-k)}{2}$	$\frac{3n^2+n+2}{2}$	$2n^2 - n + 2$
multiplications	$\frac{7n^2+9n+4}{2} + \frac{5k^2-10kn-11k}{2}$	$n(n-1)$	$\frac{7n^2-n-2}{2}$
quotients	$2n-1+2(n-k)$	$2n-1$	$4n-3$

Table 2
Leading term of the computational cost of (2).

	$k = n$	$k = 1$
T	$\frac{5}{2}n^2$	$\frac{11}{2}n^2$

Theorem 3.1. If A is SDD and $\alpha := \min_k (|a_{kk}| - \sum_{j \neq k} |a_{kj}|)$, then $\|A^{-1}\|_\infty < 1/\alpha$.

Theorem 3.1 gives a bound for the infinity norm of the inverse of an SDD matrix. This theorem, jointly with the scaling matrices introduced in Section 2, allows us to deduce Theorem 3.2.

Theorem 3.2. Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a Nekrasov matrix. Then

$$\|A^{-1}\|_\infty \leq \frac{\max_{i \in N} \left(\frac{h_i(A) + \epsilon_i}{|a_{ii}|} \right)}{\min_{i \in N} (\epsilon_i - w_i + p_i)}, \tag{2}$$

where $(\epsilon_1, \dots, \epsilon_n)$ are given by Theorems 2.1 or 2.2, $w_i := \sum_{j=1}^{i-1} |a_{ij}| \frac{\epsilon_j}{|a_{jj}|}$, and $p_i := \sum_{j=i+1}^n |a_{ij}| \frac{|a_{jj}| - h_j(A) - \epsilon_j}{|a_{jj}|}$ for all $i \in N$.

Proof. We choose a diagonal matrix S following either Theorems 2.1 or 2.2 and we deduce the following inequality:

$$\|A^{-1}\|_\infty = \|S(S^{-1}A^{-1})\|_\infty = \|S(AS)^{-1}\|_\infty \leq \|S\|_\infty \|(AS)^{-1}\|_\infty. \tag{3}$$

The matrix S is diagonal, so its infinity norm is given by $\max_{i \in N} \left(\frac{h_i(A) + \epsilon_i}{|a_{ii}|} \right)$. Since AS is SDD we can apply Theorem 3.1 to $\|(AS)^{-1}\|_\infty$. For this purpose, we need to compute for each $i = 1, \dots, n$:

$$\begin{aligned} h_i(A) + \epsilon_i - \sum_{j \neq i} |a_{ij}| \frac{h_j(A) + \epsilon_j}{|a_{jj}|} &= \epsilon_i - \sum_{j=1}^{i-1} |a_{ij}| \frac{\epsilon_j}{|a_{jj}|} + \sum_{j=i+1}^n |a_{ij}| \frac{|a_{jj}| - h_j(A) - \epsilon_j}{|a_{jj}|} \\ &= \epsilon_i - w_i + p_i, \end{aligned}$$

where we have substituted $h_i(A)$ by the expression given by (1). □

Since the diagonal matrix S satisfies $\|S\|_\infty \leq 1$, we can substitute the numerator of the bound (2) by one and obtain the following result:

Corollary 3.3. Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a Nekrasov matrix. Then

$$\|A^{-1}\|_\infty \leq \frac{1}{\min_{i \in N} (\epsilon_i - w_i + p_i)},$$

where $(\epsilon_1, \dots, \epsilon_n)$ are given by Theorems 2.1 or 2.2, $w_i := \sum_{j=1}^{i-1} |a_{ij}| \frac{\epsilon_j}{|a_{jj}|}$, and $p_i := \sum_{j=i+1}^n |a_{ij}| \frac{|a_{jj}| - h_j(A) - \epsilon_j}{|a_{jj}|}$ for $i \in N$.

In Table 1 we present the computational cost of the bound (2) using the matrix S given by Theorem 2.2. The cost depends on the index k . Two extreme cases are studied separately. The first one, $k = n$, corresponds to the simplest case, where $\epsilon_i = 0$ for $i = 1, \dots, n - 1$. The second one corresponds to $k = 1$ and it uses a diagonal matrix S with $\epsilon_i \neq 0$ for all $i \in N$. In fact, in this case the diagonal matrix S also satisfies the definition given by Theorem 2.1.

The particular cases $k = n$ and $k = 1$ have the lowest and biggest computational cost, respectively. Table 2 shows the leading term T of the computational cost in these cases.

Now we are going to introduce Algorithm 1, which allows us to compute the bound (2) choosing ϵ_i with $i \in N$ following Theorem 2.2. It corresponds to Theorem 2.1 when $k = 1$. In order to compute this bound, the algorithm needs to give some initial values to $\epsilon_1, \dots, \epsilon_n$. These parameters are initialized with either 0 or $t(|a_{ii}| - h_i(A))$, where $t \in (0, 1)$. It could be useful to choose a different scalar t for each ϵ_i . However, it is not clear how to choose their values, and for many matrices, such as those included in Section 4, we have that $\epsilon_i = 0$ for $i = 1, \dots, n - 1$. In this case, we also consider in Section 4 the possibility of choosing ϵ_n as the middle point of its interval, that is, $\epsilon_n = \frac{\Delta_n}{2}$, where $\Delta_n = |a_{nn}| - h_n(A)$.

Algorithm 1 nektoSDD - Computing bound (2).

```

Input:  $A = (a_{ij})_{1 \leq i, j \leq n}$ ,  $t$ 
for  $i = 1 : n$  do
     $h_i = \sum_{j=1}^{i-1} |a_{ij}| k_j$ 
     $r = \sum_{j=i+1}^n |a_{ij}|$ 
    if  $r == 0, J == 0$  then
         $J = i;$ 
    end if
     $h_i = h_i + r$ 
     $\Delta_i = |a_{ii}| - h_i$ 
     $k_i = h_i / |a_{ii}|$ 
end for
 $\epsilon_K = t \Delta_K$ 
 $w_1 = \dots = w_K = 0$ 
for  $i = K + 1 : n$  do
     $\epsilon_i = t \Delta_i$ 
     $p_j = \epsilon_j / |a_{jj}|$ 
     $w_i = \sum_{j=K}^{i-1} |a_{ij}| p_j$ 
    if  $w_i - \epsilon_i > 0$  then
         $M = 1/2 w_i$ 
        for  $j = K : i - 1$  do
             $\epsilon_j = \epsilon_j \epsilon_i M$ 
             $w_j = w_j \epsilon_i M$ 
        end for
         $w_i = \epsilon_i / 2$ 
    end if
end for
for  $i = n : -1 : 2$  do
     $S_i = \epsilon_i - w_i + \sum_{j=i+1}^n |a_{ij}| f_j$ 
     $f_i = (\Delta_i - \epsilon_i) / |a_{ii}|$ 
end for
 $S_1 = \epsilon_1 - w_1 + \sum_{j=2}^n |a_{1j}| f_j$ 
 $Bound = \frac{\max_{i \in N} \{k_i + \epsilon_i / |a_{ii}|\}}{\min_{i \in N} \{S_i\}}$ 

```

▷ Find the first row such that $\epsilon_i > 0$

▷ If $i \leq K$, we have that $w_i = 0$

4. Improvements, numerical tests and bounds for other norms

In the previous section, we derived the bound (2) for the infinity norm of the inverse of a Nekrasov matrix A . For this purpose, we first obtained an adequate scaling matrix S and then we applied the well-known Varah’s bound of Theorem 3.1 to the matrix AS . Nevertheless, any bound applicable to SDD matrices could be applied to AS , and a different choice would lead us to a different bound. In order to illustrate this fact, we are also going to use the bound introduced in [6] for Nekrasov matrices, which in particular improves Varah’s bound for SDD matrices (as proven in Theorem 2.4 of [6]):

$$\|A^{-1}\|_{\infty} \leq \max_{i \in N} \frac{z_i(A)}{|a_{ii}| - h_i(A)}, \tag{4}$$

$$z_1(A) := 1, \quad z_i(A) := \sum_{j=1}^{i-1} |a_{ij}| \frac{z_j(A)}{|a_{jj}|} + 1, \quad i = 2, \dots, n.$$

As in (3), the new bound for $\|A^{-1}\|_{\infty}$ reduces to the product of $\|S\|_{\infty}$ and the bound to $\|(AS)^{-1}\|_{\infty}$ obtained by (4). In fact, taking into account that $z_i(AS) = z_i(A)$ for all $i \in N$, the explicit form of this new bound is:

$$\|A^{-1}\|_{\infty} \leq \max_{i \in N} \left(\frac{h_i(A) + \epsilon_i}{|a_{ii}|} \right) \max_{i \in N} \frac{z_i(A)}{(h_i(A) + \epsilon_i - h_i(AS))}. \tag{5}$$

As shown by the following numerical experiments, this change gives a better bound whenever S follows Theorem 2.2. However, in general the substitution of Varah’s bound is going to increase the computational cost of the bound, while the bound (2) using Theorem 2.1 is not significantly improved. Analogously to (5), if better bounds than (4) for SDD matrices are obtained, then they can be also combined with our bound of Theorem 2.2 to derive sharper bounds than (5), although the computational cost can increase again.

Recent articles have studied the problem of finding bounds for the infinity norm of the inverse of a Nekrasov matrix. In [2], two bounds are introduced and tested with the following six matrices:

$$\begin{aligned}
 A_1 &= \begin{pmatrix} -7 & 1 & -0.2 & 2 \\ 7 & 88 & 2 & -3 \\ 2 & 0.5 & 13 & -2 \\ 0.5 & 3 & 1 & 6 \end{pmatrix}, & A_2 &= \begin{pmatrix} 8 & 1 & -0.2 & 3.3 \\ 7 & 13 & 2 & -3 \\ -1.3 & 6.7 & 13 & -2 \\ 0.5 & 3 & 1 & 6 \end{pmatrix}, \\
 A_3 &= \begin{pmatrix} 21 & -9.1 & -4.2 & -2.1 \\ -0.7 & 9.1 & -4.2 & -2.1 \\ -0.7 & -0.7 & 4.9 & -2.1 \\ -0.7 & -0.7 & -0.7 & 2.8 \end{pmatrix}, & A_4 &= \begin{pmatrix} 5 & 1 & 0.2 & 2 \\ 1 & 21 & 1 & -3 \\ 2 & 0.5 & 6.4 & -2 \\ 0.5 & -1 & 1 & 9 \end{pmatrix}, \\
 A_5 &= \begin{pmatrix} 6 & -3 & -2 \\ -1 & 11 & -8 \\ -7 & -3 & 10 \end{pmatrix}, & A_6 &= \begin{pmatrix} 8 & -0.5 & -0.5 & -0.5 \\ -9 & 16 & -5 & -5 \\ -6 & -4 & 15 & -3 \\ -4.9 & -0.9 & -0.9 & 6 \end{pmatrix}.
 \end{aligned}$$

In more recent works, such as [5,6,9], improvements of these bounds are developed and tested using also these matrices. Since the scaling matrices introduced in Section 2 allowed us to derive different bounds, we are going to compare them with the results obtained in some of the mentioned papers.

We have included results from Gao et al. [5,6]. The bound (4) (which corresponds to the bound 2.4 of [6]) improves those obtained in [2] for Nekrasov matrices (as proven in Theorem 2.3 of [6]). Theorem 9 of [5] gives a sharper bound in some cases, and so we also include it in our comparison.

Table 3 gathers the different bounds. The first row shows the exact infinity norm of the matrices. The data included in the second (corresponding to bound (4)) and third rows are borrowed from the articles that achieved the sharpest bounds. The other rows contain our results, obtained with bounds (2) and (5). In the last case S was given by Theorem 2.1 while in the other cases the diagonal matrix S followed Theorem 2.2. Excluding the case where $\epsilon_n = \Delta_n/2$, our bounds used an appropriate choice of parameters.

Looking at the rows corresponding to Theorem 2.2 we can observe that the obtained bounds are better for A_4 and A_5 , but they are worse in the other cases. With the choice of a diagonal matrix S following Theorem 2.1 and bound (2) we obtained a better bound for every matrix. This option seems superior to the other possibilities. However, it has an intrinsic problem: the choice of the parameters ϵ_i for $i = 1, \dots, n$. Given the right parameters, the obtained bound is excellent. But a bad choice of these values may give a useless bound. In general, it is not clear how to find the optimal values using Theorem 2.1.

Performing more numerical tests, we have seen that the bounds introduced in this paper may be particularly useful when the considered Nekrasov matrix is far from satisfying $|a_{ii}| > h_i(A)$ for some $i \in N \setminus \{n\}$. Looking at the bound for SDD matrices introduced by Varah (Theorem 3.1), we can observe that it depends on all the row sums of the comparison matrix. In particular, bounds for Nekrasov matrices based on Varah’s bound seem to be inversely proportional to $|a_{ii}| - h_i(A)$ for some indices $i \in N$. In order to illustrate this fact, we have modified one entry of all previous examples and we present the bounds obtained for the inverses of these new matrices in Table 4.

$$\begin{aligned}
 \hat{A}_1 &= \begin{pmatrix} -7 & 1 & -\mathbf{3.9} & 2 \\ 7 & 88 & 2 & -3 \\ 2 & 0.5 & 13 & -2 \\ 0.5 & 3 & 1 & 6 \end{pmatrix}, & \hat{A}_2 &= \begin{pmatrix} 8 & 1 & -0.2 & 3.3 \\ 7 & 13 & 2 & -3 \\ -\mathbf{11} & 6.7 & 13 & -2 \\ 0.5 & 3 & 1 & 6 \end{pmatrix}, \\
 \hat{A}_3 &= \begin{pmatrix} 21 & -9.1 & -4.2 & -2.1 \\ -0.7 & 9.1 & -4.2 & -\mathbf{4.2} \\ -0.7 & -0.7 & 4.9 & -2.1 \\ -0.7 & -0.7 & -0.7 & 2.8 \end{pmatrix}, & \hat{A}_4 &= \begin{pmatrix} 5 & 1 & 0.2 & 2 \\ 1 & 21 & 1 & -3 \\ 2 & 0.5 & 6.4 & -2 \\ 0.5 & -1 & \mathbf{15} & 9 \end{pmatrix}, \\
 \hat{A}_5 &= \begin{pmatrix} 6 & -3 & -2 \\ -1 & \mathbf{9} & -8 \\ -7 & -3 & 10 \end{pmatrix}, & \hat{A}_6 &= \begin{pmatrix} 8 & -0.5 & -0.5 & -0.5 \\ -\mathbf{31.9} & 16 & -5 & -5 \\ -6 & -4 & 15 & -3 \\ -4.9 & -0.9 & -0.9 & 6 \end{pmatrix}.
 \end{aligned}$$

In Table 4 we observe that bound (2) is lower than (4) and the bound of [5] even with the choice of ϵ_n as the middle point for matrices $\hat{A}_2, \hat{A}_3, \hat{A}_5$ and \hat{A}_6 . We also obtained tight bounds for the norm of the inverse of \hat{A}_1 . The remaining case, \hat{A}_4 , was built increasing significantly an entry of the last row. As a consequence, all bounds compared in Table 4 obtained weaker results than in Table 3. For \hat{A}_6 , bounds of [5] and [6] (corresponding to bound (4)) are very high while our bounds (using (2)) are all controlled. This phenomenon will be also illustrated with the following family of 3×3 matrices.

Table 3
Upper bounds of $\|A^{-1}\|_\infty$.

Matrix	A_1	A_2	A_3	A_4	A_5	A_6
Exact norm	0.1921	0.2390	0.8759	0.2707	1.1519	0.4474
(4)	0.2632	0.5365	0.9676	0.5556	1.4138	0.4928
Theorem 9 of [5]	0.2505	0.5365	0.9676	0.5038	1.4138	0.4928
(2), $\epsilon_n = \Delta_n/2$	0.6398	1.4406	1.5527	0.7264	1.2974	1.2893
(5), $\epsilon_n = \Delta_n/2$	0.4992	0.7422	1.0632	0.5596	1.2809	1.2893
(2), Theorem 2.2	0.3474	0.8894	1.3325	0.4484	1.1658	1.0796
(5), Theorem 2.2	0.3074	0.5684	0.9735	0.3817	1.1658	1.0436
(2), Theorem 2.1	0.2354	0.5260	0.9273	0.3168	1.1588	0.4527

Table 4
Upper bounds of $\|A^{-1}\|_\infty$.

Matrix	\hat{A}_1	\hat{A}_2	\hat{A}_3	\hat{A}_4	\hat{A}_5	\hat{A}_6
Exact norm	0.2385	0.9827	1.0997	0.2848	2.4545	0.9144
(4)	10.0000	16.2005	5.5357	8.7889	7.0000	266.0000
Theorem 9 of [5]	0.3979	16.2005	5.5357	8.7889	7.0000	266.0000
(2), $\epsilon_n = \Delta_n/2$	1.2345	2.2098	2.3120	17.0569	5.5208	2.6020
(5), $\epsilon_n = \Delta_n/2$	0.6144	1.2071	1.6377	3.1074	5.5208	2.6020
(2), Theorem 2.2	0.8230	1.4732	2.1018	10.2316	4.1085	2.0316
(5), Theorem 2.2	0.5344	0.9923	1.5203	3.0603	3.4717	1.9119
(2), Theorem 2.1	0.3262	1.2642	1.1479	6.6456	2.6180	2.0316

Example 4.1. Let us consider the family of matrices

$$A = \begin{pmatrix} 4 & 2 & 1 \\ \frac{4}{3} - \epsilon & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}, \tag{6}$$

where $0 < \epsilon < \frac{1}{10}$. In this case, $\|A^{-1}\|_\infty < 1.4167$, $h_1(A) = 3$, $h_2(A) = 2 - \frac{3}{4}\epsilon$ and $h_3(A) = \frac{7}{4} - \frac{3}{8}\epsilon$. Then the bounds (2.4) of [6] and Theorem 9 of [5] coincide and are equal to $\frac{16}{9\epsilon} - \frac{1}{3}$. We can observe that this bound is arbitrarily large when $\epsilon \rightarrow 0$. However, our bounds remain controlled. In fact, (2) with $\epsilon_3 = \Delta_3/2$ (and $\epsilon_1 = 0, \epsilon_2 = 0$) gives the bound $16(\frac{1-(3\epsilon/8)}{1+(3\epsilon/2)})$, (2) in Theorem 2.2 is equal to $12(\frac{1-(3\epsilon/8)}{1+(3\epsilon/2)})$ and (2) in Theorem 2.1 can become equal to 12.

We finish this section by applying Theorem 3.2 to derive bounds for other norms. The first result is obtained from applying Theorem 3.2 to the transpose matrix.

Corollary 4.2. Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a matrix with A^T Nekrasov. Then

$$\|A^{-1}\|_1 \leq \frac{\max_{i \in N} \left(\frac{h_i(A^T) + \bar{\epsilon}_i}{|a_{ii}|} \right)}{\min_{i \in N} (\bar{\epsilon}_i - \bar{w}_i + \bar{p}_i)},$$

where $\bar{\epsilon}_i, \bar{w}_i, \bar{p}_i$ are the parameters ϵ_i, w_i, p_i of Theorems 2.2 and 3.2 corresponding to A^T .

Corollary 4.3. Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a matrix with A and A^T Nekrasov and let $\sigma_n(A)$ be its minimal singular value. Then

$$\sigma_n(A) = \|A^{-1}\|_2^{-1} \geq \sqrt{\frac{\min_{i \in N} (\epsilon_i - w_i + p_i) \min_{i \in N} (\bar{\epsilon}_i - \bar{w}_i + \bar{p}_i)}{\max_{i \in N} \left(\frac{h_i(A) + \epsilon_i}{|a_{ii}|} \right) \max_{i \in N} \left(\frac{h_i(A^T) + \bar{\epsilon}_i}{|a_{ii}|} \right)}}, \tag{7}$$

where $\epsilon_i, w_i, p_i, \bar{\epsilon}_i, \bar{w}_i, \bar{p}_i$ are given in Theorems 2.2 and 3.2 and Corollary 4.2.

Proof. It is a consequence of the well-known facts that, for a nonsingular matrix M , its minimal singular value coincides with $\|M^{-1}\|_2^{-1}$ and that $\|M\|_2^2 \leq \|M\|_1 \|M\|_\infty$. \square

In the following example we apply Corollary 4.3 to the suitable matrices from the previous experiments, A_3 and A_4 . The matrix A_3 is an SDD matrix whose transpose is Nekrasov, while A_4 is an SDD matrix whose transpose is also SDD.

Example 4.4. Corollary 4.3 gives a lower bound for the minimal singular value of a Nekrasov matrix whose transpose is also a Nekrasov matrix. We can apply this result to A_3 and A_4 with the choice $\epsilon_1 = \epsilon_2 = \epsilon_3 = 0, \epsilon_4 = \Delta_4/2$, where $\Delta_4(A_3)/2 = 0.6572$ and $\Delta_4(A_4)/2 = 3.9646$. For these matrices, we have that $\sigma_n(A_3) = 1.0943$ and $\sigma_n(A_4) = 4.2327$. The bounds obtained applying (7) are $\sigma_n(A_3) > 0.3357$ and $\sigma_n(A_4) > 0.8680$.

5. Error bounds for LCP of Nekrasov matrices

Given an $n \times n$ real matrix A and $q \in \mathbb{R}^n$, these problems look for solutions $x^* \in \mathbb{R}^n$ of

$$Ax + q \geq 0, \quad x \geq 0, \quad x^T(Ax + q) = 0. \tag{8}$$

This problem (8) is usually denoted by $LCP(A, q)$. A real square matrix is called a P -matrix if all its principal minors are positive. Let us recall (see [18]) that A is a P -matrix if and only if the $LCP(A, q)$ (8) has a unique solution x^* for each $q \in \mathbb{R}^n$.

Let A be a real H -matrix with all its diagonal entries positive. Then A is a P -matrix and so we can apply the third inequality of Theorem 2.3 of [14] and obtain for any $x \in \mathbb{R}^n$ the inequality:

$$\|x - x^*\|_\infty \leq \max_{d \in [0,1]^n} \|(I - D + DA)^{-1}\|_\infty \|r(x)\|_\infty,$$

where we denote by I the $n \times n$ identity matrix, by D the diagonal matrix $D = \text{diag}(d_i)_{i=1}^n$ with $0 \leq d_i \leq 1$ for all $i = 1, \dots, n$, by x^* the solution of the $LCP(A, q)$ and by $r(x) := \min(x, Ax + q)$, where the min operator denotes the componentwise minimum of two vectors.

By (2.4) of [14], given in Theorem 2.1 of [14], when $A = (a_{ij})_{1 \leq i, j \leq n}$ is a real H -matrix with all its diagonal entries positive, then we have

$$\max_{d \in [0,1]^n} \|(I - D + DA)^{-1}\|_\infty \leq \|(\mathcal{M}(A))^{-1} \max(\Lambda, I)\|_\infty, \tag{9}$$

where we denote by $\mathcal{M}(A)$ the comparison matrix of A , by Λ the diagonal part of A ($\Lambda := \text{diag}(a_{ii})_{i=1}^n$) and by $\max(\Lambda, I) := \text{diag}(\max\{a_{ii}, 1\})_{i=1}^n$.

The next theorem, corresponding to Theorem 2.1 of [19], shows the application of obtaining scaling matrices to transform an H -matrix into an SDD matrix in order to derive error bounds for LCP.

Theorem 5.1. Suppose that $A = (a_{ij})_{1 \leq i, j \leq n}$ is an H -matrix with all its diagonal entries positive. Let $S = \text{diag}(s_i)_{i=1}^n, s_i > 0$ for all $i \in N$, be a diagonal matrix such that AS is SDD. For any $i = 1, \dots, n$, let $\tilde{\beta}_i := a_{ii}s_i - \sum_{j \neq i} |a_{ij}| s_j$. Then

$$\max_{d \in [0,1]^n} \|(I - D + DA)^{-1}\|_\infty \leq \max \left\{ \frac{\max_i \{s_i\}}{\min_i \{\tilde{\beta}_i\}}, \frac{\max_i \{s_i\}}{\min_i \{s_i\}} \right\}. \tag{10}$$

The following theorem provides an error bound for the particular LCP associated to a Nekrasov matrix.

Theorem 5.2. Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a Nekrasov matrix with all its diagonal entries positive. Let $S = \text{diag}(s_i)_{i=1}^n$ and ϵ_i ($i \in N$) be the diagonal matrix and positive real numbers, respectively, defined in Theorem 2.2. Then

$$\max_{d \in [0,1]^n} \|(I - D + DA)^{-1}\|_\infty \leq \max \left\{ \frac{1}{\min_i \{\epsilon_i - w_i + p_i\}}, \frac{1}{\min_i \{s_i\}} \right\}, \tag{11}$$

where, for each $i \in N$, p_i and w_i are defined in Theorem 3.2.

Proof. Since A is Nekrasov, $s_i < 1$ for all $i \in N$ and A is an H -matrix. So, we can apply (10) and then it is sufficient to prove that $\tilde{\beta}_i = \epsilon_i - w_i + p_i$ for all $i \in N$. For any $i \in N$, we have

$$\tilde{\beta}_i = a_{ii} \frac{h_i(A) + \epsilon_i}{a_{ii}} - \sum_{j \in N \setminus \{i\}} |a_{ij}| \frac{h_j(A) + \epsilon_j}{a_{jj}}$$

and by (1) we can write

$$\begin{aligned} \tilde{\beta}_i &= \sum_{j=1}^{i-1} |a_{ij}| \frac{h_j(A)}{a_{jj}} + \sum_{j=i+1}^n |a_{ij}| - \sum_{j \in N \setminus \{i\}} |a_{ij}| \frac{h_j(A)}{a_{jj}} + \epsilon_i - \sum_{j \in N \setminus \{i\}} \epsilon_j \frac{|a_{ij}|}{a_{jj}} \\ &= \epsilon_i - \sum_{j=1}^{i-1} \frac{|a_{ij}|}{a_{jj}} + \sum_{j=i+1}^n |a_{ij}| \left(1 - \frac{h_j(A) + \epsilon_j}{a_{jj}} \right) = \epsilon_i - w_i + p_i. \end{aligned}$$

□

As a choice of each parameter ϵ_i ($i \in N$) in Theorem 5.2, we recommend (as we already did for ϵ_n) to choose the middle point of the interval where it lies (see Theorem 2.2). This choice is applied in the following example, where we present a family of matrices for which our bound (11) is a small constant, in contrast to the bounds of [14–16], which can be arbitrarily large. Observe also that these matrices do not satisfy the necessary hypotheses to apply the bound of [1].

Example 5.3. Let us consider the family of matrices

$$A = \begin{pmatrix} K & -K + 2 & -1 \\ -K & K & 0 \\ -K & \frac{-1}{K} & K \end{pmatrix},$$

where $K > 2$. In this case, $h_1(A) = K - 1$, $h_2(A) = K - 1$ and $h_3(A) = K - 1 + \frac{K-1}{K^2}$. Then the bound (9) (of [14]) is equal to $\frac{2K^3+2K}{K^2-1}$ and the bounds of [15,16] coincide and are equal to $\frac{2K^3+2K^2}{K^2-K+1}$. We can observe that these bounds are arbitrarily large when $K \rightarrow \infty$. However, our new bound remains controlled. In fact, (11) with $\epsilon_1 = 0$, $\epsilon_2 = 1/2$ and $\epsilon_3 = \frac{2K^2-2K+3}{4K^2}$ gives the bound $\frac{4K^3}{2K^3-2K^2-2K+1}$.

Acknowledgments

This research has been partially supported by MTM2015-65433-P (MINECO/FEDER) Spanish Research Grant and by Gobierno de Aragón.

References

- [1] M. García-Esnaola, J.M. Peña, Error bounds for linear complementarity problems of Nekrasov matrices, *Numer. Algorithms* 67 (2014) 655–667.
- [2] L. Cvetković, P.-F. Dai, K.D. ski, Y.T. Li, Infinity norm bounds for the inverse of Nekrasov matrices, *Appl. Math. Comput.* 219 (2013) 5020–5024.
- [3] T. Szulc, Some remarks on a theorem of Gudkov, *Linear Algebra Appl.* 225 (1995) 221–235.
- [4] L. Cvetković, V. Kostić, K.D. ski, Max-norm bounds for the inverse of s -Nekrasov matrices, *Appl. Math. Comput.* 218 (2012) 9498–9503.
- [5] L. Gao, C. Li, Y. Li, A new upper bound on the infinity norm of the inverse of Nekrasov matrices, *J. Appl. Math.* 2014 (2014) 8. Art. ID 708128.
- [6] L.Y. Kolotilina, On bounding inverses to Nekrasov matrices in the infinity norm, *J. Math. Sci.* 199 (2014) 432–437.
- [7] L.Y. Kolotilina, Some characterizations of Nekrasov and s -Nekrasov matrices, *J. Math. Sci.* 207 (2015a) 767–775.
- [8] L.Y. Kolotilina, Bounds for the inverses of generalized Nekrasov matrices, *J. Math. Sci.* 207 (2015b) 786–794.
- [9] C. Li, H. Pei, A. Gao, Y. Li, Improvements on the infinity norm bound for the inverse of Nekrasov matrices, *Numer. Algorithms* 71 (2016) 613–630.
- [10] J. Liu, J. Zhang, L. Zhou, G. Tu, The Nekrasov diagonally dominant degree on the Schur complement of Nekrasov matrices and its applications, *Appl. Math. Comput.* 320 (2018) 251–263.
- [11] T. Szulc, L. Cvetković, M. Nedović, Scaling technique for partition-Nekrasov matrices, *Appl. Math. Comput.* 271 (2015) 201–208.
- [12] A. Berman, R.J. Plemmons, Nonnegative matrices in the mathematical sciences, in: *Classics in Applied Mathematics*, 9, SIAM, Philadelphia, 2000.
- [13] R. Bru, I. Giménez, A. Hadjidimos, Is $a \in \mathbb{C}^{n \times n}$ a general h -matrix? *Linear Algebra Appl.* 436 (2012) 364–380.
- [14] X. Chen, S. Xiang, Computation of error bounds for p -matrix linear complementarity problems, *Math. Program. Ser. A* 106 (2006) 513–525.
- [15] C. Li, P. Dai, Y. Li, New error bounds for linear complementarity problems of Nekrasov matrices and b -Nekrasov matrices, *Numer. Algorithms* 74 (2017) 997–1009.
- [16] L. Gao, C. Li, Y. Li, An improvement of the error bounds for linear complementarity problems of Nekrasov matrices, *Linear Multilinear Algebra* 66 (2018) 1505–1519.
- [17] J.M. Varah, A lower bound for the smallest singular value of a matrix, *Linear Algebra Appl.* 11 (1975) 3–5.
- [18] R.W. Cottle, J.S. Pang, R.E. Stone, *The Linear Complementarity Problems*, Academic Press, Boston MA, 1992.
- [19] M. García-Esnaola, J.M. Peña, A comparison of error bounds for linear complementarity problems of h -matrices, *Linear Algebra Appl.* 433 (2010) 956–964.

Article 5

[80] H. Orera and J. M. Peña. B_{π}^R -tensors. *Linear Algebra Appl.* 581 (2019), 247-259.

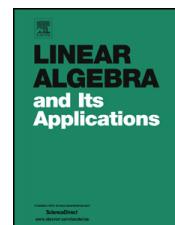


ELSEVIER

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa



B_{π}^R -tensors [☆]

H. Orera ^{*}, J.M. Peña

Departamento de Matemática Aplicada/IUMA, Universidad de Zaragoza, Spain



ARTICLE INFO

Article history:

Received 10 April 2019
 Accepted 11 July 2019
 Available online 16 July 2019
 Submitted by R. Brualdi

MSC:

15A69
 15B48

Keywords:

P -tensor
 B -tensor
 B_{π}^R -tensor
 Positive definite tensor
 Hypermatrix

ABSTRACT

B_{π}^R -matrices were introduced and analyzed in [7]. Here we extend the concept of B_{π}^R -matrix to B_{π}^R -tensor, which generalizes that of B -tensor. We analyze decompositions of these tensors and we prove that odd order B_{π}^R -tensors are P -tensors. We also prove that symmetric even order B_{π}^R -tensors are P -tensors, and so positive definite. The relationship with other classes of tensors is also analyzed.

© 2019 Elsevier Inc. All rights reserved.

1. Introduction

The class of B_{π}^R -matrices was introduced and analyzed in [7]. It contains the class of B -matrices, which we applied to the eigenvalue localization (cf. [8]) and to the Linear Complementarity Problem (cf. [4]). B_{π}^R -matrices are P -matrices under the restriction that π is a nonnegative vector (as commented in Section 2). P -matrices (P_0 -matrices)

[☆] This work was partially supported through the Spanish research grant PGC2018-096321-B-I00 (MCIU/AEI), by Gobierno de Aragón (E41-17R) and FEDER 2014-2020 “Construyendo Europa desde Aragón”.

^{*} Corresponding author.

E-mail addresses: hectororera@unizar.es (H. Orera), jmpena@unizar.es (J.M. Peña).

are matrices whose principal minors are positive (nonnegative, respectively) and they generalize positive definite (semidefinite, respectively) symmetric matrices to the non symmetric case, in the sense that a symmetric matrix is positive definite (semidefinite) if and only if it is a P -matrix (P_0 -matrix, respectively). Among many applications of P -matrices, we recall that a Linear Complementarity Problem has always a unique solution if and only if the associated matrix is a P -matrix (cf. [1]).

In [11], Song and Qi extended P -matrices (P_0 -matrices) to P -tensors (P_0 -tensors) of even order, which are positive definite (semidefinite) when the tensor is symmetric. Later, in [2], a more general definition of P -tensor (P_0 -tensor) was provided, which coincides with that of [11] for the even order tensors and includes many important structured tensors of odd order. In particular, odd order P -tensors contain B -tensors (an easily checkable class of tensors, see [5,9,10]) and strong M -tensors, which extend the corresponding classes of matrices. These last two classes of tensors also belong to the class of MB -tensors (see [6]).

In this paper, we extend the class of B_π^R -matrices to the class of B_π^R -tensors. We analyze these tensors and their relationship with other structured classes of tensors. The paper is organized as follows. In Section 2, we introduce basic concepts and notations. Section 3 provides several decompositions of B_π^R -tensors and symmetric B_π^R -tensors that will be used in the following section. In Section 3 we also provide examples showing that B_π^R -tensors are not necessarily MB -tensors. In Section 4, it is proved that odd order B_π^R -tensors are P -tensors and that symmetric even order B_π^R -tensors are P -tensors and so positive definite. Finally, we also prove that symmetric even order B_π^R -tensors are sum-of-squares tensors.

2. Basic concepts and notations

A real m th order n -dimensional tensor $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ is a multi-array of real entries $a_{i_1 \dots i_m} \in \mathbb{R}$, where $i_k \in N := \{1, \dots, n\}$ for $k = 1, \dots, m$. If the entries of the tensor \mathcal{A} are invariant under any permutation of its indices we say that \mathcal{A} is a *symmetric* tensor. Let us consider the set of entries $a_{ii_2 \dots i_m}$ for $i, i_2, \dots, i_m \in N$ as the i -th row of \mathcal{A} . Then we can define the i -th row sum of \mathcal{A} as

$$R_i(\mathcal{A}) := \sum_{i_2, \dots, i_m=1}^n a_{ii_2 \dots i_m}.$$

A tensor \mathcal{A} is called *diagonally dominant* if

$$|a_{i \dots i}| \geq \sum_{i_2, \dots, i_m \neq (i, \dots, i)}^n |a_{ii_2 \dots i_m}|, \quad i \in N. \quad (1)$$

If (1) holds strictly, then \mathcal{A} is called *strictly diagonally dominant*. We say that $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ is a B -tensor (B_0 -tensor) if

$$R_i(\mathcal{A}) > 0 (\geq 0), i \in N, \tag{2}$$

and

$$\frac{R_i(\mathcal{A})}{n^{m-1}} > a_{ij_2 \dots j_m} (\geq a_{ij_2 \dots j_m}), \forall (j_2, \dots, j_m) \neq (i, \dots, i). \tag{3}$$

Observe that, if \mathcal{A} is a B -tensor, then each diagonal entry $a_{i \dots i}$ is greater than the off-diagonal entries of its row.

We say that $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ is *nonnegative* if $a_{i_1 \dots i_m} \geq 0$ for all $i_1, \dots, i_m \in N$ and that \mathcal{A} is a Z -tensor if all its off-diagonal entries are non-positive, i.e., $a_{i_1 \dots i_m} \leq 0$ whenever $\delta_{i_1 \dots i_m} = 0$, where $\delta_{i_1 \dots i_m}$ is the *generalized Kronecker symbol* with m indices:

$$\delta_{i_1 \dots i_m} = \begin{cases} 1, & \text{if } i_1 = \dots = i_m, \\ 0, & \text{otherwise.} \end{cases}$$

Let \mathcal{I} be the identity tensor, whose off-diagonal entries are 0 and its diagonal entries are 1. A tensor $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ is called an (a *strong*) M -tensor if there exists a nonnegative tensor $\mathcal{B} = (b_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ and a positive scalar $s \geq \rho(\mathcal{B}) (> \rho(\mathcal{B}))$ such that $\mathcal{A} = s\mathcal{I} - \mathcal{B}$, where $\rho(\mathcal{B})$ is the *spectral radius* of \mathcal{B} (see page 15 of [9]). (Strictly) diagonally dominant Z -tensors are clearly (strong) M -tensors.

Let $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ and let $\beta_i(\mathcal{A})$ be given by

$$\beta_i(\mathcal{A}) := \max_{\substack{i_2, \dots, i_m \in N \\ \delta_{ii_2 \dots i_m} = 0}} \{a_{ii_2 \dots i_m}, 0\}. \tag{4}$$

Then we can decompose \mathcal{A} as follows

$$\mathcal{A} = \mathcal{B}^+ + \mathcal{C}, \tag{5}$$

where $\mathcal{B}^+ = (b_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ and $\mathcal{C} = (c_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$, with

$$b_{ii_2 \dots i_m} := a_{ii_2 \dots i_m} - \beta_i(\mathcal{A}) \text{ for } i \in N, \tag{6}$$

and

$$c_{ii_2 \dots i_m} := \beta_i(\mathcal{A}) \text{ for } i \in N. \tag{7}$$

Observe that \mathcal{A} is a B -tensor (B_0 -tensor) if and only if the tensor \mathcal{B}^+ given by (5) and (6) is a strictly diagonally dominant (diagonally dominant) Z -tensor with positive (nonnegative) diagonal entries and the tensor \mathcal{C} given by (5) and (7) is a nonnegative rank-one tensor (see page 3 of [9]). In [6] two new classes of tensors were introduced, MB_0 -tensors and MB -tensors, which clearly contain B_0 -tensors and B -tensors, respectively.

Definition 2.1. Let $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$, and let $\mathcal{A} = \mathcal{B}^+ + \mathcal{C}$ be the decomposition given by (5). Then \mathcal{A} is called an MB_0 -tensor (MB -tensor) if \mathcal{B}^+ is an M -tensor (a strong M -tensor).

A tensor \mathcal{A} is called positive semidefinite (definite) if for each (nonzero) $x \in \mathbb{R}^n$

$$\mathcal{A}x^m \geq 0 \ (\> 0),$$

where $\mathcal{A}x^m = \sum_{i_1, \dots, i_m=1}^n a_{i_1 i_2 \dots i_m} x_{i_1} \cdots x_{i_m}$. Notice that there are not any nontrivial positive semidefinite tensors when m is odd.

We now introduce the important concepts of P -tensor and P_0 -tensor, independently of the order of the tensor (cf. [2]). Let us recall that, given an m -th order tensor $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ and $x \in \mathbb{R}^n$, then $\mathcal{A}x^{m-1} \in \mathbb{R}^n$ is given by

$$(\mathcal{A}x^{m-1})_i := \sum_{i_2, \dots, i_m=1}^n a_{i i_2 \dots i_m} x_{i_2} \cdots x_{i_m}, \quad \text{for each } i = 1, \dots, n.$$

Definition 2.2. (see [2] or page 192 of [9]) A tensor $\mathcal{A} \in \mathbb{R}^{[m,n]}$ is called a P -tensor if for each nonzero $x \in \mathbb{R}^n$ there exists an index $i \in N$ such that

$$x_i^{m-1} (\mathcal{A}x^{m-1})_i > 0. \quad (8)$$

A tensor $\mathcal{A} \in \mathbb{R}^{[m,n]}$ is called a P_0 -tensor if for each nonzero $x \in \mathbb{R}^n$ there exists some index $i \in N$ such that

$$x_i \neq 0 \quad \text{and} \quad x_i^{m-1} (\mathcal{A}x^{m-1})_i \geq 0. \quad (9)$$

In [11] it was shown that in the even order case a symmetric tensor is positive definite (semidefinite) if and only if it is a P -tensor (P_0 -tensor). The following result will be used later.

Proposition 2.3. (Theorem 4.1 of [13] and Lemma 3 of [6]) Let $\mathcal{A} \in \mathbb{R}^{[m,n]}$ be a symmetric Z -tensor and let m be even. Then

1. \mathcal{A} is positive definite if and only if \mathcal{A} is a strong M -tensor.
2. \mathcal{A} is positive semidefinite if and only if \mathcal{A} is an M -tensor.

Given $v \neq 0 \in \mathbb{R}^n$, let us recall that a symmetric rank-one tensor $v^m (= v \otimes \cdots \otimes v) \in \mathbb{R}^{[m,n]}$ is defined by $(v^m)_{i_1 \dots i_m} = v_{i_1} \cdots v_{i_m}$. Then a tensor $\mathcal{A} \in \mathbb{R}^{[m,n]}$ is called *completely positive* if it can be written as

$$\mathcal{A} = \sum_{i=1}^k (u^{(i)})^m, \quad (10)$$

where k is a positive integer and $u^{(i)}$ is a nonnegative vector for $i = 1, \dots, k$. A completely positive tensor is a P_0 -tensor independently of its order (see Proposition 3.3 of [2]).

Now we introduce a class of matrices defined in [7], which will lead to a new class of tensors. Let $\pi = (\pi_1, \dots, \pi_n)$ be a nonnegative vector satisfying

$$0 < \sum_{j=1}^n \pi_j \leq 1, \tag{11}$$

let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a square real matrix with positive row sums and let $R = (R_1, \dots, R_n)$ be the vector formed by the row sums of A . Let us observe that, although in [7] there are no further restrictions on the vector π than (11), it is necessary that π is nonnegative in order to have that a B_π^R -matrix is a P -matrix (see proof of Theorem 3.4 of [7]). So, we say that A is a B_π^R -matrix if for all $i = 1, \dots, n$

$$\pi_j R_i > a_{ij}, \quad \forall j \neq i. \tag{12}$$

Definition 2.4. Let $\pi = (\pi_1, \dots, \pi_n)$ be a nonnegative vector satisfying (11), let $i_1, \dots, i_m \in N$ and let $\pi_{i_1 i_2 \dots i_k} := \pi_{i_1} \pi_{i_2} \dots \pi_{i_k}$ with $k \leq m$. Given a tensor $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m, n]}$ and the vector $R = (R_i)_{i \in N}$ formed by its row sums, we say that \mathcal{A} is a B_π^R -tensor ($(B_\pi^R)_0$ -tensor) if R is positive (nonnegative) and, for all $k \in N$,

$$\pi_{i_2 \dots i_m} R_k > a_{k i_2 \dots i_m} (\geq a_{k i_2 \dots i_m}), \quad \text{with } \delta_{k i_2 \dots i_m} = 0. \tag{13}$$

When $\pi_j = \frac{1}{n}$ for $j \in N$ this definition of a B_π^R -tensor ($(B_\pi^R)_0$ -tensor) coincides with that of a B -tensor (B_0 -tensor).

3. Decompositions of B_π^R -tensors and examples

In this section we present several decompositions of B_π^R -tensors and $(B_\pi^R)_0$ -tensors that will be used later, as well as examples of B_π^R -tensors that are not MB -tensors.

Theorem 3.1. Let $\mathcal{A} \in \mathbb{R}^{[m, n]}$ be a B_π^R -tensor. Then we can write \mathcal{A} as

$$\mathcal{A} = \mathcal{B} + \mathcal{C},$$

where \mathcal{B} is a strictly diagonally dominant M -tensor and \mathcal{C} is a nonnegative rank-one tensor.

Proof. Since π satisfies (11) there exists an index $k \in N$ such that $\pi_k > 0$. Then, for every $i \in N$ and (i_2, \dots, i_m) with $k = i_j$ for some $j \in \{2, \dots, m\}$, there exists $0 < \varepsilon_{i_2 \dots i_m} < \pi_k$ such that

$$a_{i i_2 \dots i_m} - \hat{\pi}_{i_2 \dots i_m} R_i < 0,$$

where $\hat{\pi}_{i_2 \dots i_m} = \hat{\pi}_{i_2} \cdots \hat{\pi}_{i_m}$ with

$$\hat{\pi}_{i_r} = \begin{cases} \pi_{i_r}, & r \neq j \ (i_r \neq k), \\ \pi_k - \varepsilon_{i_2 \dots i_m}, & r = j \ (i_r = k), \end{cases} \text{ for } r = 2, \dots, m.$$

Then we take $\varepsilon := \min\{\varepsilon_{i_2 \dots i_m}\}$ and we define

$$\tilde{\pi}_{i_2 \dots i_m} := \tilde{\pi}_{i_2} \cdots \tilde{\pi}_{i_m}, \text{ with } \tilde{\pi}_{i_r} = \begin{cases} \pi_{i_r}, & r \neq j, \\ \pi_k - \varepsilon, & r = j, \end{cases} \text{ for } r = 2, \dots, m.$$

We can decompose \mathcal{A} as

$$\mathcal{A} = \mathcal{B}_{\tilde{\pi}}^+ + \mathcal{C}_{\tilde{\pi}}, \tag{14}$$

where $\mathcal{B}_{\tilde{\pi}}^+ = (b_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ and $\mathcal{C}_{\tilde{\pi}} = (c_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$,

$$b_{ii_2 \dots i_m} := a_{ii_2 \dots i_m} - \tilde{\pi}_{i_2 \dots i_m} R_i, \quad i \in N,$$

and

$$c_{ii_2 \dots i_m} := \tilde{\pi}_{i_2 \dots i_m} R_i, \quad i \in N. \tag{15}$$

So, we have that $\mathcal{C}_{\tilde{\pi}}$ is a nonnegative rank-one tensor. On the other hand, we have that the i -th row sum of $\mathcal{B}_{\tilde{\pi}}^+$, with $i \in N$, is positive:

$$\begin{aligned} \sum_{i_2, \dots, i_m=1}^n b_{ii_2 \dots i_m} &= \sum_{i_2, \dots, i_m \neq k}^n (a_{ii_2 \dots i_m} - \pi_{i_2 \dots i_m} R_i) + \sum_{\substack{\exists j=2, \dots, m \\ \text{s.t. } i_j=k}}^n (a_{ii_2 \dots i_m} - \tilde{\pi}_{i_2 \dots i_m} R_i) \\ &= R_i - R_i \sum_{i_2, \dots, i_m \neq k}^n \pi_{i_2 \dots i_m} - R_i \sum_{\substack{\exists j=2, \dots, m \\ \text{s.t. } i_j=k}}^n \tilde{\pi}_{i_2 \dots i_m} \\ &> R_i \left(1 - \sum_{i_2, \dots, i_m=1}^n \pi_{i_2 \dots i_m} \right) \geq 0. \end{aligned} \tag{16}$$

Since $\mathcal{B}_{\tilde{\pi}}^+$ is a Z -tensor with positive row sums it is strictly diagonally dominant. \square

We can derive a similar decomposition for $(B_{\pi}^R)_0$ -tensors.

Theorem 3.2. *Let $\mathcal{A} \in \mathbb{R}^{[m,n]}$ be a $(B_{\pi}^R)_0$ -tensor. Then we can write \mathcal{A} as*

$$\mathcal{A} = \mathcal{B}_{\pi}^+ + \mathcal{C}_{\pi}, \tag{17}$$

where \mathcal{B}_{π}^+ is a diagonally dominant M -tensor and \mathcal{C}_{π} is a nonnegative tensor.

Proof. In order to prove this result, we can follow the proof of Theorem 3.1 with $\varepsilon = 0$. Following (14), we can decompose \mathcal{A} as

$$\mathcal{A} = \mathcal{B}_\pi^+ + \mathcal{C}_\pi,$$

where \mathcal{B}_π^+ is a Z -tensor with nonnegative row sums, or equivalently, a diagonally dominant Z -tensor. \square

Let $\mathbb{1} = \pi^m \in \mathbb{R}^{[m,n]}$ and let $J \subseteq N$. Then we denote by $\mathbb{1}^J$ a tensor $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ such that $a_{i_1 \dots i_m} = (\pi^m)_{i_1 \dots i_m} = \pi_{i_1} \cdots \pi_{i_m}$ whenever $i_j \in J$ for all $j = 1, \dots, m$ and such that all its remaining entries are zero. The tensors $\mathbb{1}^J$ play a key role in the new decomposition for symmetric B_π^R -tensors of even and odd order introduced in the next theorem. This decomposition will allow us to prove, in the following section, that a symmetric B_π^R -tensor of even order is positive definite.

Theorem 3.3. *Let $\mathcal{A} \in \mathbb{R}^{[m,n]}$ be a symmetric B_π^R -tensor. Then either \mathcal{A} is a strictly diagonally dominant symmetric Z -tensor or it can be written as*

$$\mathcal{A} = \mathcal{M} + \sum_{i=1}^s h_i \mathbb{1}^{J_i}, \tag{18}$$

where \mathcal{M} is a strictly diagonally dominant Z -tensor, s is a positive integer, $h_k > 0$, $J_k \subseteq N$ for $k = 1, \dots, s$ and $J_s \subsetneq J_{s-1} \subsetneq \dots \subsetneq J_1$.

Proof. Let $J_1 := \{i \in N \text{ such that there is at least one positive off-diagonal entry in the } i\text{-th row of } \mathcal{A}\}$. If $J_1 = \emptyset$, then \mathcal{A} is a symmetric Z -tensor with positive row sums, and so \mathcal{A} is also strictly diagonally dominant. So, let us suppose that $J_1 \neq \emptyset$. Let us define

$$\gamma_i(\mathcal{A}) := \max_{\substack{i_2, \dots, i_m \in J_1 \\ \delta_{i_2 \dots i_m} = 0}} \left\{ \frac{a_{i i_2 \dots i_m}}{\pi_{i i_2 \dots i_m}} \right\}, \text{ for } i \in J_1. \tag{19}$$

The choice of the indices implies that $\gamma_i(\mathcal{A})$ is well-defined, i.e., we avoid any division by zero. Let us suppose that $\pi_j = 0$ for some $j \in N$. In that case, from (13) we see that

$$0 = \pi_{j i_3 \dots i_m} R_{i_2} > a_{i_2 j i_3 \dots i_m}, \tag{20}$$

for any i_2, \dots, i_m with $\delta_{i_2 j i_3 \dots i_m} = 0$. Combining the symmetry of \mathcal{A} and the bound (20), we deduce that $a_{i_1 i_2 \dots i_m} < 0$ whenever $\delta_{i_1 i_2 \dots i_m} = 0$ and $i_k = j$ for some $k = 1, \dots, m$. In particular, it holds for $k = 1$, and this fact implies that $j \notin J_1$, and so, formula (19) is well-defined.

We take $h_1 := \min_{i \in J_1} \{\gamma_i(\mathcal{A})\}$. Let $j \in J_1$ be an index such that $h_1 = \gamma_j(\mathcal{A})$ and let $j_2, \dots, j_m \in N$ be the indices such that $\gamma_j(\mathcal{A}) = \frac{a_{j j_2 \dots j_m}}{\pi_{j j_2 \dots j_m}}$. Then $\mathcal{A}^{(2)} := \mathcal{A} - h_1 \mathbb{1}^{J_1}$ is also a symmetric B_π^R -tensor. Furthermore, $\mathcal{A}^{(2)}$ satisfies that $\gamma_i(\mathcal{A}^{(2)}) = \gamma_i(\mathcal{A}) - h_1 \geq 0$

for $i \in J_1$ and that $\gamma_j(\mathcal{A}^{(2)}) = 0$. In order to check that $\gamma_j(\mathcal{A}^{(2)}) = 0$ we are going to see that all the off-diagonal entries in the j th row of $\mathcal{A}^{(2)} = \left(a_{i_1 \dots i_m}^{(2)} \right)$ are nonpositive:

$$a_{j i_2 \dots i_m}^{(2)} = a_{j i_2 \dots i_m} - \frac{a_{j j_2 \dots j_m}}{\pi_j \pi_{j_2 \dots j_m}} \pi_j \pi_{i_2 \dots i_m} \leq 0, \tag{21}$$

which holds by the choice of h_1 as $\gamma_j(\mathcal{A})$. In order to see that $\mathcal{A}^{(2)}$ is a B_π^R -tensor let us first prove that it has positive row sums. In fact, for each $i \in J_1$,

$$\begin{aligned} R_i(\mathcal{A}^{(2)}) &= \sum_{i_2, \dots, i_m=1}^n a_{i i_2 \dots i_m}^{(2)} = \sum_{i_2, \dots, i_m=1}^n a_{i i_2 \dots i_m} - \sum_{i_2, \dots, i_m \in J_1} h_1 \pi_{i_2 \dots i_m} \\ &= R_i - h_1 \sum_{i_2, \dots, i_m \in J_1} \pi_{i_2 \dots i_m} = R_i - \frac{a_{j j_2 \dots j_m}}{\pi_j \pi_{j_2 \dots j_m}} \pi_i \sum_{i_2, \dots, i_m \in J_1} \pi_{i_2 \dots i_m} \\ &\geq R_i - \gamma_i(\mathcal{A}) \pi_i \sum_{i_2, \dots, i_m \in J_1} \pi_{i_2 \dots i_m} \geq R_i - \max_{\substack{i_2, \dots, i_m \in J_1 \\ \delta_{i_2 \dots i_m} = 0}} \left\{ \frac{a_{i i_2 \dots i_m}}{\pi_{i_2 \dots i_m}} \right\} > 0. \end{aligned}$$

We also need to prove that (13) holds. We have that $a_{i i_2 \dots i_m}^{(2)} = a_{i i_2 \dots i_m}$ whenever $i \notin J_1$, so let us impose that $i \in J_1$. Then, for the i -th row of $\mathcal{A}^{(2)}$, we have to see that

$$\pi_{i_2 \dots i_m} \left(R_i - \pi_i \sum_{j_2, \dots, j_m \in J_1} h_1 \pi_{j_2 \dots j_m} \right) > a_{i i_2 \dots i_m} - h_1 \pi_i \pi_{i_2 \dots i_m}. \tag{22}$$

After some computations, we deduce that condition (22) is equivalent to

$$\begin{aligned} R_i &> \frac{a_{i i_2 \dots i_m}}{\pi_{i_2 \dots i_m}} - h_1 \pi_i + \pi_i \sum_{j_2, \dots, j_m \in J_1} h_1 \pi_{j_2 \dots j_m} \\ &= \frac{a_{i i_2 \dots i_m}}{\pi_{i_2 \dots i_m}} - h_1 \pi_i \left(1 - \sum_{j_2, \dots, j_m \in J_1} \pi_{j_2 \dots j_m} \right). \end{aligned} \tag{23}$$

Since $0 < \sum_{j_2, \dots, j_m}^n \pi_{j_2 \dots j_m} \leq 1$, the inequality (23) holds from (13) and, as a consequence, (22) also holds. Finally, since both \mathcal{A} and $h_1 \mathbb{1}^{J_1}$ are symmetric, we have that $\mathcal{A}^{(2)}$ is a symmetric B_π^S -tensor, for some positive vector $S \in \mathbb{R}^n$. Moreover, we can follow the same process with $\mathcal{A}^{(2)}$. Let us define $J_2 := \{i \in N \text{ such that there is at least one positive off-diagonal entry in the } i\text{-th row of } \mathcal{A}^{(2)}\}$. By the definition of $\mathcal{A}^{(2)}$ we have that $J_2 \subsetneq J_1$. So, if we repeat this process s times (with $s \leq n$), we would end up with a tensor $\mathcal{A}^{(s+1)}$ that satisfies $J_{s+1} = \emptyset$.

This fact implies that we can write \mathcal{A} as

$$\mathcal{A} = \mathcal{A}^{(s+1)} + \sum_{i=1}^s h_i \mathbb{1}^{J_i}, \tag{24}$$

where $\mathcal{A}^{(s+1)}$ is a Z -tensor with positive row sums, and so it is strictly diagonally dominant, and the decomposition (24) corresponds to (18). \square

Following the proof of Theorem 3.3 it is straightforward to deduce the equivalent decomposition to (18) for $(B_\pi^R)_0$ -tensors.

Theorem 3.4. *Let $\mathcal{A} \in \mathbb{R}^{[m,n]}$ be a symmetric $(B_\pi^R)_0$ -tensor. Then either \mathcal{A} is a diagonally dominant symmetric Z -tensor or it can be written as*

$$\mathcal{A} = \mathcal{M} + \sum_{i=1}^s h_i \Pi^{J_i}, \tag{25}$$

where \mathcal{M} is a diagonally dominant Z -tensor, s is a positive integer, $h_k > 0$, $J_k \subseteq N$ for $k = 1, \dots, s$ and $J_s \subsetneq J_{s-1} \subsetneq \dots \subsetneq J_1$.

Let us now consider the relationship between these new classes of B_π^R -tensors and $(B_\pi^R)_0$ -tensors with other generalizations of B -tensors. As commented in Section 2, we already know that a B -tensor is a B_π^R -tensor for $\pi = (\frac{1}{n}, \dots, \frac{1}{n})$. The next result about Z -tensors that are B_π^R -tensors or $(B_\pi^R)_0$ -tensors follows from Definition 2.4.

Proposition 3.5. *Let $\mathcal{A} \in \mathbb{R}^{[m,n]}$ be a Z -tensor. Then*

1. \mathcal{A} is strictly diagonally dominant if and only if it is a B_π^R -tensor for any positive vector π satisfying (11).
2. \mathcal{A} is diagonally dominant if and only if it is a $(B_\pi^R)_0$ -tensor for any nonnegative vector π satisfying (11).

It is known that a Z -tensor is a B_0 -tensor (B -tensor) if and only if it is diagonally dominant (strictly diagonally dominant). Proposition 3.5 implies that, for the particular case of Z -tensors, the classes of B -tensors and B_π^R -tensors are the same. In particular, a Z -tensor that is a B_π^R -tensor is also an MB -tensor. We also have that a Z -tensor that is a $(B_\pi^R)_0$ -tensor is an MB_0 -tensor. But, as the following examples show, in general a B_π^R -tensor is not necessarily an MB -tensor or even an MB_0 -tensor.

Example 3.6. Let $\mathcal{A} = (a_{i_1 i_2 i_3}) \in \mathbb{R}^{[3,2]}$ be such that

$$\begin{aligned} a_{111} &= \frac{1}{2}, \quad a_{122} = 1, \quad a_{112} = a_{121} = 0, \\ a_{222} &= 20, \quad a_{212} = a_{221} = 1, \quad a_{211} = 0. \end{aligned}$$

We have that \mathcal{A} is a B_π^R -tensor with $\pi = (\frac{1}{10}, \frac{9}{10})$ but it is not an MB_0 -tensor. If we decompose \mathcal{A} using (5)

$$\mathcal{A} = \mathcal{B}^+ + \mathcal{C},$$

we have that \mathcal{B}^+ is not an M -tensor. In order to see this fact, it is sufficient to check its first row. Since $\beta_1(\mathcal{A}) = 1$,

$$b_{111} = -\frac{1}{2}, b_{122} = 0, b_{121} = b_{112} = -1. \quad (26)$$

It is known that the diagonal entries of an M -tensor must be nonnegative (see Proposition 15 of [3]). This is also true for the case of even order tensors. Let $\mathcal{A} = (a_{i_1 i_2 i_3 i_4}) \in \mathbb{R}^{[4,2]}$ be such that

$$\begin{aligned} a_{1111} &= \frac{1}{2}, a_{1222} = 1, a_{1112} = a_{1121} = a_{1211} = a_{1221} = a_{1122} = a_{1212} = 0, \\ a_{2222} &= 20, a_{2122} = a_{2212} = a_{2221} = 1, a_{2111} = a_{2211} = a_{2121} = a_{2112} = 0. \end{aligned}$$

Then we have that \mathcal{A} is a B_π^R -tensor with $\pi = (\frac{1}{10}, \frac{9}{10})$ but it is not an MB_0 -tensor. The reasoning given in the previous example can also be applied to this case in order to check that \mathcal{A} is not an MB_0 -tensor.

The following section will analyze the relationship of B_π^R -tensors and $(B_\pi^R)_0$ -tensors with P -tensors and P_0 -tensors.

4. B_π^R -tensors and P -tensors

The following result shows that all B_π^R -tensors of odd order are P -tensors.

Theorem 4.1. *Let $\mathcal{A} \in \mathbb{R}^{[m,n]}$ be a B_π^R -tensor with m odd. Then \mathcal{A} is a P -tensor.*

Proof. Following Theorem 3.1 we can decompose \mathcal{A} by (14) as

$$\mathcal{A} = \mathcal{B}_\pi^+ + \mathcal{C}_\pi,$$

where \mathcal{B}_π^+ is a strong M -tensor since it is a Z -tensor with positive row sums. Given $x \neq 0 \in \mathbb{R}^n$, for any $i \in N$ one can derive from (15)

$$(\mathcal{C}_\pi x^{m-1})_i = R_i(\tilde{\pi}_1 x_1 + \dots + \tilde{\pi}_n x_n)^{m-1} \geq 0,$$

and so we deduce that

$$x_i^{m-1} (\mathcal{C}_\pi x^{m-1})_i \geq 0. \quad (27)$$

Let $i \in N$ be such that $x_i^{m-1} (\mathcal{B}_\pi^+ x^{m-1})_i > 0$. Then we can use the decomposition (14) to see that

$$x_i^{m-1} (\mathcal{A}x^{m-1})_i = x_i^{m-1} (\mathcal{C}_\pi x^{m-1})_i + x_i^{m-1} (\mathcal{B}_\pi^+ x^{m-1})_i > 0,$$

and so, we conclude that \mathcal{A} is a P -tensor. \square

Analogously, a $(B_\pi^R)_0$ -tensor of odd order is a P_0 -tensor.

Theorem 4.2. *Let $\mathcal{A} \in \mathbb{R}^{[m,n]}$ be a $(B_\pi^R)_0$ -tensor with m odd. Then \mathcal{A} is a P_0 -tensor.*

Proof. In order to prove this result we can use the decomposition given by Theorem 3.2,

$$\mathcal{A} = \mathcal{B}_\pi^+ + \mathcal{C}_\pi,$$

where \mathcal{B}_π^+ is a diagonally dominant Z -tensor, and as a consequence, an M -tensor. Given $x \neq 0 \in \mathbb{R}^n$, let $i \in N$ be such that $x_i^{m-1} (\mathcal{B}_\pi^+ x^{m-1})_i \geq 0$. Then we can derive

$$x_i^{m-1} (\mathcal{A}x^{m-1})_i = x_i^{m-1} (\mathcal{C}_\pi x^{m-1})_i + x_i^{m-1} (\mathcal{B}_\pi^+ x^{m-1})_i \geq 0,$$

and \mathcal{A} is a P_0 -tensor. \square

A B_π^R -matrix is a P -matrix, and so, a B_π^R -tensor of order 2 is a P -tensor. However, in general a B -tensor of even order $m \geq 4$ is not a P -tensor (see Proposition 3.1 of [12]). But, as the following result shows, a symmetric B_π^R -tensor of even order is also a P -tensor.

Theorem 4.3. *Let $\mathcal{A} \in \mathbb{R}^{[m,n]}$ be a symmetric B_π^R -tensor. If m is even, then \mathcal{A} is positive definite, and so, a P -tensor.*

Proof. By Theorem 3.3, either \mathcal{A} is a symmetric strong M -tensor or we can decompose \mathcal{A} by (18) as the sum of a symmetric strong M -tensor, \mathcal{M} , and a completely positive tensor, $\sum_{i=1}^s h_i \Pi^{J_i}$. By Proposition 2.3, a symmetric strong M -tensor with even order is positive definite, so let us suppose that we are in the case where the decomposition (18) of \mathcal{A} is not trivial. Let us define $\pi_{J_k} = (\tilde{\pi}_1, \dots, \tilde{\pi}_n)$, where $\tilde{\pi}_i := \pi_i$ if $i \in J_k$ and $\tilde{\pi}_i := 0$ otherwise. By (18) we have that

$$\begin{aligned} \mathcal{A}x^m &= \mathcal{M}x^m + \sum_{i=1}^s h_i \Pi^{J_i} x^m = \mathcal{M}x^m + \sum_{i=1}^s h_i (\pi_{J_i}^T x)^m \\ &= \mathcal{M}x^m + \sum_{i=1}^s h_i (\tilde{\pi}_1 x_1 + \dots + \tilde{\pi}_m x_m)^m \geq \mathcal{M}x^m > 0, \end{aligned}$$

and \mathcal{A} is positive definite, and so, a P -tensor. \square

The analogous result for $(B_\pi^R)_0$ -tensors also holds.

Theorem 4.4. Let $\mathcal{A} \in \mathbb{R}^{[m,n]}$ be a symmetric $(B_\pi^R)_0$ -tensor. If m is even, then \mathcal{A} is positive semidefinite, and so, a P_0 -tensor.

Proof. By Theorem 3.4, either \mathcal{A} is a symmetric M -tensor or we can decompose \mathcal{A} following (25) as the sum of a symmetric M -tensor, \mathcal{M} , and a completely positive tensor, $\sum_{i=1}^s h_i \Pi^{J_i}$. Since \mathcal{A} is symmetric and m is even this decomposition implies that \mathcal{A} is positive semidefinite and a P_0 -tensor. \square

Suppose that $\mathcal{A} \in \mathbb{R}^{[m,n]}$ is a symmetric tensor with m even. Let

$$f_{\mathcal{A}}(x) = \mathcal{A}x^m = \sum_{i_1, \dots, i_m=1}^n a_{i_1 \dots i_m} x_{i_1} \cdots x_{i_m},$$

where $x \in \mathbb{R}^n$. If it is possible to write

$$f_{\mathcal{A}}(x) = \sum_{j=1}^r f_j(x)^2,$$

where f_j for $j = 1, \dots, r$ are homogeneous polynomials of degree $\frac{m}{2}$, then f is called a *sum-of-squares* (SOS) polynomial, and the symmetric tensor \mathcal{A} is called a *sum-of-squares* (SOS) tensor. Applications of SOS tensors can be seen in [9].

Theorem 4.5. Let $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ be a symmetric $(B_\pi^R)_0$ -tensor with m even. Then \mathcal{A} is an SOS tensor.

Proof. We have already seen that we can decompose \mathcal{A} using (25), and so we can write \mathcal{A} as the sum of a symmetric M -tensor, which is an SOS tensor, and $\sum_{i=1}^s h_i \Pi^{J_i}$. Let us see that Π^{J_k} is an SOS tensor. Let us define $\pi_{J_k} = (\tilde{\pi}_1, \dots, \tilde{\pi}_n)$, where $\tilde{\pi}_i := \pi_i$ if $i \in J_k$ and $\tilde{\pi}_i := 0$ otherwise. Then:

$$\begin{aligned} f_{\Pi^{J_k}}(x) &= \sum_{i_1, \dots, i_m=1}^n \tilde{\pi}_{i_1 \dots i_m} x_{i_1} \cdots x_{i_m} = \sum_{i_1, \dots, i_m=1}^n \tilde{\pi}_{i_1} \cdots \tilde{\pi}_{i_m} x_{i_1} \cdots x_{i_m} \\ &= (\tilde{\pi}_{i_1} x_{i_1} + \dots + \tilde{\pi}_{i_n} x_{i_n})^m. \quad \square \end{aligned}$$

Declaration of Competing Interest

There is no competing interest.

References

- [1] A. Berman, R.J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Classics in Applied Mathematics, vol. 9, SIAM, Philadelphia, 2000.

- [2] W. Ding, Z. Luo, L. Qi, P -tensors, P_0 -tensors and their applications, *Linear Algebra Appl.* 555 (2018) 336–354.
- [3] W. Ding, L. Qi, Y. Wei, M -tensors and nonsingular M -tensors, *Linear Algebra Appl.* 439 (2013) 3264–3278.
- [4] M. García-Esnaola, J.M. Peña, Error bounds for linear complementarity problems of B -matrices, *Appl. Math. Lett.* 22 (2009) 1071–1075.
- [5] C. Li, Y. Li, Double B -tensors and quasi-double B -tensors, *Linear Algebra Appl.* 466 (2015) 343–356.
- [6] C. Li, L. Qi, Y. Li, MB -tensors and MB_0 -tensors, *Linear Algebra Appl.* 484 (2015) 141–153.
- [7] M. Neumann, J.M. Peña, O. Pryporova, Some classes of nonsingular matrices and applications, *Linear Algebra Appl.* 438 (2013) 1936–1945.
- [8] J.M. Peña, A class of P -matrices with applications to the localization of the eigenvalues of a real matrix, *SIAM J. Matrix Anal. Appl.* 22 (2001) 1027–1037.
- [9] L. Qi, Z. Luo, *Tensor Analysis*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2017.
- [10] L. Qi, Y. Song, An even order symmetric B tensor is positive definite, *Linear Algebra Appl.* 457 (2014) 303–312.
- [11] Y. Song, L. Qi, Properties of some classes of structured tensors, *J. Optim. Theory Appl.* 165 (2015) 854–873.
- [12] P. Yuan, L. You, Some remarks on P , P_0 , B and B_0 tensors, *Linear Algebra Appl.* 459 (2014) 511–521.
- [13] L. Zhang, L. Qi, G. Zhou, M -tensors and some applications, *SIAM J. Matrix Anal. Appl.* 35 (2014) 437–452.

Article 6

- [18] J. Delgado, H. Orera and J. M. Peña. Accurate bidiagonal decomposition and computations with generalized Pascal matrices. *J. Comput. Appl. Math.* 391 (2021), Paper No. 113443, 10 pp.



Accurate bidiagonal decomposition and computations with generalized Pascal matrices

J. Delgado^{*,1}, H. Orera¹, J.M. Peña¹

Departamento de Matemática Aplicada, Universidad de Zaragoza, Spain

ARTICLE INFO

Article history:

Received 11 February 2020

Received in revised form 12 November 2020

MSC:

65F05

65F15

65G50

15A23

05A05

11B65

Keywords:

Generalized Pascal matrices

High relative accuracy

Total positivity

ABSTRACT

This paper provides an accurate method to obtain the bidiagonal factorization of many generalized Pascal matrices, which in turn can be used to compute with high relative accuracy the eigenvalues, singular values and inverses of these matrices. Numerical examples are included.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Finding classes of matrices relevant in applications for which algebraic computations can be performed with high relative accuracy (HRA) is an active research topic of great interest in recent years. This goal has been achieved for some subclasses of totally positive matrices (see, for instance, [1–7]). Let us recall that a matrix is called *totally positive* (TP) if all their minors are nonnegative and, if they are all positive, then the matrix is called *strictly totally positive* (STP). These classes of matrices play an important role in many fields such as approximation theory, statistics, mechanics, computer-aided geometric design, economics, combinatorics or biology (see [8–10]). For the subclasses of TP matrices mentioned above, their bidiagonal decomposition (see Section 2) was obtained with HRA, and then the algorithms given in [11–13] permit to compute with HRA many algebraic calculations: all their eigenvalues and singular values, their inverses, or the solution of some linear systems. Recall that a real value x is obtained with HRA if the relative error of the computed value \tilde{x} satisfies $\|x - \tilde{x}\|/\|x\| < Ku$, where K is a positive constant independent of the arithmetic precision and u is the unit round-off. It is well known that an algorithm can be performed with HRA if all the included subtractions are of initial data, that is, if it only includes products, divisions, sums of numbers of the same sign and subtractions of the initial data (cf. [1,12,14]).

The lower triangular Pascal matrix $P = (p_{ij})_{1 \leq i, j \leq n+1}$ (with $p_{ij} = \binom{i-1}{j-1}$ for $1 \leq j \leq i \leq n+1$ and $p_{ij} := 0$ when $j > i$) and the symmetric Pascal matrix $R = (r_{ij})_{1 \leq i, j \leq n+1}$ (with $r_{ij} = \binom{i+j-2}{j-1}$) are naturally derived from the Pascal triangle. The

* Corresponding author.

E-mail address: jorgedel@unizar.es (J. Delgado).

¹ This work was partially supported through the Spanish research grant PGC2018-096321-B-I00 (MCIU/AEI), by Gobierno de Aragón (E41-17R) and Feder 2014–2020 “Construyendo Europa desde Aragón”.

matrix $R = PP^T$ is also called Pascal matrix. This paper deals with some classes of matrices (see [15–18]) generalizing the lower triangular Pascal matrix and the symmetric Pascal matrix. These generalized classes of Pascal matrices arise in applications such as filter design, probability, combinatorics, signal processing or electrical engineering (see [19] and its references). In [19], one can also see some concrete applications of solving linear systems with these matrices. Let us recall that the bidiagonal decomposition of a Pascal matrix is well known and has the remarkable property that it is formed by 1's (see [12,20]). In this paper, we show that the bidiagonal decompositions of these generalized Pascal matrices can be obtained with HRA, and so the remaining algebraic calculations mentioned above can be also computed with HRA. Although Pascal matrices are ill-conditioned (see [20]) and the bidiagonal decompositions of their generalizations are not as simple as those with the Pascal matrix, we can still guarantee the mentioned algebraic calculations with HRA.

In Section 2 we present auxiliary results concerning the bidiagonal decomposition of nonsingular TP matrices and some basic definitions of generalized Pascal matrices. In Section 3 we obtain the bidiagonal decomposition of generalized triangular Pascal matrices and of lattice path matrices, which in turn contain many classical generalized Pascal matrices. In many cases, we prove that they are TP or STP and show that the algebraic calculations mentioned above can be computed with HRA. Section 4 includes numerical experiments showing the great accuracy of the proposed method. Finally, Section 5 summarized the main conclusions of the paper.

2. Auxiliary results and basic definitions

Neville elimination (NE) is an alternative procedure to Gaussian elimination that produces zeros in a column of a matrix by adding to each row an appropriate multiple of the previous one. This elimination procedure is very useful when dealing with some classes of matrices such as TP matrices. For more details on NE see [21,22]. Given a nonsingular matrix $A = (a_{ij})_{1 \leq i, j \leq n}$, the Neville elimination procedure consists of $n - 1$ steps and leads to the following sequence of matrices:

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

where U is an upper triangular matrix.

The matrix $\tilde{A}^{(k)} = (\tilde{a}_{ij}^{(k)})_{1 \leq i, j \leq n}$ is obtained from the matrix $A^{(k)} = (a_{ij}^{(k)})_{1 \leq i, j \leq n}$ by a row permutation that moves to the bottom the rows with a zero entry in column k below the main diagonal. For nonsingular TP matrices, it is always possible to perform NE without row exchanges (see [21]). If a row permutation is not necessary at the k th step, we have that $\tilde{A}^{(k)} = A^{(k)}$. The entries of $A^{(k+1)} = (a_{ij}^{(k+1)})_{1 \leq i, j \leq n}$ can be obtained from $\tilde{A}^{(k)} = (\tilde{a}_{ij}^{(k)})_{1 \leq i, j \leq n}$ using the formula:

$$a_{ij}^{(k+1)} = \begin{cases} \tilde{a}_{ij}^{(k)} - \frac{\tilde{a}_{ik}^{(k)}}{\tilde{a}_{i-1,k}^{(k)}} \tilde{a}_{i-1,j}^{(k)}, & \text{if } k \leq j < i \leq n \text{ and } \tilde{a}_{i-1,k}^{(k)} \neq 0, \\ \tilde{a}_{ij}^{(k)}, & \text{otherwise,} \end{cases} \tag{2}$$

for $k = 1, \dots, n - 1$. The (i, j) pivot of the NE of A is given by

$$p_{ij} = \tilde{a}_{ij}^{(j)}, \quad 1 \leq j \leq i \leq n.$$

If $i = j$ we say that p_{ii} is a *diagonal pivot*. The (i, j) multiplier of the NE of A , with $1 \leq j \leq i \leq n$, is defined as

$$m_{ij} = \begin{cases} \frac{\tilde{a}_{ij}^{(j)}}{\tilde{a}_{i-1,j}^{(j)}} = \frac{p_{ij}}{p_{i-1,j}}, & \text{if } \tilde{a}_{i-1,j}^{(j)} \neq 0, \\ 0, & \text{if } \tilde{a}_{i-1,j}^{(j)} = 0. \end{cases}$$

The multipliers satisfy that

$$m_{ij} = 0 \Rightarrow m_{hj} = 0 \quad \forall h > i.$$

Nonsingular TP matrices can be expressed as a product of nonnegative bidiagonal matrices. The following theorem (see Theorem 4.2 and p. 120 of [22]) introduces this representation, which is called the *bidiagonal decomposition*.

Theorem 1 (Cf. Theorem 4.2 of [22]). *Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a nonsingular TP matrix. Then A admits the following representation:*

$$A = F_{n-1}F_{n-2} \cdots F_1 D G_1 \cdots G_{n-2}G_{n-1}, \tag{3}$$

where D is the diagonal matrix $\text{diag}(p_{11}, \dots, p_{nn})$ with positive diagonal entries and F_i, G_i are the nonnegative bidiagonal matrices given by

$$F_i = \begin{pmatrix} 1 & & & & & & & & \\ 0 & 1 & & & & & & & \\ & & \ddots & \ddots & & & & & \\ & & & 0 & 1 & & & & \\ & & & & m_{i+1,1} & 1 & & & \\ & & & & & & \ddots & & \\ & & & & & & & \ddots & \\ & & & & & & & & m_{n,n-i} & 1 \end{pmatrix}, \tag{4}$$

$$G_i = \begin{pmatrix} 1 & 0 & & & & & & & \\ & 1 & \ddots & & & & & & \\ & & \ddots & \ddots & & & & & \\ & & & 0 & 1 & & & & \\ & & & & \tilde{m}_{i+1,1} & 1 & & & \\ & & & & & & 1 & \ddots & \\ & & & & & & & \ddots & \tilde{m}_{n,n-i} & 1 \end{pmatrix}, \tag{5}$$

for all $i \in \{1, \dots, n - 1\}$. If, in addition, the entries m_{ij} and \tilde{m}_{ij} satisfy

$$\begin{aligned} m_{ij} = 0 &\Rightarrow m_{hj} = 0 \quad \forall h > i, \\ \tilde{m}_{ij} = 0 &\Rightarrow \tilde{m}_{hj} = 0 \quad \forall h > i, \end{aligned} \tag{6}$$

then the decomposition is unique.

In the bidiagonal decomposition given by (3), (4) and (5), the entries m_{ij} and p_{ii} are the multipliers and diagonal pivots, respectively, corresponding to the NE of A (see Theorem 4.2 of [22] and the comment below it) and the entries \tilde{m}_{ij} are the multipliers of the NE of A^T (see p. 116 of [22]). The following result shows that the bidiagonal decomposition also characterizes STP matrices.

Theorem 2 (Cf. Theorem 4.3 of [22]). *A nonsingular $n \times n$ matrix A is STP if and only if it can be factorized in the form (3) with D a diagonal matrix with positive diagonal entries, F_i, G_i given by (4) and (5), and the entries m_{ij} and \tilde{m}_{ij} are positive numbers. This factorization is unique.*

Let us recall that an algorithm can be performed with high relative accuracy if it only includes products, divisions, sums of numbers of the same sign and subtractions of initial data (cf. [1,14]). In [11,12], assuming that the bidiagonal decomposition of a nonsingular TP matrix A is known to HRA, Plamen Koev designed efficient algorithms for computing to HRA the eigenvalues, singular values and the inverse of A as well as the solution to linear systems of equations $Ax = b$ whenever b has a pattern of alternating signs.

In [11] the matrix notation $\mathcal{BD}(A)$ was introduced to represent the bidiagonal decomposition of a nonsingular TP matrix,

$$(\mathcal{BD}(A))_{ij} = \begin{cases} m_{ij}, & \text{if } i > j, \\ \tilde{m}_{ji}, & \text{if } i < j, \\ p_{ii}, & \text{if } i = j. \end{cases} \tag{7}$$

In general, more matrices can be written as a product of bidiagonal matrices following (3). Throughout this paper, we will use the notation $\mathcal{BD}(A)$ given by (7) to denote the bidiagonal decomposition (3)–(5) of a general matrix A .

Finally, let us introduce the following classical generalizations of Pascal matrices.

Definition 3 (See [15,18]). For a real number x , the generalized Pascal matrix of the first kind, $P_n[x]$, is defined as the $(n + 1) \times (n + 1)$ lower triangular matrix with 1's on the main diagonal and

$$(P_n[x])_{ij} := x^{i-j} \binom{i-1}{j-1}, \quad 1 \leq j \leq i \leq n+1$$

and the symmetric generalized Pascal $(n + 1) \times (n + 1)$ matrix $R_n[x]$ is given by

$$(R_n[x])_{ij} := x^{i+j-2} \binom{i+j-2}{j-1}, \quad 1 \leq i, j \leq n+1.$$

For $x, y \in \mathbb{R}$ we define the $(n + 1) \times (n + 1)$ matrix $R_n[x, y]$

$$(R_n[x, y])_{ij} := x^{i-1} y^{i-1} \binom{i+j-2}{j-1}, \quad 1 \leq i, j \leq n+1.$$

Observe that $R_n[x] = R_n[x, x]$, that $P_n[1]$ is the lower triangular Pascal matrix and that $R_n[1]$ is the symmetric Pascal matrix.

Definition 4 (See [16]). For $x, y \in \mathbb{R}$, the extended generalized Pascal matrix $\Phi_n[x, y]$ is defined as

$$(\Phi_n[x, y])_{ij} = x^{i-j}y^{i+j-2} \binom{i-1}{j-1}, \quad 1 \leq j \leq i \leq n+1$$

and the extended generalized symmetric Pascal matrix $\Psi_n[x, y]$ is given by

$$(\Psi_n[x, y])_{ij} = x^{i-j}y^{i+j-2} \binom{i+j-2}{j-1}, \quad 1 \leq i, j \leq n+1.$$

In the next section we are going to deduce the bidiagonal decomposition of more general classes of matrices. As a consequence, we can also obtain the bidiagonal decomposition of the matrices $P_n[x]$, $R_n[x, y]$, $\Phi_n[x, y]$ and $\Psi_n[x, y]$.

3. Bidiagonal decomposition of generalized Pascal matrices

3.1. Generalized triangular Pascal matrices

Let x and λ be two real numbers and let n be a nonnegative integer. We define the notation $x^{n|\lambda}$ as follows:

$$x^{n|\lambda} = \begin{cases} x(x+\lambda) \cdots (x+(n-1)\lambda), & \text{if } n > 0, \\ 1, & \text{if } n = 0. \end{cases} \tag{8}$$

In [17], the generalized lower triangular Pascal matrix $P_{n,\lambda}[x]$ is defined by

$$(P_{n,\lambda}[x])_{i,j} := x^{(i-j)|\lambda} \binom{i-1}{j-1}, \quad 1 \leq j \leq i \leq n+1, \tag{9}$$

where n is a natural number and λ and x are real numbers. Observe that the particular case $\lambda = 0$ leads to the generalized Pascal matrix of the first kind $P_{n,0}[x] = P_n[x]$. The following result provides the bidiagonal decomposition of the generalized Pascal matrix $P_{n,\lambda}[x]$.

Theorem 5. Given $x, \lambda \in \mathbb{R}$ and $n \in \mathbb{N}$, let $P_{n,\lambda}[x]$ be the $(n+1) \times (n+1)$ lower triangular matrix given by (9).

(i) If $x \neq k\lambda$ for $k = -n+1, \dots, 0, \dots, n-1$, then we have that

$$(\mathcal{BD}(P_{n,\lambda}[x]))_{ij} = \begin{cases} 1, & i = j, \\ x + (i-2j)\lambda, & i > j, \\ 0, & i < j. \end{cases} \tag{10}$$

(ii) If $x = k\lambda$ for some $k \in \{0, \dots, n-1\}$, then we have that

$$(\mathcal{BD}(P_{n,\lambda}[x]))_{ij} = \begin{cases} 1, & i = j, \\ x + (i-2j)\lambda, & i > j, j \leq k, \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

(iii) If $x = -k\lambda$ for some $k \in \{0, \dots, n-1\}$, then we have that

$$(\mathcal{BD}(P_{n,\lambda}[x]))_{ij} = \begin{cases} 1, & i = j, \\ x + (i-2j)\lambda, & 0 < i-j \leq k, \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

Proof. Let us first assume that $x \neq k\lambda$ for $k = -n+1, \dots, 0, \dots, n-1$. We are going to perform the first step of the Neville elimination of $A = (a_{ij})_{1 \leq i, j \leq n+1}$, where $a_{ij} := (P_{n,\lambda}[x])_{i,j}$ for $i, j = 1, \dots, n+1$:

$$a_{ij}^{(2)} = a_{ij} - \frac{a_{i1}}{a_{i-1,1}} a_{i-1,j} = a_{ij} - (x + (i-2)\lambda) a_{i-1,j}, \quad i > j \geq 1.$$

Applying (9) to the previous formula, $a_{ij}^{(2)}$ can be written as

$$a_{ij}^{(2)} = x^{(i-j)|\lambda} \binom{i-1}{j-1} - (x + (i-2)\lambda) x^{(i-j-1)|\lambda} \binom{i-2}{j-1}.$$

By formula (8), we have that

$$\begin{aligned}
 a_{ij}^{(2)} &= \left((x + (i - j - 1)\lambda) \binom{i - 1}{j - 1} - (x + (i - 2)\lambda) \binom{i - 2}{j - 1} \right) x^{(i-j-1)\lambda} \\
 &= \left(x \binom{i - 2}{j - 2} + \frac{(i - j - 1)(i - 1)!}{(j - 1)!(i - j)!} \lambda - \frac{(i - 2)(i - 2)!}{(j - 1)!(i - j - 1)!} \lambda \right) x^{(i-j-1)\lambda}.
 \end{aligned}$$

After some computations we deduce that

$$a_{ij}^{(2)} = \left(x \binom{i - 2}{j - 2} - \lambda \binom{i - 2}{j - 2} \right) x^{(i-j-1)\lambda} = \binom{i - 2}{j - 2} (x - \lambda)^{(i-j)\lambda}.$$

We can observe that $a_{ij}^{(2)} = (P_{n,\lambda}[x])_{ij}^{(2)} = (P_{n,\lambda}[x - \lambda])_{i-1,j-1}$ for $i > j \geq 2$ and, hence, $(P_{n,\lambda}[x])^{(2)}[2, \dots, n + 1] = (P_{n,\lambda}[x - \lambda])[1, \dots, n]$. Then we can deduce that $(P_{n,\lambda}[x])_{ij}^{(k+1)} = (P_{n,\lambda}[x - k\lambda])_{i-k,j-k}$ for $i > j \geq k + 1$ and that the multipliers for the k th step of the NE of $P_{n,\lambda}[x]$ will be given by $x - (k - 1)\lambda + (i - k - 1)\lambda$ for $i = k + 1, \dots, n + 1$, and so, we conclude that (10) holds.

Let us now assume that $x = k\lambda$ for any $k \in \{0, \dots, n - 1\}$. Following the above proof we can see that $(P_{n,\lambda}[x])_{ij}^{(k+1)} = (P_{n,\lambda}[0])_{i-k,j-k}$ and the NE finishes at the $k + 1$ step. Hence, (ii) holds.

Finally, if $x = -k\lambda$ for any $k \in \{0, \dots, n - 1\}$, then $x^{(i-j)\lambda} = 0$ for $i - j > k$. Then the $n - k$ lower subdiagonals are already zero and the associated multipliers are also zero since the elimination procedure is not carried out on those entries. So, we conclude that (iii) holds.

Remark 6. It can be checked that the computational cost for the bidiagonal decomposition in (10) is of $\mathcal{O}(n^2)$ elementary operations. For the bidiagonal decompositions in (11) and (12) the computational costs are of $\mathcal{O}(k^2)$ and of $\mathcal{O}(k \cdot n)$ elementary operations, respectively.

The following corollary characterizes the matrices $P_{n,\lambda}[x]$ that are TP.

Corollary 7. Let $P_{n,\lambda}[x]$ be the lower triangular matrix given by (9) with $x, \lambda \in \mathbb{R}$ and with $n \in \mathbb{N}$. Then $P_{n,\lambda}[x]$ is a TP matrix if and only if one of the following conditions holds:

- (i) $x \geq (n - 1)|\lambda|$.
- (ii) $x = k|\lambda|$ for $k = 0, \dots, n - 1$.

Proof. By Theorem 5 we know that $P_{n,\lambda}[x]$ admits a factorization as a product of bidiagonal matrices. If (i) or (ii) holds, then all the bidiagonal matrices are nonnegative and so TP, and hence, its product is also TP (see for example Theorem 3.1 of [8]). Conversely, if $P_{n,\lambda}[x]$ is TP, since it is also nonsingular, it admits a unique bidiagonal decomposition by Theorem 1. Moreover, this bidiagonal decomposition will be given by Theorem 5 and the m_{ij} 's will be nonnegative. Hence, either (i) or (ii) holds.

The previous definition of $P_{n,\lambda}[x]$ is generalized in [17] for two variables x, y as follows:

$$(P_{n,\lambda}[x, y])_{i,j} := x^{(i-j)\lambda} y^{(j-1)\lambda} \binom{i - 1}{j - 1}. \tag{13}$$

Let us also define $P_n[x, y] := P_{n,0}[x, y]$. It is straightforward to see that the matrix $P_{n,\lambda}[x, y]$ can be expressed as the product of $P_{n,\lambda}[x]$ and a diagonal matrix:

$$P_{n,\lambda}[x, y] = P_{n,\lambda}[x] \text{diag}(1, y^{1\lambda}, \dots, y^{n\lambda}). \tag{14}$$

In [17] a further generalization of $P_{n,\lambda}[x, y]$ is given in terms of an arbitrary sequence $\mathbf{a} = \{a_n\}_{n \geq 0}$

$$(P_{n,\lambda}[x, y, \mathbf{a}])_{i,j} := a_{j-1} x^{(i-j)\lambda} y^{(j-1)\lambda} \binom{i - 1}{j - 1}, \tag{15}$$

and so we also derive

$$P_{n,\lambda}[x, y, \mathbf{a}] = P_{n,\lambda}[x] \text{diag}(a_0, a_1 y^{1\lambda}, \dots, a_n y^{n\lambda}). \tag{16}$$

Observe that the matrix $P_{n,\lambda}[x, y] = P_{n,\lambda}[x, y, \mathbf{1}]$, where $\mathbf{1}$ is the sequence formed by 1's. By (16) and Theorem 5, we can deduce the bidiagonal decomposition of the matrix $\mathcal{BD}(P_{n,\lambda}[x, y, \mathbf{a}])$. For example, if $x \neq k\lambda$ for $k = -n + 1, \dots, 0, \dots, n - 1$, its bidiagonal decomposition is given by

$$(\mathcal{BD}(P_{n,\lambda}[x, y, \mathbf{a}]))_{ij} = \begin{cases} a_{j-1} y^{(j-1)\lambda}, & i = j, \\ x + (i - 2j)\lambda, & i > j, \\ 0, & i < j. \end{cases} \tag{17}$$

3.2. Lattice path matrices

Let $Lp_n(\alpha, \beta, \gamma) = (k_{ij})_{1 \leq i, j \leq n+1}$ be the $(n+1) \times (n+1)$ lattice path matrix such that its entries are given by the recurrence relation

$$\alpha k_{i-1, j} + \beta k_{i-1, j} + \gamma k_{i-1, j-1} = k_{ij}, \quad 2 \leq i, j \leq n+1, \tag{18}$$

with $k_{ij} = \alpha^{j-1}$ for $j \in \{1, \dots, n+1\}$ and $k_{i1} = \beta^{i-1}$ for $i \in \{1, \dots, n+1\}$. These matrices were considered in [18]. Other related classes of matrices were considered in [23], where it was also shown that some of those matrices are TP. In Theorem 2.3 of [18] it is also shown that $Lp_n(\alpha, \beta, \gamma)$ admits the following factorization

$$Lp_n(\alpha, \beta, \gamma) = P_n[\alpha] D_{\alpha\beta+\gamma}^n (P_n[\beta])^T, \tag{19}$$

where $D_{\alpha\beta+\gamma}^n = \text{diag}(1, \alpha\beta + \gamma, \dots, (\alpha\beta + \gamma)^n)$ and $P_n[\delta] = P_{n,0}[\delta]$. Observe that the matrix $Lp_n(\alpha, \beta, \gamma)$ is nonsingular if and only if $\alpha\beta + \gamma \neq 0$. In the following result, we deduce the bidiagonal decomposition of $Lp_n(\alpha, \beta, \gamma)$.

Theorem 8. Let $Lp_n(\alpha, \beta, \gamma) = (k_{ij})_{1 \leq i, j \leq n+1}$ be the matrix whose entries are defined by (18) with $\alpha\beta + \gamma \neq 0$. Then its bidiagonal decomposition is given by

$$(\mathcal{BD}(Lp_n(\alpha, \beta, \gamma)))_{ij} = \begin{cases} (\alpha\beta + \gamma)^{i-1}, & \text{if } i = j, \\ \alpha, & \text{if } i > j, \\ \beta, & \text{if } i < j. \end{cases} \tag{20}$$

Proof. By (19), the matrix $Lp_n(\alpha, \beta, \gamma)$ can be decomposed as the product of a lower triangular matrix, a diagonal matrix and an upper triangular matrix. Hence, we can deduce its bidiagonal decomposition from the bidiagonal decomposition of these three factors. Since $P_n[\alpha] = P_{n,0}[\alpha]$, by Theorem 5 we have that

$$(\mathcal{BD}(P_n[\alpha]))_{ij} = \begin{cases} 1, & \text{if } i = j, \\ \alpha, & \text{if } i > j, \\ 0, & \text{if } i < j. \end{cases}$$

Analogously, the bidiagonal decomposition of $(P_n[\beta])^T$ is given by

$$(\mathcal{BD}((P_n[\beta])^T))_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i > j, \\ \beta, & \text{if } i < j. \end{cases}$$

Therefore, we conclude that

$$Lp_n(\alpha, \beta, \gamma) = P_n[\alpha] D_{\alpha\beta+\gamma}^n (P_n[\beta])^T = \bar{F}_n \cdots \bar{F}_1 D_{\alpha\beta+\gamma}^n \bar{G}_1 \cdots \bar{G}_n, \tag{21}$$

where \bar{F}_k is the lower bidiagonal matrix given by (4) with all multipliers equal to α and \bar{G}_k is the upper bidiagonal matrix given by (5) with all multipliers equal to β . So, (20) holds.

The following corollary considers a case where $Lp_n(\alpha, \beta, \gamma)$ is STP and shows that its bidiagonal decomposition can be computed to HRA.

Corollary 9. Let $Lp_n(\alpha, \beta, \gamma) = (k_{ij})_{1 \leq i, j \leq n+1}$ be the matrix whose entries are defined by (18). If $\alpha, \beta > 0$ and $\alpha\beta + \gamma > 0$, then $Lp_n(\alpha, \beta, \gamma) = (k_{ij})_{1 \leq i, j \leq n+1}$ is an STP matrix. Moreover, if $\gamma \geq 0$, then its bidiagonal decomposition (20) can be computed to HRA and it can be used to obtain the eigenvalues, singular values and the inverse of $Lp_n(\alpha, \beta, \gamma)$ with HRA as well as the solution of the linear systems $Lp_n(\alpha, \beta, \gamma)x = b$, where $b = (b_1, \dots, b_{n+1})$ has alternating signs.

Proof. By Theorem 2, $Lp_n(\alpha, \beta, \gamma) = (k_{ij})_{1 \leq i, j \leq n+1}$ is an STP matrix. With the additional condition $\gamma \geq 0$, $\mathcal{BD}(Lp_n(\alpha, \beta, \gamma))$ can be computed with a subtraction-free algorithm, and hence, with HRA, which in turn guarantees that the algebraic computations stated in the statement of this corollary can be performed with HRA (see Section 5 or Section 3 of [12]).

Remark 10. In order to compute the bidiagonal decomposition (20), $n + 1$ elementary operations are necessary, that is, a computational cost of $\mathcal{O}(n)$ elementary operations.

The class of lattice path matrices, $Lp_n(\alpha, \beta, \gamma)$, contains the generalizations of Pascal matrices given by definitions 3 and 4, and so, from their bidiagonal decomposition we can deduce the bidiagonal decomposition of these matrices. In particular, in Theorem 3.1 of [18] the following relationship was proved:

$$Lp_n(\alpha, \beta, \gamma) = \begin{cases} P_n[x, y], & \text{if } \alpha = 0, \beta = y, \gamma = x, \\ R_n[x, y], & \text{if } \alpha = x, \beta = y, \gamma = 0, \\ \Phi_n[x, y], & \text{if } \alpha = 0, \beta = xy, \gamma = y^2, \\ \Psi_n[x, y], & \text{if } \alpha = y/x, \beta = xy, \gamma = 0. \end{cases} \tag{22}$$

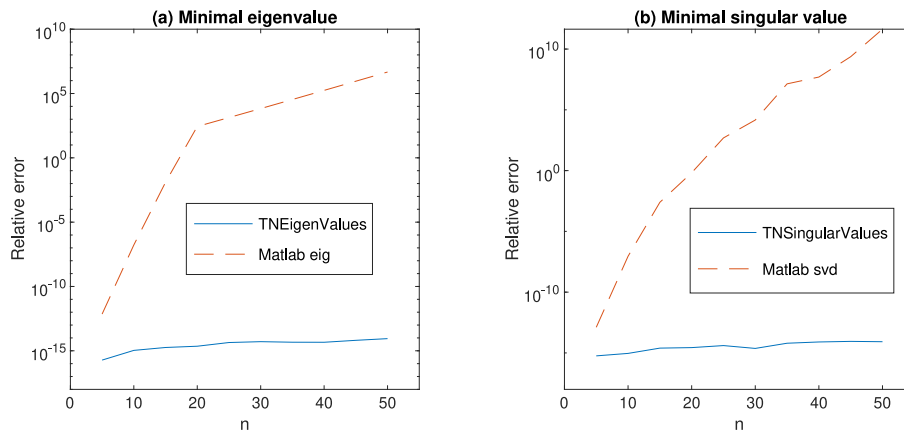


Fig. 1. Relative error for the minimal eigenvalues and singular values of $Lp_n(\sqrt{2}, \sqrt{3}, \sqrt{5})$.

Taking into account (22), we can use Theorem 8 to obtain their bidiagonal decomposition. We can also apply Corollary 9 to study the cases when they are STP and when their bidiagonal decomposition can be obtained with HRA.

Corollary 11.

- (i) If $x, y > 0$, then $R_n[x, y]$ is STP and its bidiagonal decomposition can be computed to HRA.
- (ii) If $xy > 0$, then $\Psi_n[x, y]$ is STP and its bidiagonal decomposition can be computed to HRA.

4. Numerical experiments

In [11,12], assuming that the parameterization $\mathcal{BD}(A)$ of a nonsingular TP matrix A is known, Plamen Koev presented algorithms to solve the following algebraic problems for A : computation of the eigenvalues and the singular values of A , computation of A^{-1} and solution of the systems of linear equations $Ax = b$. In [24] Marco and Martínez presented another algorithm for the computation of A^{-1} from $\mathcal{BD}(A)$. If, in addition, $\mathcal{BD}(A)$ is known to HRA, then the algorithms solve these algebraic problems to HRA (in the case of linear systems only when b has a pattern of alternating signs). Koev implemented the corresponding algorithms for Matlab and Octave, which are available in the software library *TNTool* in [13]. The functions are `TNEigenValues` for the eigenvalues, `TNSingularValues` for the singular values, `TNInverseExpand` for the inverse and `TNSolve` for the solution of linear system of equations. The functions require as input argument the data determining the bidiagonal decomposition (3)–(5) of A , or equivalently, $\mathcal{BD}(A)$ given by (7), and, in the case of `TNSolve`, in addition, the vector b .

Remark 12. The computational cost for both `TNSolve` and `TNInverseExpand` is $\mathcal{O}(n^2)$ elementary operations (see [12] and Section 4 of [24]) and for the other two functions, `TNEigenValues` and `TNSingularValues`, is $\mathcal{O}(n^3)$ elementary operations. Hence, taking into account Remarks 6 and 10, the total computational cost for solving a linear system or computing the inverse with the matrices corresponding to these bidiagonal computations is $\mathcal{O}(n^2)$ elementary operations, and so we have fast algorithms, whereas the total computational cost of computing the eigenvalues or the singular values is $\mathcal{O}(n^3)$ elementary operations.

4.1. HRA computations with lattice path matrices

If $\alpha, \beta > 0$ and $\gamma \geq 0$, by Corollary 9, the matrices $Lp_n(\alpha, \beta, \gamma)$ are STP and their bidiagonal decompositions can be computed to HRA, and so the algebraic computations mentioned before can also be performed to HRA.

Let us consider the matrices $Lp_n(\sqrt{2}, \sqrt{3}, \sqrt{5})$ for $n = 5, 10, \dots, 50$. First, we have computed the eigenvalues and the singular values of these matrices with Mathematica using a precision of 100 digits. We have also computed approximations to the eigenvalues of those matrices in Matlab with `eig` and also with `TNEigenValues` using the bidiagonal decomposition provided by (20). Then we have computed the relative errors of the approximations obtained considering the eigenvalues obtained with Mathematica as exact computations.

In Fig. 1(a) we can see the relative error for the minimal eigenvalue of each matrix $Lp_n(\sqrt{2}, \sqrt{3}, \sqrt{5})$, $n = 5, 10, \dots, 50$, for both `eig` and `TNEigenValues`. We can observe that Matlab function `eig` does not provide an acceptable approximation of the minimal eigenvalue of the matrices $Lp_n(\sqrt{2}, \sqrt{3}, \sqrt{5})$ for $n \geq 15$ in contrast to the accurate approximations provided by the HRA computations of `TNEigenValues`.

We have also computed approximations to the singular values of the matrices $Lp_n(\sqrt{2}, \sqrt{3}, \sqrt{5})$, $n = 5, 10, \dots, 50$, in Matlab with `svd` and also with `TNSingularValues`. Then we have computed the relative errors of the approximations

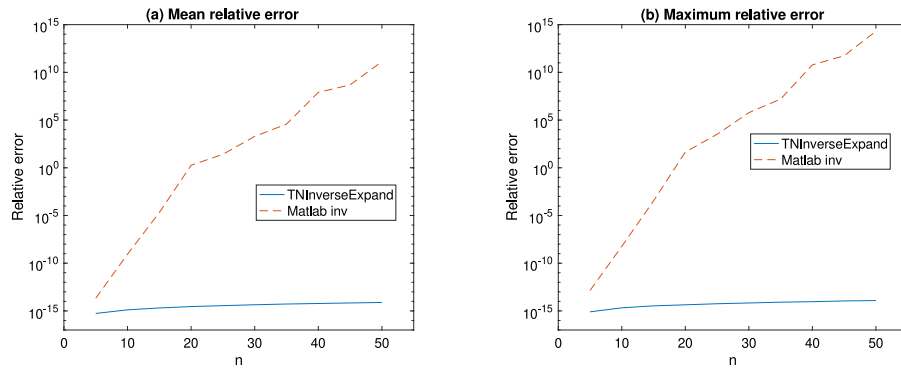


Fig. 2. Relative errors for $Lp_n(\sqrt{2}, \sqrt{3}, \sqrt{5})^{-1}$, $n = 5, 10, \dots, 50$.

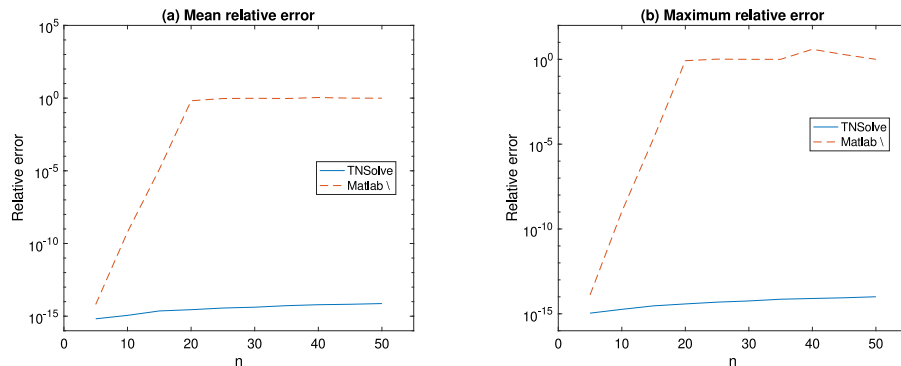


Fig. 3. Relative errors for the systems $Lp_n(\sqrt{2}, \sqrt{3}, \sqrt{5})x = b_n$, $n = 5, 10, \dots, 50$.

obtained considering the singular values obtained with Mathematica as exact computations. In Fig. 1(b) we can see the relative error for the minimal singular value of each matrix $Lp_n(\sqrt{2}, \sqrt{3}, \sqrt{5})$ for both `svd` and `TNSingularValues`. As in the case of the eigenvalues, `TNSingularValues` provides very accurate approximations to the minimal singular values in contrast to the poor results provided by `svd`.

We have also computed with Matlab approximations to the inverses of the matrices $Lp_n(\sqrt{2}, \sqrt{3}, \sqrt{5})$, $n = 5, 10, \dots, 50$, with `inv` and with `TNInverseExpand` using the bidiagonal decomposition given by (20). The inverses of these matrices have been computed with Mathematica using a precision of 100 digits. Then we have computed the corresponding componentwise relative errors. Finally we have obtained the mean and maximum componentwise relative errors. Fig. 2(a) shows the mean relative error and (b) shows the maximum relative error. We can also observe in this case that the results obtained with `TNInverseExpand` are much more accurate than those obtained with `inv`. In fact, the approximations obtained with `inv` are not acceptable for $n > 15$.

Now we consider the linear systems $Lp_n(\sqrt{2}, \sqrt{3}, \sqrt{5})x = b_n$ for $n = 5, 10, \dots, 50$, where $b_n \in \mathbb{R}^n$ has the absolute value of its entries randomly generated as integers in the interval $[1, 1000]$, but with alternating signs. We have computed approximations to the solution x of the linear systems with Matlab, the first one using `TNSolve` and the bidiagonal decomposition given by (20), and the second one using the Matlab command `A\b`. By using Mathematica with a precision of 100 digits we have computed the solution of the systems and then we have computed the componentwise relative errors for the two approximations obtained with Matlab. Then we have obtained the mean and maximum componentwise relative error. Fig. 3(a) shows the mean relative error and (b) shows the maximum relative error. Again, the results obtained with HRA algorithms are very accurate in contrast to the results obtained with the usual Matlab command.

We also consider the linear systems $Lp_n(\sqrt{2}, \sqrt{3}, \sqrt{5})x = \hat{b}_n$ for $n = 5, 10, \dots, 50$, where now $\hat{b}_n \in \mathbb{R}^n$ has its entries randomly generated as integers in the interval $[-1000, 1000]$. Fig. 4(a) shows the mean relative error and (b) shows the maximum relative error. In this case, we cannot guarantee that the solution of the linear systems provided by `TNSolve` can be computed to HRA. However, the results obtained with `TNSolve` are very accurate in contrast to the results obtained with the usual Matlab command.

4.2. Accurate computations for generalized triangular Pascal matrices

Let us consider the lower triangular matrices $P_{n,1}[3/2]$ for $n = 5, 10, \dots, 50$, given by (9) with $\chi = 3/2$ and $\lambda = 1$. Unfortunately, by Corollary 7, these matrices are not TP and we cannot assure that their bidiagonal decomposition can be computed to HRA. So we cannot guarantee that the algebraic computations mentioned above can be performed to

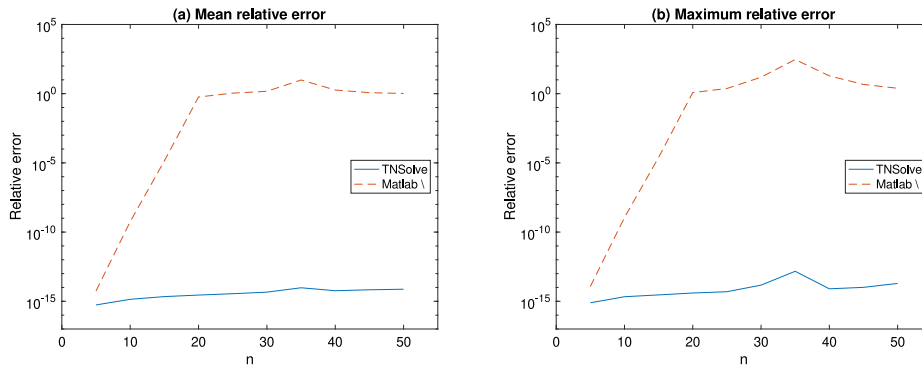


Fig. 4. Relative errors for the systems $Lp_n(\sqrt{2}, \sqrt{3}, \sqrt{5})x = \tilde{b}_n$, $n = 5, 10, \dots, 50$.

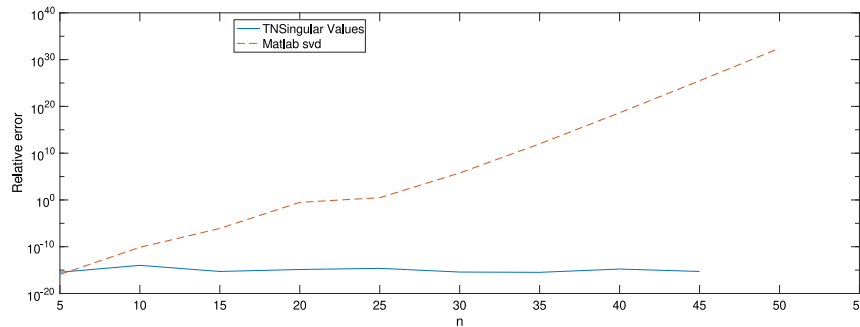


Fig. 5. Relative error for the minimal singular values of $P_{n,1}[3/2]$.

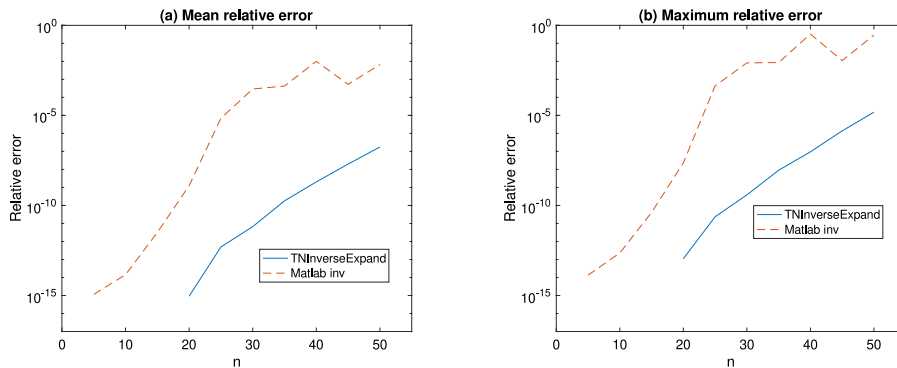


Fig. 6. Relative errors for $P_{n,1}[3/2]^{-1}$, $n = 5, 10, \dots, 50$.

HRA neither. Nevertheless, let us also compare the numerical accuracy of TNSingularValues, TNInverseExpand and TNSolve versus the usual Matlab commands svd, inv and \, respectively.

First, we have computed the singular values of these matrices with Mathematica using a precision of 100 digits. We have also computed approximations to the singular values of the matrices $P_{n,1}[3/2]$ with Matlab function svd and also with TNSingularValues and the corresponding $\mathcal{BD}(P_{n,1}[3/2])$ given in Theorem 5. Then we have computed the relative errors of the approximations obtained considering the singular values obtained with Mathematica as exact computations. In Fig. 5 we can see the relative error for the minimal singular value of each matrix for both svd and TNSingularValues.

We have also computed with Matlab approximations to $P_{n,1}[3/2]^{-1}$, $n = 5, 10, \dots, 50$, with inv and with TNInverseExpand using $\mathcal{BD}(P_{n,1}[3/2])$. With Mathematica we have computed the inverse of these matrices with exact arithmetic. Then we have computed the corresponding componentwise relative errors. Finally we have obtained the mean and maximum componentwise relative error. Fig. 6(a) shows the mean relative error and (b) shows the maximum relative error. We can also observe in this case that the results obtained with TNInverseExpand are much more accurate than those obtained with inv.

Finally we consider the linear systems $P_{n,1}[3/2]x = b_n$, $n = 5, 10, \dots, 50$, where $b_n \in \mathbb{R}^n$ has its entries randomly generated as integers in the interval $[-1000, 1000]$. We have computed approximations to the solution x of the linear system with Matlab, the first one using TNSolve and $\mathcal{BD}(P_{n,1}[3/2])$, and the second one using the Matlab command A\b.

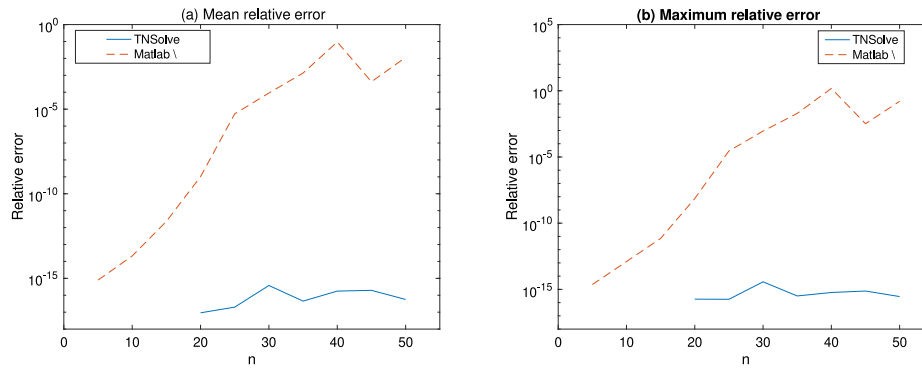


Fig. 7. Relative errors for the systems $P_{n,1}[3/2]x = b_n$, $n = 5, 10, \dots, 50$.

By using Mathematica with exact arithmetic we have computed the exact solution of the systems and then we have computed the componentwise relative errors for the two approximations obtained with Matlab. Fig. 7(a) shows the mean relative error and (b) shows the maximum relative error. Again, the results obtained with TNSolve are very accurate in contrast to the results obtained with the usual Matlab command.

5. Conclusions

Pascal matrices and some generalizations considered in this paper arise in many applications, as commented in the introduction. It is well known that Pascal matrices and their generalizations are ill-conditioned (see [20]). However, we show in this paper that we can compute with HRA all their eigenvalues and all their singular values, and also the inverses of these matrices as well as the solutions of some linear systems. In fact, our numerical experiments show that we can considerably improve the accuracy obtained with the usual Matlab commands. The crucial tool has been to obtain the bidiagonal decomposition of the generalized Pascal matrices with HRA and then apply the corresponding algorithms given in [11–13]. Let us also remark that, in spite of its much greater accuracy, the procedure presented in this paper has a computational cost similar to the usual algorithms used to solve these problems.

References

- [1] J. Demmel, P. Koev, The accurate and efficient solution of a totally positive generalized Vandermonde linear system, *SIAM J. Matrix Anal. Appl.* 27 (2005) 142–152.
- [2] A. Marco, J.J. Martínez, Accurate computations with Said-Ball-Vandermonde matrices, *Linear Algebra Appl.* 432 (2010) 2894–2908.
- [3] A. Marco, J.J. Martínez, Accurate computations with totally positive Bernstein-Vandermonde matrices, *Electron. J. Linear Algebra* 26 (2013) 357–380.
- [4] J. Delgado, J.M. Peña, Fast and accurate algorithms for Jacobi-Stirling matrices, *Appl. Math. Comput.* 236 (2014) 253–259.
- [5] J. Delgado, J.M. Peña, Accurate computations with collocation matrices of q -Bernstein polynomials, *SIAM J. Matrix Anal. Appl.* 36 (2015) 880–893.
- [6] J. Delgado, H. Orera, J.M. Peña, Accurate computations with Laguerre matrices, *Numer. Linear Algebra Appl.* 26 (2019) e2217, 10 pp.
- [7] J. Delgado, H. Orera, J.M. Peña, Accurate algorithms for Bessel matrices, *J. Sci. Comput.* 80 (2019) 1264–1278.
- [8] T. Ando, Totally positive matrices, *Linear Algebra Appl.* 90 (1987) 165–219.
- [9] M. Gasca, C.A. Micchelli (Eds.), Total positivity and its applications, in: *Mathematics and its Applications*, Kluwer Acad. Publ., Dordrecht, 1996.
- [10] A. Pinkus, Totally positive matrices, in: *Tracts in Mathematics*, vol. 181, Cambridge University Press, Cambridge, UK, 2010.
- [11] P. Koev, Accurate eigenvalues and SVDs of totally nonnegative matrices, *SIAM J. Matrix Anal. Appl.* 27 (2005) 1–23.
- [12] P. Koev, Accurate computations with totally nonnegative matrices, *SIAM J. Matrix Anal. Appl.* 29 (2007) 731–751.
- [13] P. Koev, 2020, <http://www.math.sjsu.edu/~koev/software/TNTTool.html>. (Accessed 11 February 2020).
- [14] J. Demmel, I. Dumitriu, O. Holtz, P. Koev, Accurate and efficient expression evaluation and linear algebra, *Acta Numer.* 17 (2008) 87–145.
- [15] Z. Zhang, The linear algebra of the generalized Pascal matrix, *Linear Algebra Appl.* 250 (1997) 51–60.
- [16] Z. Zhang, M. Liu, An extension of the generalized Pascal matrix and its algebraic properties, *Linear Algebra Appl.* 271 (1998) 169–177.
- [17] M. Bayat, H. Teimoori, The linear algebra of the generalized Pascal functional matrix, *Linear Algebra Appl.* 295 (1999) 81–89.
- [18] I.-P. Kim, LDU decomposition of an extension matrix of the Pascal matrix, *Linear Algebra Appl.* 434 (2011) 2187–2196.
- [19] X.-G. Lv, T.-Z. Huang, Z.-G. Ren, A new algorithm for linear systems of the Pascal type, *J. Comput. Appl. Math.* 225 (2009) 309–315.
- [20] P. Alonso, J. Delgado, R. Gallego, J.M. Peña, Conditioning and accurate computations with Pascal matrices, *J. Comput. Appl. Math.* 252 (2013) 21–26.
- [21] M. Gasca, J.M. Peña, Total positivity and Neville Elimination, *Linear Algebra Appl.* 165 (1992) 25–44.
- [22] M. Gasca, J.M. Peña, On factorizations of totally positive matrices, in: M. Gasca, C.A. Micchelli (Eds.), *Total Positivity and its Applications*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996, pp. 109–130.
- [23] F. Brenti, Combinatorics and total positivity, *J. Combin. Theory Ser. A* 71 (1995) 175–218.
- [24] A. Marco, J.J. Martínez, Accurate computation of the Moore-Penrose inverse of strictly totally positive matrices, *J. Comput. Appl. Math.* 350 (2019) 299–308.

Article 7

- [20] J. Delgado, H. Orera and J. M. Peña. High relative accuracy with matrices of q -integers. Numer. Linear Algebra Appl. 28 (2021), Paper No. e2383, 20 pp.

High relative accuracy with matrices of q -integers

Jorge Delgado¹  | Héctor Orera¹ | Juan M. Peña¹

¹Departamento de Matemática Aplicada, Universidad de Zaragoza, Zaragoza, Spain

Correspondence

Jorge Delgado, Departamento de Matemática Aplicada, Universidad de Zaragoza, Escuela Universitaria Politécnica de Teruel, Teruel 44071, Spain.
Email: jorgedel@unizar.es

Funding information

Gobierno de Aragón, Grant/Award Number: E41_17R; MCIU/AEI, Grant/Award Number: PGC2018-096321-B-I00

Abstract

This article shows that the bidiagonal decomposition of many important matrices of q -integers can be constructed to high relative accuracy (HRA). This fact can be used to compute with HRA the eigenvalues, singular values, and inverses of these matrices. These results can be applied to collocation matrices of q -Laguerre polynomials, q -Pascal matrices, and matrices formed by q -Stirling numbers. Numerical examples illustrate the theoretical results.

KEYWORDS

bidiagonal decomposition, high relative accuracy, q -integers, quantum calculus, quantum orthogonal polynomials, total positivity

1 | INTRODUCTION

Quantum calculus (see Reference 1) uses q -integers, q -binomial coefficients (see Section 3), and other q -analogues of classical calculus. In particular, it has led to the use of matrices of q -integers. This article shows that many algebraic computations (eigenvalues, singular values, and inverses) with these matrices can be performed with high relative accuracy (HRA). An important source of these matrices comes from classical matrices very useful in Combinatorics, such as Pascal matrices or Jacobi–Stirling matrices (see Reference 2). For classical Pascal matrices, it is known how to perform accurately the algebraic computations mentioned before (see References 3,4). In this article, we first consider the accurate computations with q -Pascal matrices. Then we consider matrices formed with q -Stirling numbers (see Reference 5) of the first and second kind. In this article, we also guarantee the accurate computation for the collocation matrices of some systems of functions. Let us recall that in Reference 6 this goal was achieved for collocation matrices of q -Bernstein polynomials, in Reference 7 for collocation matrices of h -Bernstein basis and in Reference 8 for collocation matrices of Laguerre polynomials. In fact, another field where quantum calculus has played an important role is that of orthogonal polynomials, where quantum orthogonal polynomials have been considered (see Reference 9). Here, we also guarantee HRA for the mentioned algebraic computations with collocation matrices of q -Laguerre polynomials.

In all cases considered in this article, a key tool has been to prove the total positivity of the matrices. Let us recall that a matrix is called *totally positive* (TP) (*strictly totally positive* [STP], respectively) if all its minors are nonnegative (positive, respectively). These matrices are also called in the literature totally nonnegative and totally positive, respectively. TP matrices arise in many fields such as approximation theory, statistics, economy, mechanics, computer-aided geometric design, or combinatorics (see References 10–12). Nonsingular TP matrices satisfy the remarkable property that they admit a bidiagonal decomposition. This bidiagonal decomposition is the start point for the algorithms of Reference 13 to carry out the mentioned algebraic computations with HRA. Let us also remark that the bidiagonal decomposition of the q -Pascal matrices obtained in this article is not (for $q \neq 1$) as simple as that of Pascal matrices shown in References 4,14, where all the entries are ones. Finally, in contrast to the HRA computation of the bidiagonal decomposition of the collocation matrices of generalized Laguerre polynomials (see Reference 8), its extension

to the bidiagonal decomposition of the collocation matrices of generalized q -Laguerre polynomials requires additional conditions.

The article is organized as follows. Section 2 presents basic notations and auxiliary results concerning the bidiagonal decomposition of nonsingular TP matrices. Section 3 is devoted to q -Pascal matrices. Section 4 provides the bidiagonal decompositions of matrices with q -Stirling numbers and it uses a general result on the bidiagonal decomposition of the inverse of a triangular TP matrix. Section 5 focuses on accurate computations with collocation matrices of q -Laguerre polynomials. Section 6 illustrates the theoretical results of the article with numerical experiments. They show the HRA of the calculation of the inverse, eigenvalues, or singular values of A or the solution of linear systems $Ax = b$ such that b has alternating signs. Finally, Section 7 summarizes the main conclusions of the article.

2 | AUXILIARY RESULTS

Let $D = (d_{ij})_{1 \leq i, j \leq n}$ be a diagonal matrix, which can be denoted by $D = \text{diag}(d_1, \dots, d_n)$, where $d_i = d_{ii}$ for $i = 1, \dots, n$. Using this notation the $n \times n$ identity matrix is defined as $I_n = \text{diag}(1, \dots, 1)$. Let us denote by $E_i(x)$, with $i = 2, \dots, n$, the $n \times n$ lower elementary bidiagonal matrix whose $(i, i - 1)$ entry is x :

$$E_i(x) = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & x & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix}. \quad (1)$$

The matrix $E_i^T(x) = (E_i(x))^T$ is called upper elementary bidiagonal matrix. The matrices $E_k(x)$ satisfy the property

$$E_i(x)E_j(y) = E_j(y)E_i(x), \quad (2)$$

unless $|i - j| = 1$ with $xy \neq 0$.

Neville elimination (NE) is an alternative procedure to Gaussian elimination that produces zeros in a column of a matrix by adding to each row an appropriate multiple of the previous one. Given a nonsingular matrix $A = (a_{ij})_{1 \leq i, j \leq n}$, the NE procedure consists of $n - 1$ steps and leads to the following sequence of matrices:

$$A =: A^{(1)} \rightarrow \tilde{A}^{(1)} \rightarrow A^{(2)} \rightarrow \tilde{A}^{(2)} \rightarrow \dots \rightarrow A^{(n)} = \tilde{A}^{(n)} = U, \quad (3)$$

where U is an upper triangular matrix.

The matrix $\tilde{A}^{(k)} = (\tilde{a}_{ij}^{(k)})_{1 \leq i, j \leq n}$ is obtained from the matrix $A^{(k)} = (a_{ij}^{(k)})_{1 \leq i, j \leq n}$ by a row permutation that moves to the bottom the rows with a zero entry in column k below the main diagonal. For nonsingular TP matrices, it is always possible to perform NE without row exchanges (see Reference 15). If a row permutation is not necessary at the k th step, we have that $\tilde{A}^{(k)} = A^{(k)}$. The entries of $A^{(k+1)} = (a_{ij}^{(k+1)})_{1 \leq i, j \leq n}$ can be obtained from $\tilde{A}^{(k)} = (\tilde{a}_{ij}^{(k)})_{1 \leq i, j \leq n}$ using the formula:

$$a_{ij}^{(k+1)} = \begin{cases} \tilde{a}_{ij}^{(k)} - \frac{\tilde{a}_{ik}^{(k)}}{\tilde{a}_{i-1,k}^{(k)}} \tilde{a}_{i-1,j}^{(k)}, & \text{if } k \leq j < i \leq n \text{ and } \tilde{a}_{i-1,k}^{(k)} \neq 0, \\ \tilde{a}_{ij}^{(k)}, & \text{otherwise,} \end{cases} \quad (4)$$

for $k = 1, \dots, n - 1$. The (i, j) pivot of the NE of A is given by

$$p_{ij} = \tilde{a}_{ij}^{(j)}, \quad 1 \leq j \leq i \leq n.$$

If $i=j$ we say that p_{ii} is a *diagonal pivot*. The (i,j) *multiplier* of the NE of A , with $1 \leq j \leq i \leq n$, is defined as

$$m_{ij} = \begin{cases} \frac{\tilde{a}_{ij}^{(j)}}{\tilde{a}_{i-1,j}^{(j)}} = \frac{p_{ij}}{p_{i-1,j}}, & \text{if } \tilde{a}_{i-1,j}^{(j)} \neq 0, \\ 0, & \text{if } \tilde{a}_{i-1,j}^{(j)} = 0. \end{cases}$$

The multipliers satisfy that

$$m_{ij} = 0 \Rightarrow m_{hj} = 0 \quad \forall h > i.$$

Nonsingular TP matrices can be expressed as a product of nonnegative bidiagonal matrices. The following theorem (see theorem 4.2 and p. 120 of Reference 16) introduces this representation, which is called the *bidiagonal decomposition*.

Theorem 1 (cf. theorem 4.2 of Reference 16). *Let $A = (a_{ij})_{1 \leq i,j \leq n}$ be a nonsingular TP matrix. Then A admits the following representation:*

$$A = F_{n-1}F_{n-2} \cdots F_1 D G_1 \cdots G_{n-2}G_{n-1}, \tag{5}$$

where D is the diagonal matrix $\text{diag}(p_{11}, \dots, p_{nn})$ with positive diagonal entries and F_i, G_i are the nonnegative bidiagonal matrices given by

$$F_i = \begin{pmatrix} 1 & & & & & & \\ & 0 & 1 & & & & \\ & & \ddots & \ddots & & & \\ & & & 0 & 1 & & \\ & & & & m_{i+1,1} & 1 & \\ & & & & & \ddots & \ddots \\ & & & & & & m_{n,n-i} & 1 \end{pmatrix}, \tag{6}$$

$$G_i = \begin{pmatrix} 1 & 0 & & & & & \\ & 1 & \ddots & & & & \\ & & \ddots & 0 & & & \\ & & & 1 & \tilde{m}_{i+1,1} & & \\ & & & & 1 & \ddots & \\ & & & & & \ddots & \tilde{m}_{n,n-i} \\ & & & & & & 1 \end{pmatrix}, \tag{7}$$

for all $i \in \{1, \dots, n-1\}$. If, in addition, the entries m_{ij} and \tilde{m}_{ij} satisfy

$$\begin{aligned} m_{ij} = 0 &\Rightarrow m_{hj} = 0 \quad \forall h > i, \\ \tilde{m}_{ij} = 0 &\Rightarrow \tilde{m}_{hj} = 0 \quad \forall h > i, \end{aligned} \tag{8}$$

then the decomposition is unique.

In the bidiagonal decomposition given by (5)–(7), the entries m_{ij} and p_{ii} are the multipliers and diagonal pivots, respectively, corresponding to the NE of A (see theorem 4.2 of Reference 16 and the comment below it) and the entries \tilde{m}_{ij} are the multipliers of the NE of A^T (see p. 116 of Reference 16). The following result shows that the bidiagonal decomposition also characterizes STP matrices.

Theorem 2 (cf. theorem 4.3 of Reference 16). *A nonsingular $n \times n$ matrix A is STP if and only if it can be factorized in the form (5) with D a diagonal matrix with positive diagonal entries, F_i, G_i given by (6) and (7), and the entries m_{ij} and \tilde{m}_{ij} positive numbers. This factorization is unique.*

The bidiagonal decomposition can be used to represent more classes of matrices. The following remark shows which hypotheses of Theorem 1 are sufficient for the uniqueness of a factorization following (5).

Remark 1. If we consider the factorization given by (5)–(8) without any further requirement than the nonsingularity of D , by proposition 2.2 of Reference 17 the uniqueness of (5) holds.

In Reference 3, the matrix notation $BD(A)$ was introduced to represent the bidiagonal decomposition of a nonsingular TP matrix,

$$(BD(A))_{ij} = \begin{cases} m_{ij}, & \text{if } i > j, \\ \tilde{m}_{ji}, & \text{if } i < j, \\ p_{ii}, & \text{if } i = j. \end{cases} \tag{9}$$

Throughout this article, $BD(A)$ will denote the bidiagonal decomposition of a matrix that satisfies the hypotheses of Remark 1. The following remark gives the relationship between the bidiagonal decompositions of a matrix and of its transpose.

Remark 2. If A is a TP matrix, then A^T is also TP. Transposing formula (5) of Theorem 1 we obtain the unique bidiagonal decomposition of A^T :

$$A^T = G_{n-1}^T \cdots G_1^T D F_1^T \cdots F_{n-1}^T,$$

where F_i and G_i , $i \in \{1, \dots, n-1\}$, are the bidiagonal lower and upper triangular nonnegative matrices given in (6) and (7), respectively. It can also be checked that

$$BD(A^T) = BD(A)^T.$$

An algorithm can be performed with HRA if it does not include subtractions (except for the initial data), that is, if it only includes products, divisions, sums of numbers of the same sign, subtractions of numbers of opposite sign and subtractions of the initial data (cf. References 3,18). In particular, a subtraction-free algorithm provides results with HRA. In Reference 3, assuming that the parameters of $BD(A)$ are known with HRA, Koev presented algorithms for computing the eigenvalues of the matrix A , the singular values of the matrix A , the inverse of the matrix A and the solution of linear systems of equations $Ax = b$ where b has a pattern of alternating signs to HRA.

In the following sections, we are going to present the bidiagonal decomposition of some matrices of q -integers. This factorization will allow us to compute to HRA their inverses, singular values, and eigenvalues as well as the solution to some linear systems of equations.

3 | q -INTEGERS AND q -PASCAL MATRICES

Given a positive real number q and a natural number r we define the q -integer $[r]$ (see References 1,9) as

$$[r] = \begin{cases} 1 + q + \dots + q^{r-1} = \frac{1-q^r}{1-q}, & \text{if } q \neq 1, \\ r, & \text{if } q = 1, \end{cases}$$

the q -factorial $[r]!$ as

$$[r]! = \begin{cases} [r][r-1] \dots [1], & \text{if } q \neq 1, \\ r!, & \text{if } q = 1, \end{cases}$$

the q -shifted factorial as

$$(a; q)_0 = 1, \quad (a; q)_n = \prod_{k=1}^n (1 - aq^{k-1}), \quad n \in N, \quad a \in R, \quad q \in (0, 1)$$

and the q -binomial coefficient $\begin{bmatrix} i \\ j \end{bmatrix}$ as

$$\begin{bmatrix} i \\ j \end{bmatrix} = \frac{[i]!}{[j]![i-j]}.$$

The q -binomial coefficients satisfy the following recurrence relations

$$\begin{bmatrix} i \\ j \end{bmatrix} = \begin{bmatrix} i-1 \\ j-1 \end{bmatrix} + q^j \begin{bmatrix} i-1 \\ j \end{bmatrix}, \tag{10}$$

$$\begin{bmatrix} i \\ j \end{bmatrix} = q^{i-j} \begin{bmatrix} i-1 \\ j-1 \end{bmatrix} + \begin{bmatrix} i-1 \\ j \end{bmatrix}, \tag{11}$$

and they also satisfy a q -analogue of the Vandermonde identity:

$$\begin{bmatrix} m+n \\ k \end{bmatrix} = \sum_{j=0}^k q^{(k-j)(m-j)} \begin{bmatrix} m \\ j \end{bmatrix} \begin{bmatrix} n \\ k-j \end{bmatrix}. \tag{12}$$

Let us also define the lower triangular matrix of q -binomial coefficients, $P_{L,q}$, whose nonzero entries are given by

$$(P_{L,q})_{ij} = \begin{bmatrix} i-1 \\ j-1 \end{bmatrix}, \quad 1 \leq j \leq i \leq n+1, \tag{13}$$

and its upper triangular counterpart $P_{U,q} := P_{L,q}^T$. Our first result gives the bidiagonal decomposition of $P_{L,q}$ with HRA. In particular, it also shows that it is a TP matrix (the total positivity of $P_{L,q}$ was already known, see p. 198 of Reference 19).

Theorem 3. *Let $P_{L,q}$ be the $(n+1) \times (n+1)$ matrix given by (13). Then $P_{L,q}$ is TP and the bidiagonal decomposition of $P_{L,q}$ is given by*

$$(BD(P_{L,q}))_{ij} = \begin{cases} 1, & i=j, \\ q^{j-1}, & i>j, \\ 0, & \text{otherwise,} \end{cases} \tag{14}$$

which can be computed to HRA.

Proof. We are going to see that the pivots of the NE of $P_{L,q}$ are given by

$$p_{ij} = q^{(i-j)(j-1)}, \quad 1 \leq j \leq i \leq n+1 \tag{15}$$

and that the multipliers are given by

$$m_{ij} = q^{j-1}, \quad 1 \leq j < i \leq n+1. \tag{16}$$

Let $A := P_{L,q}$ and let $A^{(k)} = (a_{ij}^{(k)})_{1 \leq i,j \leq n+1}$ be the matrix obtained after performing $k-1$ steps of the NE of A for $k=2, \dots, n+1$. Let us first prove by induction on $k \in \{2, \dots, n+1\}$ that

$$a_{ij}^{(k)} = q^{(i-j)(k-1)} \begin{bmatrix} i-k \\ j-k \end{bmatrix}, \quad k \leq j \leq i \leq n+1. \tag{17}$$

For $k=2$, using the first step of NE and (11), we can see that

$$a_{ij}^{(2)} = a_{ij} - \frac{a_{i1}}{a_{i-1,1}} a_{i-1,j} = a_{ij} - a_{i-1,j} = \begin{bmatrix} i-1 \\ j-1 \end{bmatrix} - \begin{bmatrix} i-2 \\ j-1 \end{bmatrix} = \begin{bmatrix} i-2 \\ j-2 \end{bmatrix} q^{i-j},$$

for $2 \leq j \leq i \leq n+1$. So, now let us assume that (17) holds for some $k \in \{2, \dots, n\}$ and let us perform the k th step of the NE to prove that (17) holds for $k+1$:

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - \frac{a_{ik}^{(k)}}{a_{i-1,k}^{(k)}} a_{i-1,j}^{(k)}, \quad k+1 \leq j \leq i \leq n+1.$$

By the induction hypothesis we have that

$$\begin{aligned} a_{ij}^{(k+1)} &= a_{ij}^{(k)} - \frac{q^{(i-k)(k-1)} \begin{bmatrix} i-k \\ k-k \end{bmatrix}}{q^{(i-1-k)(k-1)} \begin{bmatrix} i-1-k \\ k-k \end{bmatrix}} a_{i-1,j}^{(k)} = q^{(i-j)(k-1)} \begin{bmatrix} i-k \\ j-k \end{bmatrix} - q^{k-1} q^{(i-1-j)(k-1)} \begin{bmatrix} i-1-k \\ j-k \end{bmatrix} \\ &= q^{(i-j)(k-1)} \left(\begin{bmatrix} i-k \\ j-k \end{bmatrix} - \begin{bmatrix} i-1-k \\ j-k \end{bmatrix} \right). \end{aligned}$$

Applying (11) we deduce that

$$a_{ij}^{(k+1)} = q^{(i-j)k} \begin{bmatrix} i-(k+1) \\ j-(k+1) \end{bmatrix},$$

and hence, (17) holds for $k+1$. Finally, we conclude that the pivot $p_{ij} = a_{ij}^{(j)}$ is given by (17) for $k=j$. Therefore, since $m_{ij} = \frac{p_{ij}}{p_{i-j}}$ for $i > j$, (15) and (16) hold. Then $BD(P_{L,q})$ can be computed through a subtraction-free algorithm by (14). In addition, by (14) $P_{L,q}$ can be written as a product of bidiagonal nonnegative (and hence TP) matrices and then, by theorem 3.1 of Reference 10, $P_{L,q}$ is TP. ■

Let us recall that the $(n+1) \times (n+1)$ Pascal matrix $P = \left(\begin{bmatrix} i+j-2 \\ j-1 \end{bmatrix} \right)_{1 \leq i, j \leq n+1}$ can be expressed as the product of the $(n+1) \times (n+1)$ lower triangular Pascal matrix P_L (whose (i, j) entry is $\begin{bmatrix} i-1 \\ j-1 \end{bmatrix}$ if $i \geq j$ and its transpose:

$$P = P_L P_L^T.$$

This decomposition can be used to deduce the bidiagonal decomposition of P from $BD(P_L)$. Following the same strategy, we can deduce the bidiagonal decomposition of the matrix whose (i, j) entry is the q -binomial coefficient $[i+j-2]_{i-1}$. Let us define the symmetric matrix of q -binomial coefficients P_q :

$$(P_q)_{i,j} = \begin{bmatrix} i+j-2 \\ i-1 \end{bmatrix}, \quad 1 \leq i, j \leq n+1. \quad (18)$$

Proposition 1. *Let P_q be the matrix of q -binomial coefficients given by (18). Then P_q is STP and its bidiagonal decomposition is given by*

$$(BD(P_q))_{i,j} = \begin{cases} q^{(j-1)^2}, & i = j, \\ q^{j-1}, & i > j, \\ q^{i-1}, & \text{otherwise.} \end{cases} \quad (19)$$

Proof. By (12) the q -binomial coefficient $[i+j-2]_{i-1}$ can be written as

$$\begin{bmatrix} i+j-2 \\ i-1 \end{bmatrix} = \sum_{r=0}^{i-1} q^{(i-1-r)^2} \begin{bmatrix} i-1 \\ r \end{bmatrix} \begin{bmatrix} j-1 \\ i-1-r \end{bmatrix} = \sum_{t=0}^{i-1} q^{t^2} \begin{bmatrix} i-1 \\ i-1-t \end{bmatrix} \begin{bmatrix} j-1 \\ t \end{bmatrix} = \sum_{t=0}^{i-1} q^{t^2} \begin{bmatrix} i-1 \\ t \end{bmatrix} \begin{bmatrix} j-1 \\ t \end{bmatrix}.$$

This identity implies that the matrix P_q can be factorized as

$$P_q = P_{L,q} \text{diag}(q^{(j-1)^2})_{1 \leq j \leq n+1} P_{U,q}. \quad (20)$$

From (20) we can deduce $BD(P_q)$ since we know the bidiagonal decomposition of the three factors. Formula (14) gives the bidiagonal decomposition of $P_{L,q}$. Moreover, (14) jointly with Remark 2 allow us to deduce that

$$(BD(P_{U,q}))_{i,j} = \begin{cases} 1, & i = j, \\ 0, & i > j, \\ q^{i-1}, & i < j. \end{cases} \quad (21)$$

Therefore, (20) can be written as

$$P = \bar{F}_n \bar{F}_{n-1} \cdots \bar{F}_1 \text{diag}(q^{(j-1)^2})_{1 \leq j \leq n+1} \bar{G}_1 \cdots \bar{G}_{n-1} \bar{G}_n,$$

where \bar{F}_k, \bar{G}_k are bidiagonal matrices following (6) and (7), respectively, with $k = 1, \dots, n$. The multipliers are $m_{ij} = q^{j-1}$ and $\tilde{m}_{ji} = q^{i-1}$, and hence, by the uniqueness of the bidiagonal decomposition, (19) holds. By Theorem 2, taking into account that $m_{ij}, p_{ii}, \tilde{m}_{ji} > 0$, we deduce that P_q is STP. ■

The bidiagonal decomposition of P_q can be computed to HRA and as a consequence, it serves as a parameterization to perform some algebraic computations with this matrix to HRA.

Corollary 1. *Let P_q be the matrix of q -binomial coefficients given by (18). Then we can compute $BD(P_q)$ with HRA and hence, the following computations can be performed with HRA: all the eigenvalues and singular values, the inverse of P_q , and the solution of the linear systems $P_q x = b$ where $b = (b_0, \dots, b_n)$ has alternating signs.*

Proof. The subtractions in formula (19) are of integers, and hence, they can be computed in an exact way. Therefore, $BD(P_q)$ can be computed with HRA and used to perform also with HRA the algebraic computations mentioned in the statement of this corollary. ■

In the following section, we are going to present a result that shows the relationship between a triangular TP matrix and its inverse. This result will be used to deduce the bidiagonal decomposition of $P_{L,q}^{-1}$.

4 | BIDIAGONAL FACTORIZATION OF THE INVERSE OF A TRIANGULAR TP MATRIX AND Q-STIRLING NUMBERS

In this section, we shall obtain the accurate bidiagonal decomposition of matrices S_q with some q -analogs of the Stirling numbers. We start with matrices C_q with the unsigned q -analogs of the Stirling numbers of the first kind. Since the corresponding matrices B_q of the q -analogs of the Stirling numbers of the second kind are inverses of matrices with the q -analogs of the Stirling numbers of the first kind, we shall analyze the bidiagonal decomposition of a triangular TP matrix previously to the bidiagonal decomposition of matrices B_q .

The q -Stirling numbers of the second kind, $B_q = (b_{ij})_{1 \leq i, j \leq n+1}$, are given by the recurrence relation (see Reference 5)

$$b_{ij} = b_{i-1, j-1} + [j - 1]b_{i-1, j}, \tag{22}$$

with $b_{00} = 1, b_{i0} = 0$ for $i > 0$ and $b_{0j} = 0$ for $j > 0$. The q -Stirling numbers of the first kind, $S_q = (s_{ij})_{1 \leq i, j \leq n+1}$, follow the relationship (see Reference 5)

$$s_{ij} = s_{i-1, j-1} - [i - 1]s_{i-1, j}, \tag{23}$$

with $s_{00} = 1, s_{i0} = 0$ for $i > 0$ and $s_{0j} = 0$ for $j > 0$. Let us define the unsigned q -Stirling numbers of the first kind, $C_q = (c_{ij})_{1 \leq i, j \leq n+1}$, by the following relationship

$$c_{ij} = c_{i-1, j-1} + [i - 1]c_{i-1, j}, \tag{24}$$

with $c_{00} = 1, c_{i0} = 0$ for $i > 0$ and $c_{0j} = 0$ for $j > 0$. The entries of S_q are equal in absolute value to those of $C_q = (c_{ij})_{1 \leq i, j \leq n}$ given by (24). The difference lies on their sign pattern: S_q has a checkerboard pattern of alternating signs while $C_q \geq 0$. We are going to deduce the bidiagonal decomposition of the matrix C_q . In particular, this proposition also serves as a proof that C_q is a TP matrix.

Proposition 2. *Let $C_q = (c_{ij})_{1 \leq i, j \leq n+1}$ be the matrix whose (i, j) entry is the unsigned q -Stirling number of the first kind c_{ij} given by (24). Then C_q is TP and*

$$BD(C_q) = \begin{cases} 1, & i = j, \\ [i - j], & i > j, \\ 0, & \text{otherwise.} \end{cases}$$

Proof. Using (24), let us perform the first step of the NE of C_q :

$$c_{ij}^{(2)} = c_{ij} - \frac{c_{i1}}{c_{i-1,1}} c_{i-1,j} = c_{ij} - [i-1]c_{i-1,j} = c_{i-1,j-1}, \quad 2 \leq j \leq i \leq n+1.$$

We see that $p_{11} = 1$ and $m_{i1} = [i-1]$ for $i > 1$. Moreover, the matrix obtained after one step of the NE satisfies that $C_q^{(2)}[2, \dots, n+1] = C_q[1, \dots, n]$. Hence, we deduce that $p_{jj} = 1$ and $m_{ij} = [i-j]$ for $i > j \geq 2$. Observe now that the (unique) bidiagonal factorization of C_q corresponds to (5) with D and all matrices G_i equal to the identity matrix and the matrices F_i given by (6) and $m_{ij} = [i-j]$ for $i > j \geq 2$. So, C_q is a product of bidiagonal nonnegative (and hence TP) matrices and then, by theorem 3.1 of Reference 10, C_q is TP. ■

By using (23) instead of (24), the same proof of Proposition 2 leads to

$$BD(S_q) = \begin{cases} 1, & i = j, \\ -[i-j], & i > j, \\ 0, & \text{otherwise.} \end{cases} \quad (25)$$

In spite that S_q is not a TP matrix, it is closely related to this class of matrices since it is the inverse of the matrix B_q .

Theorem 4. (Theorem 3.16 of Reference 5) *The two q -Stirling numbers viewed as matrices are inverses of each other:*

$$\sum_k s_{ik} b_{kj} = \delta_{ij},$$

where $\delta_{ij} := 1$ if $i=j$ and $\delta_{ij} := 0$ if $i \neq j$.

The following result gives the bidiagonal decomposition of the inverse of a lower triangular TP matrix A in terms of $BD(A)$ whenever the multipliers of the NE of A are nonzero.

Theorem 5. *Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a lower triangular TP matrix such that*

$$(BD(A))_{ij} = \begin{cases} m_{ij} > 0, & i > j, \\ 0, & i < j, \\ 1, & i = j. \end{cases} \quad (26)$$

Then the bidiagonal decomposition of its inverse is given by

$$(BD(A^{-1}))_{ij} = \begin{cases} -m_{i,i-j}, & i > j, \\ 0, & i < j, \\ 1, & i = j. \end{cases} \quad (27)$$

Proof. Since D and G_i for $i = 1, \dots, n$ are equal to the $n \times n$ identity matrix I_n , we can use (5) and (26) to factorize the matrix A as:

$$A = F_{n-1} \dots F_1 = \{E_n(m_{n,1})\} \{E_{n-1}(m_{n-1,1})E_n(m_{n,2})\} \dots \{E_2(m_{2,1}) \dots E_n(m_{n,n-1})\}.$$

As a direct consequence, A^{-1} can be written as the following product

$$A^{-1} = \{E_n(-m_{n,n-1}) \dots E_2(-m_{2,1})\} \{E_n(-m_{n,n-2}) \dots E_3(-m_{3,1})\} \dots \{E_n(-m_{n,2}) E_{n-1}(-m_{n-1,1})\} \{E_n(-m_{n,1})\}. \quad (28)$$

Using (2) we can rewrite (28) with a permutation of the matrices $E_i(x)$:

$$A^{-1} = \{E_n(-m_{n,n-1}) \dots E_3(-m_{3,2})\} \{E_n(-m_{n,n-2}) \dots E_4(-m_{4,2})\} \dots \{E_n(-m_{n,2})\} \{E_2(-m_{2,1})E_3(-m_{3,1}) \dots E_{n-1}(-m_{n-1,1})E_n(-m_{n,1})\}. \quad (29)$$

The matrix $E_2(-m_{2,1}) \dots E_n(-m_{n,1})$ would be the first factor of $BD(A^{-1})$. Following this argumentation we can keep reordering the matrices $E_i(x)$ of (28) until we obtain (27). ■

An analogous result is true for upper triangular TP matrices.

Corollary 2. *Let A be an upper triangular TP matrix such that*

$$(BD(A))_{ij} = \begin{cases} 0, & i > j, \\ \tilde{m}_{ji} > 0, & i < j, \\ 1, & i = j. \end{cases} \tag{30}$$

Then the bidiagonal decomposition of its inverse is given by

$$(BD(A^{-1}))_{ij} = \begin{cases} 0, & i > j, \\ -\tilde{m}_{i-j,i}, & i < j, \\ 1, & i = j. \end{cases} \tag{31}$$

Proof. We just need to apply Theorem 5 to A^T . ■

The following example shows that the strict positivity of the multipliers in Theorem 5 (or analogously in Corollary 2) is necessary.

Example 1. Let $A = (a_{ij})_{1 \leq i,j \leq 4}$ be the matrix:

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

Applying NE we see that its bidiagonal decomposition is given by

$$BD(A) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix},$$

which means that

$$A = E_4(1)E_3(1)E_2(1)E_4(1). \tag{32}$$

From (32) we deduce that

$$A^{-1} = (E_4(1)E_3(1)E_2(1)E_4(1))^{-1} = E_4(-1)E_2(-1)E_3(-1)E_4(-1), \tag{33}$$

or using the notation (9),

$$BD(A^{-1}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & -1 & 1 \end{pmatrix}.$$

Hence, requiring that the multipliers are nonzero is necessary since $BD(A^{-1})$ does not satisfy (27).

By using (25), Theorem 5 and the arguments of the proof of Proposition 2, we deduce the following result.

Corollary 3. Let $B_q = (b_{ij})_{1 \leq i, j \leq n+1}$ be the matrix whose (i, j) entry is the q -Stirling number of the second kind b_{ij} given by (22). Then B_q is TP and

$$BD(B_q) = \begin{cases} 1, & i = j, \\ [j], & i > j, \\ 0, & \text{otherwise.} \end{cases}$$

Theorem 5 can be used to deduce the bidiagonal decomposition of more matrices. For example, applying it to $P_{L,q}$ we see that

$$BD((P_{L,q})^{-1}) = \begin{cases} 1, & i = j, \\ -q^{i-j-1}, & i > j, \\ 0, & \text{otherwise.} \end{cases}$$

The matrix $BD(A)$ with the bidiagonal factorization of a nonsingular TP matrix A gives also the bidiagonal factorization of the matrix A^{-1} . Taking this into account, Marco and Martínez presented in Section 4 of Reference 20 a fast and accurate algorithm (called `TNInverseExpand`) for computing A^{-1} starting from $BD(A)$. It will be used and recalled in Section 6.

5 | q -LAGUERRE POLYNOMIALS

In this section, we consider the q -Laguerre polynomials $L_{n,q}^{(\alpha)}$ (see p. 552 of Reference 9). These polynomials are given by

$$L_{n,q}^{(\alpha)}(x) = \frac{(q^{\alpha+1}; q)_n}{(q; q)_n} \sum_{k=0}^n \begin{bmatrix} n \\ k \end{bmatrix} q^{\alpha k + k^2} \frac{(-x)^k}{(q^{\alpha+1}; q)_k}. \tag{34}$$

Let $M := (L_{j-1,q}^{(\alpha)}(t_{i-1}))_{1 \leq i, j \leq n+1}$ be the collocation matrix of the q -Laguerre polynomials at the nodes $(0 >) t_0 > t_1 > \dots > t_n$ and let $R_{q,\alpha}, J$ and D_q be the following $(n+1) \times (n+1)$ diagonal matrices :

$$R_{q,\alpha} = \text{diag}((q^{\alpha+1}; q)_{i-1})_{1 \leq i \leq n+1}, \tag{35}$$

$$J = \text{diag}((-1)^{i-1})_{1 \leq i \leq n+1}, \tag{36}$$

$$D_q = \text{diag}(q^{\alpha(i-1) + (i-1)^2})_{1 \leq i \leq n+1}. \tag{37}$$

The following result shows the strict total positivity of M and guarantees HRA for many algebraic computations with M whenever α is a nonnegative integer.

Theorem 6. Let $M := (L_{j-1,q}^{(\alpha)}(t_{i-1}))_{1 \leq i, j \leq n+1}$ for $(0 >) t_0 > t_1 > \dots > t_n$ with $\alpha > -1$ and $0 < q < 1$, let $P_{U,q}$ be the $(n+1) \times (n+1)$ upper triangular matrix given by the transpose of $P_{L,q}$ in (13), and let $R_{q,\alpha}, J$ and D_q be the diagonal matrices given by (35)–(37), respectively. Then

- (i) M is an STP matrix.
- (ii) If $\alpha \in \mathbb{N} \cup \{0\}$, given the nodes t_i ($0 \leq i \leq n$) we can compute $BD(M)$ with HRA and hence, the following computations can be performed with HRA: all the eigenvalues and singular values, the inverse of M , and the solution of the linear systems $Mx = b$ where $b = (b_0, \dots, b_n)$ has alternating signs.

Proof. Let $A = (a_{ij})_{1 \leq i, j \leq n+1}$ be the matrix of change of basis between the basis of the q -Laguerre polynomials and the monomial basis:

$$(L_{0,q}^{(\alpha)}(t), L_{1,q}^{(\alpha)}(t), \dots, L_{n,q}^{(\alpha)}(t)) = (1, t, \dots, t^n)A.$$

By (34) its entries are given by

$$a_{ij} = \frac{(q^{\alpha+1}; q)_{j-1}}{(q; q)_{j-1}} \begin{bmatrix} j-1 \\ i-1 \end{bmatrix} \frac{q^{\alpha(i-1)+(i-1)^2}}{(q^{\alpha+1}; q)_{i-1}} (-1)^{i-1}. \tag{38}$$

Therefore, we can write A as the following product

$$A = JD_q R_{q,\alpha}^{-1} P_{U,q} R_{q,0}^{-1} R_{q,\alpha},$$

and, given $V := (t_{i-1}^{j-1})_{1 \leq i,j \leq n+1}$, the collocation matrix M can be written as

$$M = VJD_q R_{q,\alpha}^{-1} P_{U,q} R_{q,0}^{-1} R_{q,\alpha}. \tag{39}$$

Since $0 < -t_0 < \dots < -t_n$, $VJ = ((-t_{i-1})^{j-1})_{1 \leq i,j \leq n+1}$ is a Vandermonde matrix with strictly increasing positive nodes. Hence, VJ is STP (see p. 12 of Reference 11). The upper triangular matrix $P_{U,q}$ is a nonsingular TP matrix and so $D_q R_{q,\alpha}^{-1} P_{U,q} R_{q,0}^{-1} R_{q,\alpha}$ is also nonsingular TP because D_q , $R_{q,\alpha}^{-1}$, $R_{q,0}^{-1}$, and $R_{q,\alpha}$ are diagonal matrices with positive diagonal entries. We can write (39) as

$$M = BC, \quad B := VJ, \quad C := D_q R_{q,\alpha}^{-1} P_{U,q} R_{q,0}^{-1} R_{q,\alpha}, \tag{40}$$

and so, by theorem 3.1 of Reference 10, M is STP because it is the product of an STP matrix and a nonsingular TP matrix and (i) holds.

Moreover, we can construct the bidiagonal decomposition of a Vandermonde matrix with strictly increasing positive nodes with HRA (see Section 3 of Reference 14). In our case, we can compute $BD(B)$ from the parameters $(0 < -t_0 < \dots < -t_n)$ with HRA. We need to obtain $BD(C)$ to compute $BD(BC)$. In a previous section, we have seen that $BD(P_{U,q})$ is given by (21), which means that $P_{U,q}$ can be expressed as the following product

$$P_{U,q} = \bar{G}_1 \dots \bar{G}_n,$$

where \bar{G}_k with $1 \leq k \leq n$ is the bidiagonal upper triangular matrix

$$\bar{G}_k = \begin{pmatrix} 1 & \bar{g}_1^{(k)} & & & \\ & \ddots & \ddots & & \\ & & \ddots & \bar{g}_n^{(k)} & \\ & & & & 1 \end{pmatrix},$$

with $\bar{g}_i^{(k)} = q^{i-k}$ for $i \geq k$ and $\bar{g}_i^{(k)} = 0$ for $i < k$. Since C is a nonsingular triangular TP matrix, it admits by Theorem 1 a bidiagonal decomposition:

$$C = DG_1 \dots G_n,$$

where D is a positive diagonal matrix and G_k ($1 \leq k \leq n$) are bidiagonal upper triangular matrices with 1's on the main diagonal. We are going to obtain $BD(C)$ from $BD(P_{U,q})$. First, we need to deduce $BD(P_{U,q}\bar{D})$, where $\bar{D} = R_{q,0}^{-1} R_{q,\alpha}$. By the relationship between C and $P_{U,q}$ given by (40) we can write C as

$$C = D_q R_{q,\alpha}^{-1} \bar{G}_1 \dots \bar{G}_n R_{q,0}^{-1} R_{q,\alpha}. \tag{41}$$

The factorization (41) only differs from the expression of the bidiagonal decomposition on the right factor $\bar{D} := R_{q,0}^{-1} R_{q,\alpha}$. Let us denote by G_k the bidiagonal upper triangular matrix with 1's on the main diagonal such that

$$\bar{G}_k \bar{D} = \bar{D} G_k.$$

Using the notation $\bar{D} = \text{diag}(d_0, d_1, \dots, d_n)$ and $g_i^{(k)}$ for the $(i, i+1)$ entry of G_k , we have that $g_i^{(k)} = 0$ for $i < k$ and that, for $i \geq k$,

$$g_i^{(k)} = \frac{d_{i+1}}{d_i} q^{i-k} = \frac{(q^{\alpha+1}; q)_i}{(q; q)_i} \frac{(q; q)_{i-1}}{(q^{\alpha+1}; q)_{i-1}} q^{i-k} = \frac{1 - q^{i+\alpha}}{1 - q^i} q^{i-k}. \tag{42}$$

Hence, the factorization (41) has the form

$$C = D_q R_{q,\alpha}^{-1} \bar{D} G_1 \dots G_n.$$

Let us define $D := D_q R_{q,\alpha}^{-1} \bar{D}$. Since $D = D_q R_{q,\alpha}^{-1} R_{q,0}^{-1} R_{q,\alpha} = D_q R_{q,0}^{-1}$ is a diagonal matrix, by the uniqueness of the bidiagonal decomposition we conclude that $\mathcal{BD}(C)$ is given by

$$C = D G_1 \dots G_n,$$

where $D = \text{diag}\left(\frac{q^{\alpha(i-1)+(i-1)^2}}{(q;q)_{i-1}}\right)_{1 \leq i \leq n+1}$ and G_k are the bidiagonal upper triangular matrices whose nonzero extradiagonal entries are defined by (42) for $1 \leq k \leq n$. Finally, let us check that we can compute $\mathcal{BD}(C)$ with HRA whenever $\alpha \in \mathbb{N} \cup \{0\}$. The diagonal entries of $D = D_q R_{q,0}^{-1}$ can be obtained with HRA following the definitions (37) and (35). Since $\alpha \in \mathbb{N} \cup \{0\}$, we can rewrite (42) to compute the remaining entries as follows:

$$g_i^{(k)} = \frac{1 - q^{i+\alpha}}{1 - q^i} q^{i-k} = \frac{\sum_{t=0}^{i+\alpha-1} q^t}{\sum_{t=0}^{i-1} q^t} q^{i-k}. \quad (43)$$

Therefore, we can compute $\mathcal{BD}(C)$ with HRA, and since we also know $\mathcal{BD}(B) = \mathcal{BD}(VJ)$ with HRA, we can construct $\mathcal{BD}(M)$ by (40) through the subtraction-free algorithm 5.1 of Reference 3, and hence, with HRA. Finally, we can use $\mathcal{BD}(M)$ to perform the algebraic computations mentioned in the statement (ii) to HRA. ■

In the following corollary, we extend the cases where Theorem 6 assures the HRA.

Corollary 4. *Let $M := (L_{j-1,q}^{(\alpha)}(t_{i-1}))_{1 \leq i,j \leq n+1}$ for $(0 >) t_0 > t_1 > \dots > t_n$ with $\alpha > -1$ and $0 < q < 1$. Given the parametrization t_i ($0 \leq i \leq n$), if $\alpha \in \mathbb{Q}$ we can compute $\mathcal{BD}(M)$ with HRA and hence, the following computations can be performed with HRA: all the eigenvalues and singular values, the inverse of M , and the solution of the linear systems $Mx = b$ where $b = (b_0, \dots, b_n)$ has alternating signs.*

Proof. Given the irreducible fraction $\alpha = \frac{a}{b}$, we can rewrite (42) as

$$g_i^{(k)} = \frac{1 - q^{i+\alpha}}{1 - q^i} q^{i-k} = \frac{\sum_{t=0}^{bi+\alpha-1} q^{\frac{t}{b}}}{\sum_{t=0}^{bi-1} q^{\frac{t}{b}}} q^{i-k}. \quad (44)$$

Hence, we can compute $g_i^{(k)}$ with a subtraction-free algorithm, and, following the argumentation given in the proof of Theorem 6, we see that we can compute $\mathcal{BD}(M)$ with HRA and use it to compute with HRA the eigenvalues, singular values, and inverse of M as well as the solution of the linear systems $Mx = b$ whenever b has alternating signs. ■

6 | NUMERICAL EXPERIMENTS

Assuming that the parameterization $\mathcal{BD}(A)$ of a nonsingular TP matrix A is known with HRA, in Reference 3 Koev devised algorithms to compute the inverse, the eigenvalues and the singular values of A and the solution of linear systems of equations $Ax = b$ where b has a pattern of alternating signs. In Reference 20 Marco and Martínez presented another algorithm for the computation of A^{-1} from $\mathcal{BD}(A)$. These algorithms were implemented to be used with Matlab and Octave in the software library *TNTTool* available in Reference 13. The corresponding functions are `TNEigenvalues`, `TNSingularValues`, `TNSolve`, and `TNInverseExpand`, respectively. These four functions require, as input argument, the data determining the bidiagonal decomposition $\mathcal{BD}(A)$ of A given by (9), to HRA. `TNSolve` also requires a second argument, the vector b of the linear system $Ax = b$ to be solved.

The computational cost for both `TNSolve` and `TNInverseExpand` is $\mathcal{O}(n^2)$ elementary operations (see Reference 3 and Section 4 of Reference 20) and for `TNEigenValues` and `TNSingularValues` is $\mathcal{O}(n^3)$ elementary operations.

6.1 | q -Pascal matrices

In Proposition 1, it has been provided the bidiagonal decomposition of the q -Pascal matrices defined by (18). In Corollary 1, it has been proved that the bidiagonal decomposition can be computed to HRA and so the eigenvalues and singular values, the inverse of P_q , and the solution of some linear systems with the coefficient matrix P_q . The pseudocode providing $BD(P_q)$ to HRA can be seen in Algorithm 1. We have also included the value $q = 1$, which corresponds to Pascal matrices. In this case, Algorithm 1 provides the bidiagonal decomposition of Pascal matrices to HRA (see Reference 4).

Algorithm 1. Computation of the bidiagonal decomposition of P_q to HRA

Require: $q \in (0, 1]$, order n of the matrix

Ensure: $BD(P_q)$ bidiagonal decomposition of P_q to HRA

```

for  $i = 1 : n$  do
  for  $j = 1 : i - 1$  do
     $(BD(P_q))_{ij} = q^{j-1}$ 
  end for
   $(BD(P_q))_{ii} = q^{(i-1)^2}$ 
  for  $j = i + 1 : n$  do
     $(BD(P_q))_{ij} = q^{i-1}$ 
  end for
end for

```

We have implemented the previous algorithm to be used in Matlab and Octave in a function `TNBDqPascal`. The bidiagonal decompositions with HRA of q -Pascal matrices obtained with `TNBDqPascal` can be used with `TNInverseExpand`, `TNEigenValues`, `TNSingularValues`, and `TNSolve` to obtain accurate solutions for the above mentioned algebraic problems.

Remark 3. It can be checked that the computational cost of Algorithm 1 is of $\mathcal{O}(n^2)$ elementary operations. Taking into account this fact and the computational cost for the methods `TNSolve` and `TNInverseExpand`, `TNBDqPascal` with these two functions provides algorithms of $\mathcal{O}(n^2)$ elementary operations to solve a linear system of equations $P_q x = b$ and to compute the inverse P_q^{-1} , in contrast to the $\mathcal{O}(n^3)$ elementary operations of the standard algorithms for those problems. For the case of eigenvalues and singular values of P_q , taking into account the computational cost of `TNEigenValues` and `TNSingularValues`, `TNBDqPascal` with these two functions provides $\mathcal{O}(n^3)$ algorithms to compute the eigenvalues and singular values of P_q .

Now we include some numerical experiments illustrating the high accuracy. Let us consider the q -Pascal matrix of order 21 given by (18) with $q = 0.5$.

First, we have computed in Matlab, by using `TNBDqPascal`, the bidiagonal decomposition of the considered q -Pascal matrix P_q to HRA. Then we have used that bidiagonal decomposition for computing the eigenvalues of P_q with `TNEigenValues`. We also compute their approximations with the Matlab function `eig`. We have also computed the eigenvalues of P_q by using mathematica with a 200 digits precision. Then we compute the relative errors corresponding to the approximations $\hat{\lambda}_i$ of the eigenvalues λ_i obtained with both methods `eig` and `TNEigenValues` with `TNBDqPascal`, considering the eigenvalues provided by Mathematica as exact. We have ordered the eigenvalues in the following way: $\lambda_1 > \lambda_2 \dots > \lambda_{21}$. In Table 1, the relative errors can be seen. We observe that the HRA method provides very accurate approximations in contrast to the poor approximations provided by the usual method for the lower eigenvalues. Recall that, since P_q is symmetric, its singular values coincide with its eigenvalues.

We have also computed with Matlab approximations to P_q^{-1} with both `inv` and with `TNInverseExpand` using the bidiagonal decomposition given by `TNBDqPascal`. With mathematica, we have computed the inverse of this matrix with exact arithmetic. Then we have computed the corresponding componentwise relative errors. Finally, we have obtained the mean and maximum componentwise relative error for both methods. The results can be seen in Table 2.

i	$\frac{ \lambda_i - \hat{\lambda}_i }{ \lambda_i }$ HRA	$\frac{ \lambda_i - \hat{\lambda}_i }{ \lambda_i }$ eig
1	$2.2e - 16$	$1.0e + 00$
2	$4.2e - 16$	$1.0e + 00$
3	$2.5e - 16$	$1.0e + 00$
4	$4.8e - 16$	$1.0e + 00$
5	$7.7e - 16$	$1.0e + 00$
6	$4.3e - 16$	$1.0e + 00$
7	$1.2e - 15$	$1.0e + 00$
8	$1.2e - 16$	$8.5e - 01$
9	$4.8e - 16$	$7.1e + 03$
10	$2.5e - 16$	$1.2e + 09$
11	$1.2e - 16$	$9.2e + 14$
12	$2.5e - 16$	$2.2e + 21$
13	$9.1e - 16$	$3.9e + 28$
14	$8.0e - 16$	$5.6e + 36$
15	$1.9e - 15$	$1.8e + 48$
16	$2.4e - 15$	$2.2e + 60$
17	$6.4e - 16$	$2.9e + 72$
18	$2.2e - 15$	$4.4e + 84$
19	$1.1e - 16$	$7.6e + 96$
20	$1.7e - 16$	$1.5e + 109$
21	$9.6e - 16$	$1.1e + 123$

TABLE 1 Relative errors for the eigenvalues of P_q

Abbreviation: HRA, high relative accuracy.

	HRA method	inv
mean rel. error	$9.4585e - 17$	1.0000
maximum rel. error	$5.1298e - 16$	1.0000

TABLE 2 Relative errors for P_q^{-1}

Abbreviation: HRA, high relative accuracy.

	HRA method	A\b
mean rel. error	$1.5656e - 16$	1
maximum rel. error	$5.5342e - 16$	1

TABLE 3 Relative errors for $P_q x = b$

Abbreviation: HRA, high relative accuracy.

Now we consider the linear system $P_q x = b$ where $b \in \mathbb{R}^{21}$ has the absolute value of its entries randomly generated as integers in the interval $[1, 1000]$, but with alternating signs. We have computed approximations to the solution x of the linear system with Matlab, the first one using `TNSolve` and the bidiagonal decomposition of the q -Pascal matrices obtained with `TNBDqPascal`, and the second one using the Matlab command `A\b`. By using Mathematica with exact arithmetic we have computed the exact solution of the systems and then we have computed the componentwise relative errors for the two approximations obtained with Matlab. Then we have obtained the mean componentwise relative error and the maximum componentwise relative error. Table 3 shows these relative errors. Again, the results obtained with HRA algorithms are very accurate in contrast to the poor results obtained with the usual Matlab command.

6.2 | Matrices of q -Stirling numbers

In Proposition 2, it has been proved the matrix $C_q = (c_{ij})_{1 \leq i, j \leq n}$ formed by the unsigned q -Stirling numbers (24) is TP, and it has also been provided the bidiagonal decomposition of the matrix. This bidiagonal decomposition can be computed to HRA and so the eigenvalues and singular values, the inverse of C_q , and the solution of some linear systems with coefficient matrix C_q . The pseudocode providing $BD(C_q)$ to HRA can be seen in Algorithm 2.

Algorithm 2. Computation of the bidiagonal decomposition of C_q to HRA

Require: $q \in (0, 1]$, order n of the matrix

Ensure: $BD(C_q)$ bidiagonal decomposition of C_q to HRA

```

for  $i = 1 : n$  do
  for  $j = 1 : i - 1$  do
     $(BD(P_q))_{ij} = [i - j]$ 
  end for
   $(BD(P_q))_{ii} = 1$ 
  for  $j = i + 1 : n$  do
     $(BD(P_q))_{ij} = 0$ 
  end for
end for

```

We have implemented the previous algorithm to be used in Matlab and Octave in a function `TNBDunsQStir1`. The bidiagonal decompositions with HRA of C_q obtained with that function can be used with `TNInverseExpand`, `TNEigenValues`, `TNSingularValues`, and `TNSolve` to obtain accurate solutions for the above mentioned algebraic problems.

Remark 4. Algorithm 2 consists of the computations of $[0], [1], \dots, [n-1]$. Then, taking into account that for $q < 1$ $[r] = q \times [r-1] + 1$, it can be checked that the computational cost of Algorithm 2 is $\mathcal{O}(n)$ elementary operations. Hence, `TNBDunsQStir1` with `TNSolve` and `TNInverseExpand` provides algorithms to solve a linear system $C_q x = b$ and to compute the inverse C_q^{-1} with $\mathcal{O}(n^2)$ elementary operations. In addition, `TNBDunsQStir1` with `TNEigenValues` and `TNSingularValues` provides algorithms for computing the eigenvalues and singular values of C_q with a computational cost of $\mathcal{O}(n^3)$ elementary operations.

Now we include some numerical experiments illustrating the high accuracy.

Let us consider the matrix C_q of order 20 with $q = 0.5$. First we have computed in Matlab, by using `TNBDunsQStir1`, the bidiagonal decomposition of the considered matrix C_q to HRA. Then we have used that bidiagonal decomposition for computing the singular values of C_q with `TNSingularValues`. We also compute their approximations with the Matlab function `svd`. We have also computed the singular values of C_q by using Mathematica with a 200 digits precision. Then we compute the relative errors corresponding to the approximations $\hat{\sigma}_i$ of the singular values σ_i obtained with both methods `svd` and `TNSingularValues` with `TNBDunsQStir1`, considering the singular values provided by Mathematica as exact. We have ordered the singular values in the following way: $\sigma_1 > \sigma_2 \dots > \sigma_{20}$. The relative errors can be seen in Table 4. We observe that the HRA method provides very accurate approximations in contrast to the poor approximations provided by the usual method for the lower singular values.

We have also computed with Matlab approximations to C_q^{-1} with both `inv` and with `TNInverseExpand` using the bidiagonal decomposition given by `TNBDunsQStir1`. With mathematica, we have computed the inverse of this matrix with exact arithmetic. Then we have computed the corresponding componentwise relative errors. Finally, we have obtained the mean and maximum componentwise relative error for both methods. The results can be seen in Table 5.

Now we consider the linear system $C_q x = b$ where $b \in \mathbb{R}^{20}$ has the absolute value of its entries randomly generated as integers in the interval $[1, 1000]$, but with alternating signs. We have computed approximations to the solution x of the linear system with Matlab, the first one using `TNSolve` and the bidiagonal decomposition of the matrices C_q obtained with `TNBDunsQStir1`, and the second one using the Matlab command `A\b`. By using Mathematica with exact arithmetic

i	$\frac{ \sigma_i - \hat{\sigma}_i }{ \sigma_i }$ HRA	$\frac{ \sigma_i - \hat{\sigma}_i }{ \sigma_i }$ svd
1	2.804e-16	1.402e-16
2	8.0692e-16	2.0173e-16
3	0	3.9227e-16
4	2.9638e-16	1.393e-14
5	7.3461e-16	5.5096e-16
6	1.9151e-16	4.2266e-13
7	0	4.4546e-12
8	1.2725e-16	2.6327e-11
9	1.5293e-16	1.7817e-10
10	4.41709e-16	1.61523e-10
11	6.63314e-16	4.89389e-09
12	8.30966e-16	3.74874e-09
13	9.58919e-16	3.23951e-09
14	7.59915e-16	5.7262e-09
15	6.81317e-16	2.55909e-08
16	3.52401e-16	1.90929e-08
17	1.22622e-16	8.05222e-08
18	3.17177e-16	2.92423e-08
19	6.68987e-16	5.43857e-08
20	8.45409e-16	5.02386e-09

TABLE 4 Relative errors for the singular values of C_q

Abbreviation: HRA, high relative accuracy.

	HRA method	inv
mean rel. error	1.6095e-18	6.6032e-06
maximum rel. error	2.1819e-16	2.1926e-03

TABLE 5 Relative errors for C_q^{-1}

Abbreviation: HRA, high relative accuracy.

	HRA method	A\b
mean rel. error	3.8540e-17	7.1034e-12
maximum rel. error	2.1309e-16	8.4793e-11

TABLE 6 Relative errors for $C_q x = b$

Abbreviation: HRA, high relative accuracy.

we have computed the exact solution of the systems and then we have computed the componentwise relative errors for the two approximations obtained with Matlab. Then we have obtained the mean componentwise relative error and the maximum componentwise relative error. Table 6 shows these relative errors. Again, the results obtained with HRA algorithms are very accurate in contrast to the poor results obtained with the usual Matlab command.

6.3 | Collocation matrices of q -Laguerre polynomials

In the proof of Theorem 6 it has been shown how to compute the bidiagonal decomposition of the collocation matrix $M := (L_{j-1,q}^{(\alpha)}(t_{i-1}))_{1 \leq i,j \leq n+1}$ to HRA for $(0 >) t_0 > t_1 > \dots > t_n$ with $\alpha \in \mathbb{N} \cup \{0\}$ and $0 < q < 1$. According to that proof $M = BC$,

where $B = VJ$ is a TP Vandermonde matrix with node sequence $-\mathbf{t} = (-t_i)_{i=0}^n$. Then, by using $\text{TNVandBD}(-\mathbf{t})$ of library *TNTool* $\text{BD}(VJ)$ can be obtained to HRA. By formula (43) in the proof, the bidiagonal decomposition of the matrix C can also be obtained to HRA for the considered parameters. In the library *TNTool*, Koev also provided the function $\text{TNProduct}(B1, B2)$, which, given the bidiagonal decompositions $B1$ and $B2$ to HRA of two TP matrices B and C , provides the bidiagonal decomposition of the TP matrix BC to HRA. Taking into account these facts, the pseudocode providing $\text{BD}(M)$ to HRA can be seen in Algorithm 3.

Algorithm 3. Computation of the bidiagonal decomposition of M to HRA

Require: $\mathbf{t} = (t_i)_{i=0}^n$ such that $0 > t_0 > t_1 > \dots > t_n$, $q \in (0, 1)$ and $\alpha \in \mathbb{N} \cup \{0\}$

Ensure: $\text{BD}(M)$ bidiagonal decomposition of M to HRA

$\text{BD}(B) = \text{TNVandBD}(-\mathbf{t})$

for $i = 1 : n + 1$ **do**

$$(\text{BD}(C))_{ii} = \frac{q^{\alpha(i-1)+(i-1)^2}}{(q; q)_{i-1}}$$

end for

for $k = 1 : n$ **do**

for $i = 1 : n - k$ **do**

$$(\text{BD}(C))_{i,i+k} = \frac{\sum_{r=0}^{i+k+\alpha-2} q^r}{\sum_{r=0}^{i+k-2} q^r} q^{i-1}$$

end for

end for

$\text{BD}(M) = \text{TNProduct}(\text{BD}(B), \text{BD}(C))$

We have implemented the previous algorithm to be used in Matlab and Octave in a function TNBDqLaguerre . The bidiagonal decompositions with HRA of q -Laguerre matrices obtained with TNBDqLaguerre can be used with TNInverseExpand , TNEigenValues , TNSingularValues , and TNSolve to obtain accurate solutions for the above mentioned algebraic problems.

Remark 5. Taking into account the computation of $q^{n+\alpha+n^2}$, that $(q; q)_r = (q; q)_{r-1} \times (1 - q^r)$ and that $\sum_{r=0}^{i+k+\alpha-2} q^r = q \times \sum_{r=0}^{i+k+\alpha-3} q^r + 1$ it can be deduced that the computational cost of Algorithm 3 is $\mathcal{O}(n(n + \alpha))$ elementary operations.

Now we include some numerical experiments illustrating the high accuracy.

Let us consider the q -Laguerre matrices M_n of order n given by the collocation matrices of the q -Laguerre polynomials $(L_{0,q}^{(2)}(x), \dots, L_{n-1,q}^{(2)}(x))$ at the nodes $(-i)_{1 \leq i \leq n}$, that is,

$$M_n = (L_{j-1,q}^{(2)}(-i))_{1 \leq i,j \leq n}, \quad (45)$$

for $n = 2, \dots, 30$.

First we have computed in Matlab by using TNBDqLaguerre , the bidiagonal decomposition of the matrices M_n to HRA. Then we have used that bidiagonal decomposition of M_n for computing their eigenvalues and their singular values with TNEigenValues and TNSingularValues , respectively. In the case of eigenvalues, we also compute their approximations with the Matlab function eig . We have also computed the eigenvalues of M_n by using Mathematica with a 500 digits precision. Then we compute the relative errors corresponding to the approximations of the eigenvalues obtained with both methods eig and TNEigenValues with TNBDqLaguerre , considering the eigenvalues provided by Mathematica as exact. We have observed that the approximations of all the eigenvalues obtained with TNBDqLaguerre are very accurate, whereas the approximations of the lower eigenvalues obtained with the command eig are not very accurate. In particular, the lower the eigenvalue is, the more inaccurate the approximation obtained with eig is. To illustrate this fact, Figure 1 shows the relative errors of the approximations to the lowest eigenvalue of the matrices M_2, \dots, M_{30} obtained by both eig and TNEigenValues with TNBDqLaguerre . We can observe in the figure that our method provides very accurate results in contrast to the very poor results provided by eig .

For the case of singular values, we have also computed their approximations with the Matlab function svd . To show the accuracy of the approximations to the singular values computed in both ways we calculate the singular values of the matrices M_n with Mathematica using a precision of 500 digits. As in the case of eigenvalues, we observed that the lower

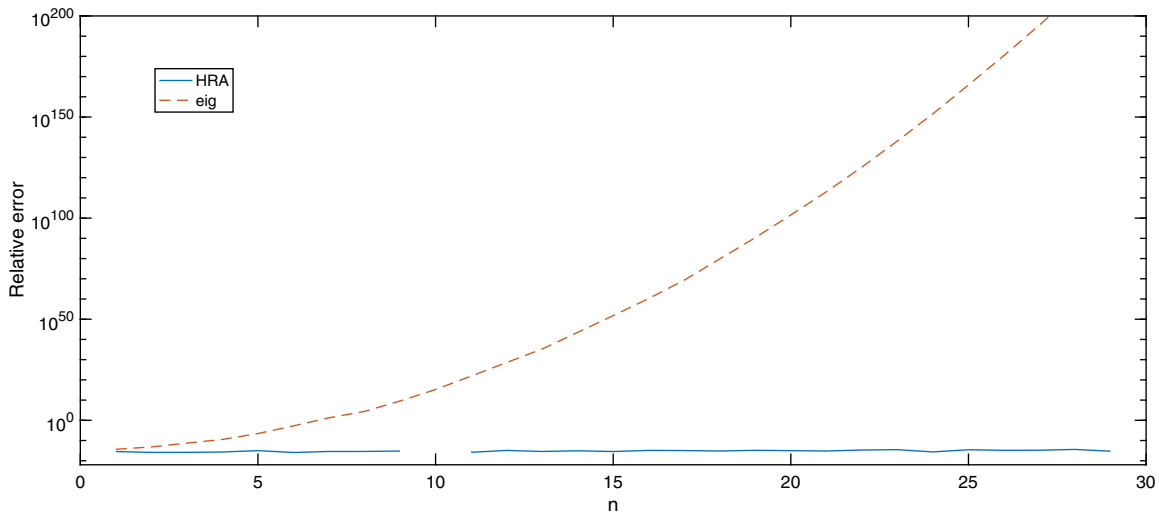


FIGURE 1 Relative errors for the lowest eigenvalue of q -Laguerre matrices

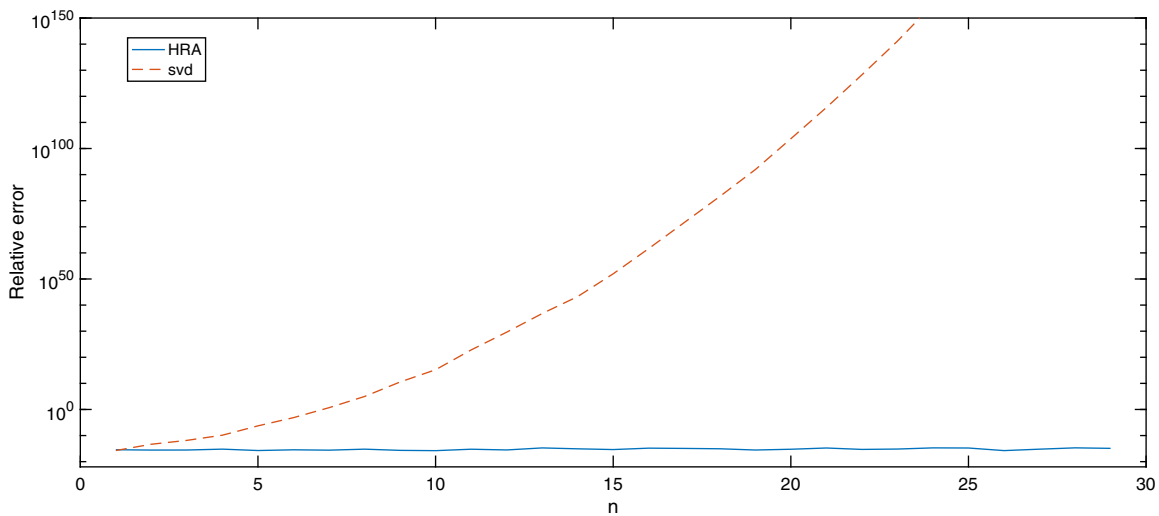


FIGURE 2 Relative errors for the lowest singular value of q -Laguerre matrices

the singular value is, the more inaccurate the approximation obtained with `svd` is, whereas the approximations of all the singular values provided by the new method are very accurate. Figure 2 shows the relative errors of the approximations to the lowest singular value of the matrices M_2, \dots, M_{30} obtained by both `svd` and `TNSingularValues` with `TNBDqLaguerre`. We can observe in the figure that HRA algorithm provides very accurate results. By contrast, `svd` provides very inaccurate results.

We have also computed with Matlab approximations to $M_n^{-1}, n = 2, \dots, 30$, with `inv` and with `TNInverseExpand` using the bidiagonal decomposition given by `TNBDqLaguerre`. With mathematica, we have computed the inverse of these q -Laguerre matrices with exact arithmetic. Then we have computed the corresponding componentwise relative errors. Finally, we have obtained the mean and maximum componentwise relative error. Figure 3(a) shows the mean relative error and (b) shows the maximum relative error. We can also observe in this case that the results obtained with `TNInverseExpand` are much more accurate than the ones obtained with `inv`.

Now we consider the linear systems $M_n x = b_n, n = 2, \dots, 30$, where M_n is the q -Laguerre matrix of order n previously defined in (45) and $b_n \in \mathbb{R}^n$ has the absolute value of its entries randomly generated as integers in the interval $[1, 1000]$, but with alternating signs. We have computed approximations to the solution x of the linear system with Matlab, the first one using `TNSolve` and the bidiagonal decomposition of the q -Laguerre matrices obtained with `TNBDqLaguerre`, and the second one using the Matlab command `A\b`. By using Mathematica with exact arithmetic we have computed the exact solution of the systems and then we have computed the componentwise relative errors for the two approximations

FIGURE 3 Relative errors for M_n^{-1} , $n = 2, \dots, 30$

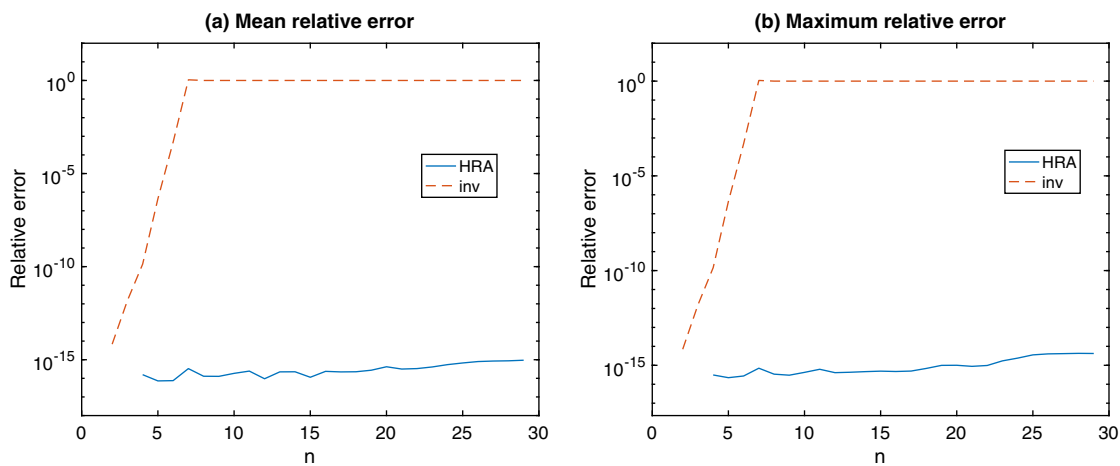
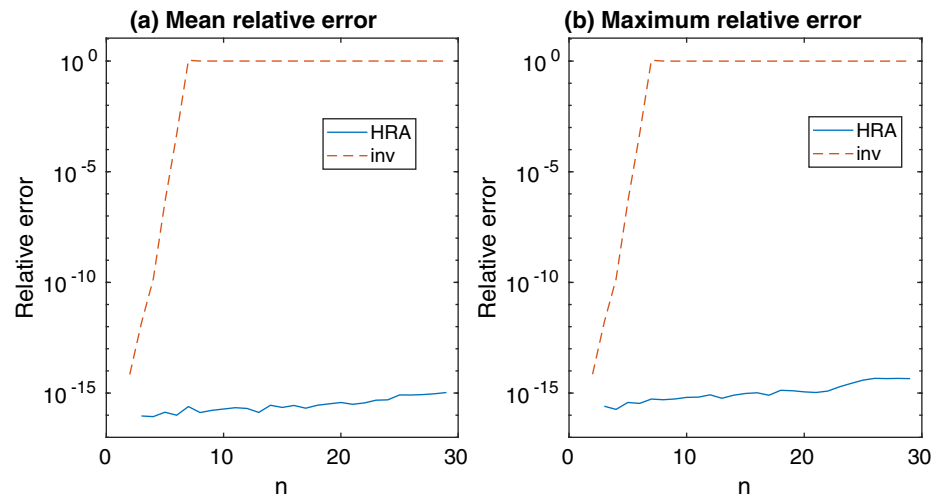


FIGURE 4 Relative errors for the systems $M_n x = b_n$, $n = 2, \dots, 30$

obtained with Matlab. Then we have obtained the mean and maximum componentwise relative error. Figure 4 shows these mean and maximum relative errors. Again, the results obtained with HRA algorithms are very accurate in contrast to the results obtained with the usual Matlab command.

7 | CONCLUSIONS

The bidiagonal decomposition of a triangular TP matrix can be used to derive explicitly the bidiagonal decomposition of its inverse. The bidiagonal decomposition of q -Pascal matrices, q -Stirling matrices, and a large family of collocation matrices of generalized q -Laguerre polynomials are constructed with HRA. They can be used to compute with HRA the eigenvalues, singular values, and inverses of these matrices, as well as the solutions of linear systems $Mx = b$, where M is a matrix of these classes and b is a vector with alternating signs. Numerical examples illustrate the accuracy of the proposed methods in contrast to the poor results obtained with the corresponding Matlab commands.

ACKNOWLEDGMENTS

This research was partially funded by the Spanish research grant PGC2018- 096321-B-I00 (MCIU/AEI), by Gobierno de Aragón (E41_17R).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Jorge Delgado  <https://orcid.org/0000-0003-2156-9856>

REFERENCES

1. Kac V, Cheung P. Quantum calculus. New York, NY: Springer; 2002.
2. Delgado J, Peña JM. Fast and accurate algorithms for Jacobi-Stirling matrices. *Appl Math Comput*. 2014;236:253–9.
3. Koev P. Accurate computations with totally nonnegative matrices. *SIAM J Matrix Anal Appl*. 2007;29:731–51.
4. Alonso P, Delgado J, Gallego R, Peña JM. Conditioning and accurate computations with Pascal matrices. *J Comput Appl Math*. 2013;252:21–6.
5. Ernst T. q -Stirling numbers, an umbral approach. *Adv Dyn Syst Appl*. 2008;3:251–82.
6. Delgado J, Peña JM. Accurate computations with collocation matrices of q -Bernstein polynomials. *SIAM J Matrix Anal Appl*. 2015;36:880–93.
7. Marco A, Martínez JJ, Viaña R. Accurate bidiagonal decomposition of totally positive h -Bernstein-Vandermonde matrices and applications. *Linear Algebra Appl*. 2019;579:320–35.
8. Delgado J, Orera H, Peña JM. Accurate computations with Laguerre matrices. *Numer Linear Algebra Appl*. 2019;26:e2217 10 pp.
9. Ismail M. Classical and quantum orthogonal polynomials in one variable. *Encyclopedia of Mathematics and Its Applications*. Vol 98. Cambridge, MA: Cambridge University Press; 2005.
10. Ando T. Totally positive matrices. *Linear Algebra Appl*. 1987;90:165–219.
11. Fallat SM, Johnson CR. Totally nonnegative matrices. *Princeton Series in Applied Mathematics*. Vol 35. Princeton, NJ: Princeton University Press; 2011.
12. Pinkus A. Totally positive matrices. *Tracts in Mathematics*. Vol 181. Cambridge, UK: Cambridge University Press; 2010.
13. Koev P. <http://www.math.sjsu.edu/~koev/software/TNTool.html>. Accessed 15 Jan 2020.
14. Koev P. Accurate eigenvalues and SVDs of totally nonnegative matrices. *SIAM J Matrix Anal Appl*. 2005;27:1–23.
15. Gasca M, Peña JM. Total positivity and Neville elimination. *Linear Algebra Appl*. 1992;165:25–44.
16. Gasca M, Peña JM. On factorizations of totally positive matrices. In: Gasca M, Micchelli CA, editors. *Total positivity and its applications*. Dordrecht, Netherlands: Kluwer Academic Publishers; 1996. p. 109–30.
17. Barreras A, Peña JM. Accurate computations of matrices with bidiagonal decomposition using methods for totally positive matrices. *Numer Linear Algebra Appl*. 2013;20:413–24.
18. Demmel J, Koev P. The accurate and efficient solution of a totally positive generalized Vandermonde linear system. *SIAM J Matrix Anal Appl*. 2005;27:142–52.
19. Brenti F. Combinatorics and total positivity. *J Combin Theory Ser A*. 1995;71:175–218.
20. Marco A, Martínez JJ. Accurate computation of the Moore-Penrose inverse of strictly totally positive matrices. *J Comput Appl Math*. 2019;350:299–308.

How to cite this article: Delgado J, Orera H, Peña JM. High relative accuracy with matrices of q -integers. *Numer Linear Algebra Appl*. 2021;e2383. <https://doi.org/10.1002/nla.2383>

Article 8

- [19] J. Delgado, H. Orera and J. M. Peña. Optimal properties of tensor product of B-bases. *Appl. Math. Lett.* 121 (2021), Paper No. 107473, 5 pp.

Optimal properties of tensor product of B-bases[☆]Jorge Delgado^{a,*}, Héctor Orera^b, J.M. Peña^b^a Departamento de Matemática Aplicada/IUMA, Escuela de Ingeniería y Arquitectura de Zaragoza, Universidad de Zaragoza, Calle María de Luna, 3, Zaragoza, 50018, Spain^b Departamento de Matemática Aplicada/IUMA, Facultad de Ciencias, Universidad de Zaragoza, Calle de Pedro Cerbuna, 12, Zaragoza, 50009, Spain

ARTICLE INFO

Article history:

Received 25 May 2021

Accepted 16 June 2021

Available online 24 June 2021

Keywords:

Tensor product

B-basis

Totally positive basis

Conditioning

ABSTRACT

It is proved the optimal conditioning for the ∞ -norm of collocation matrices of the tensor product of normalized B-bases among the tensor product of all normalized totally positive bases of the corresponding space of functions. Bounds for the minimal eigenvalue and singular value and illustrative numerical examples are also included.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction and main results

Given a system of functions $u = (u_0, \dots, u_n)$ defined on $I \subseteq \mathbb{R}$, the *collocation matrix* of u at $t_0 < \dots < t_m$ in I is given by $(u_j(t_i))_{i=0, \dots, m}^{j=0, \dots, n}$. If $\sum_{i=0}^n u_i(t) = 1$ for all $t \in I$, then we say that the system is *normalized*. If all collocation matrices of u have all their minors nonnegative, then we say that the system is *totally positive* (TP). Normalized totally positive (NTP) systems play a crucial role in Computer Aided Geometric Design because they lead to shape preserving representations. Among all NTP bases of a space, the basis with optimal shape preserving properties is the *normalized B-basis* [1,2]. The Bernstein basis of polynomials and the B-spline basis are examples of normalized B-bases of their corresponding spaces. In this paper we extend some optimal properties of normalized B-bases given in [2] to their corresponding tensor products. Recall that, given two systems $u^1 = (u_0^1, \dots, u_m^1)$ and $u^2 = (u_0^2, \dots, u_n^2)$ of functions defined on $[a, b]$ and $[c, d]$, respectively, the system $u^1 \otimes u^2 := (u_i^1(x) \cdot u_j^2(y))_{i=0, \dots, m}^{j=0, \dots, n}$ is called a tensor product system and generates a tensor product surface. The *Kronecker product* of two square matrices $A = (a_{ij})_{1 \leq i, j \leq m}$ and

[☆] This work was partially supported by MCIU/AEI through the Spanish research grant PGC2018-096321-B-I00 and by Gobierno de Aragón (E41-17R).

* Corresponding author.

E-mail addresses: jorgedel@unizar.es (J. Delgado), hectororera@unizar.es (H. Orera), jmpena@unizar.es (J.M. Peña).

$B = (b_{ij})_{1 \leq i, j \leq n}$, $A \otimes B$, is defined to be the $mn \times mn$ block matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1m}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mm}B \end{pmatrix}.$$

Given the collocation matrices $B_1 := (u_j^1(x_i))_{0 \leq i, j \leq m}$ and $B_2 := (u_j^2(y_i))_{0 \leq i, j \leq n}$ of u^1 and u^2 , $B_1 \otimes B_2$ is the collocation matrix of $u^1 \otimes u^2$ at $((x_i, y_j)_{j=0, \dots, n})_{i=0, \dots, m}$.

Given two square real matrices $A = (a_{ij})_{1 \leq i, j \leq n}$ and $B = (b_{ij})_{1 \leq i, j \leq n}$, $A \leq B$ denotes that $a_{ij} \leq b_{ij}$ for all i, j . Given a complex matrix $C = (c_{ij})_{1 \leq i, j \leq n}$, A is said to *dominate* C if $|c_{ij}| \leq a_{ij}$ for all i, j . If matrices A and B are nonsingular, by Corollary 4.2.11 of [3] we have that $A \otimes B$ is nonsingular and

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}. \tag{1}$$

The next result shows the optimal properties of a collocation matrix of the tensor product of normalized B-bases among all the corresponding collocation matrices of the tensor product of NTP bases of the spaces.

Theorem 1. *Let $u^1 = (u_0^1, \dots, u_m^1)$ be an NTP basis on $[a, b]$ of a space of functions \mathcal{U}_1 , $u^2 = (u_0^2, \dots, u_n^2)$ be an NTP basis on $[c, d]$ of a space of functions \mathcal{U}_2 and let $v^1 = (v_0^1, \dots, v_m^1)$ and $v^2 = (v_0^2, \dots, v_n^2)$ be the normalized B-bases of \mathcal{U}_1 and \mathcal{U}_2 , respectively. Given the increasing sequences of nodes $\mathbf{t} = (t_i)_{i=0}^m$ on $[a, b]$ and $\mathbf{r} = (r_i)_{i=0}^n$ on $[c, d]$, the nonsingular collocation matrices A_1 and M_1 of the bases u^1 and v^1 , respectively, at \mathbf{t} , and A_2 and M_2 of the bases u^2 and v^2 , respectively, at \mathbf{r} , the following properties hold:*

- (i) *The matrix $|(A_1 \otimes A_2)^{-1}|$ dominates $(M_1 \otimes M_2)^{-1}$.*
- (ii) *The minimal eigenvalue (resp., singular value) of $A_1 \otimes A_2$ is bounded above by the minimal eigenvalue (resp., singular value) of $M_1 \otimes M_2$.*
- (iii) $\kappa_\infty(M_1 \otimes M_2) \leq \kappa_\infty(A_1 \otimes A_2)$.

Proof.

- (i) By Corollary 1 of [2], $|A_1^{-1}|$ dominates $|M_1^{-1}|$ and $|A_2^{-1}|$ dominates $|M_2^{-1}|$. Hence, $|A_1^{-1}| \otimes |A_2^{-1}|$ dominates $|M_1^{-1}| \otimes |M_2^{-1}|$, and, since $|(A_1 \otimes A_2)^{-1}| = |A_1^{-1} \otimes A_2^{-1}| = |A_1^{-1}| \otimes |A_2^{-1}|$ by (1), $|(A_1 \otimes A_2)^{-1}|$ dominates $(M_1 \otimes M_2)^{-1}$.
- (ii) Let B_1 be an $n \times n$ matrix and B_2 an $m \times m$ matrix. If λ is an eigenvalue of B_1 and μ is an eigenvalue of B_2 , then $\lambda\mu$ is an eigenvalue of $B_1 \otimes B_2$ and every eigenvalue of $B_1 \otimes B_2$ arises as such a product of eigenvalues of B_1 and B_2 (see Theorem 4.2.12 of [3]). By Corollary 2 of [2], we have that $\lambda_{\min}(A_1) \leq \lambda_{\min}(M_1)$ and that $\lambda_{\min}(A_2) \leq \lambda_{\min}(M_2)$. Hence,

$$\lambda_{\min}(M_1 \otimes M_2) = \lambda_{\min}(M_1)\lambda_{\min}(M_2) \geq \lambda_{\min}(A_1)\lambda_{\min}(A_2) = \lambda_{\min}(A_1 \otimes A_2).$$

The case of singular values is analogous to that of eigenvalues recalling that every nonzero singular value of $B_1 \otimes B_2$ is the product of a singular value of B_1 and a singular value of B_2 (see Theorem 4.2.15 of [3]).

- (iii) First, let us see that the infinity norm of the Kronecker product of two matrices $A = (a_{ij})_{1 \leq i, j \leq m}$ and $B = (b_{ij})_{1 \leq i, j \leq n}$ satisfies that $\|A \otimes B\|_\infty = \|A\|_\infty \|B\|_\infty$:

$$\|A \otimes B\|_\infty = \max_{0 \leq i \leq nm-1} \sum_{j=1}^m |a_{t+1, j}| \left(\sum_{k=1}^n |b_{r+1, k}| \right), \text{ where } t = \left\lfloor \frac{i}{n} \right\rfloor, r = i - tn. \tag{2}$$

Denoting $R_t := \sum_{j=1}^m |a_{tj}|$ and $S_r = \sum_{k=1}^n |b_{rk}|$ we can rewrite (2) as

$$\|A \otimes B\|_\infty = \max_{0 \leq tn+r \leq nm-1} R_{t+1} S_{r+1} = \max_{1 \leq t \leq m} R_t \max_{1 \leq r \leq n} S_r = \|A\|_\infty \|B\|_\infty.$$

Hence, the condition number satisfies by (1) that

$$\begin{aligned} \kappa_\infty(B_1 \otimes B_2) &= \|B_1 \otimes B_2\|_\infty \|(B_1 \otimes B_2)^{-1}\|_\infty \\ &= \|B_1\|_\infty \|B_2\|_\infty \|B_1^{-1}\|_\infty \|B_2^{-1}\|_\infty = \kappa_\infty(B_1) \kappa_\infty(B_2). \end{aligned}$$

By Corollary 2 of [2], we have that $\kappa_\infty(M_1) \leq \kappa_\infty(A_1)$ and $\kappa_\infty(M_2) \leq \kappa_\infty(A_2)$. So, we conclude that $\kappa_\infty(M_1 \otimes M_2) \leq \kappa_\infty(A_1 \otimes A_2)$. \square

2. Numerical tests

In this section two numerical examples illustrating the theoretical results will be presented. The first example will be constructed by performing the tensor product of three different NTP bases $u^n = (u_0^n, \dots, u_n^n)$ of the space $\mathcal{P}_n([0, 1])$ of polynomials of degree not greater than n , which were used in [2]. A second example will be presented considering the tensor product of rational bases $r^n = (r_0^n, \dots, r_n^n)$ constructed from the three NTP bases considered in the first example with positive weights and the tensor product of rational monomial bases (the monomial basis is TP in $[0, 1]$) also with positive weights. In fact, if u^n is a TP basis, it can be checked that the rational basis (r_0^n, \dots, r_n^n) , $r_i^n(x) = w_i u_i^n(x) / (\sum_{j=0}^n w_j u_j^n(x))$, with weights $w_i^n > 0$, is NTP. The basis $u^n = (b_0^n, \dots, b_n^n)$ formed by the Bernstein polynomials of degree n (see Example 6 a) in [2]) is the normalized B-basis of $\mathcal{P}_n([0, 1])$ and the corresponding rational Bernstein basis r_B^n defined by $r_i^n(x) = w_i b_i^n(x) / (\sum_{j=0}^n w_j b_j^n(x))$ with $w_i > 0$, $i = 0, \dots, n$, is the normalized B-basis of its spanned space $\langle r_B^n \rangle$.

We will also consider the Said–Ball basis $s^n = (s_0^n, \dots, s_n^n)$ and the DP basis $c^n = (c_0^n, \dots, c_n^n)$, which are both NTP basis. The Said–Ball basis (see [4]) is defined by

$$s_i^n(x) = \binom{\lfloor n/2 \rfloor + i}{i} x^i (1-x)^{\lfloor n/2 \rfloor + 1}, \quad 0 \leq i \leq \lfloor (n-1)/2 \rfloor,$$

$s_i^n(x) = s_{n-i}^n(1-x)$, $\lfloor n/2 \rfloor + 1 \leq i \leq n$, and, if n is even

$$s_{n/2}^n(x) = \binom{n}{n/2} x^{n/2} (1-x)^{n/2},$$

where $\lfloor m \rfloor$ is the greatest integer less than or equal to m . The DP basis is given by $c_0^n(x) = (1-x)^n$, $c_n^n(x) = x^n$, $c_i^n(x) = x(1-x)^{n-i}$, $1 \leq i \leq \lfloor n/2 \rfloor - 1$, $c_i^n(x) = x^i(1-x)$, $\lfloor (n+1)/2 \rfloor + 1 \leq i \leq n-1$, and, if n is even $c_{n/2}^n(x) = 1 - x^{\frac{n}{2}+1} - (1-x)^{\frac{n}{2}+1}$, and, if n is odd,

$$c_{\frac{n-1}{2}}^n(x) = x(1-x)^{\frac{n+1}{2}} + \frac{1}{2} \left[1 - x^{\frac{n+1}{2}+1} - (1-x)^{\frac{n+1}{2}+1} \right], \quad c_{\frac{n+1}{2}}^n(x) = c_{\frac{n-1}{2}}^n(1-x).$$

Let $(t_i^n)_{i=1}^{n+1}$ be the sequence of points given by $t_i = i/(n+2)$ for $i = 1, \dots, n+1$. Let us consider the Kronecker products of the collocation matrices of the Bernstein, Said–Ball and DP bases of $\mathcal{P}_n([0, 1])$ for $n = 3, 4, 5$ at $(t_i^n)_{i=1}^{n+1}$ by themselves: $M^n \otimes M^n$, $B_1^n \otimes B_1^n$ and $B_2^n \otimes B_2^n$, respectively. Then, the computation of the eigenvalues and the singular values of these matrices have been carried out with Mathematica using a precision of 100 digits. We can see the corresponding minimal eigenvalues and singular values in Table 1. It can be observed that the minimal eigenvalue, resp. singular value, of $M^n \otimes M^n$ is higher than the minimal eigenvalue, resp. singular value, of $B_1^n \otimes B_1^n$ and $B_2^n \otimes B_2^n$ as Theorem 1 has stated.

We have also computed $k_\infty(M^n \otimes M^n)$, $k_\infty(B_1^n \otimes B_1^n)$ and $k_\infty(B_2^n \otimes B_2^n)$ for $n = 3, 4, 5$. Table 2 shows the results. It can be observed that $k_\infty(M^n \otimes M^n) \leq k_\infty(B_i^n \otimes B_i^n)$ for $i = 1, 2$, as it has been shown in Theorem 1.

As it has been said before, the rational Said–Ball, DP and monomial bases with positive weights are NTP. Taking a sequence of positive weights $(w_i^n)_{i=0}^n$ and taking into account that $\sum_{j=0}^n w_j^n b_j^n(x) \in \mathcal{P}_n([0, 1])$ and

Table 1
The minimal eigenvalue and singular value of $M^n \otimes M^n$, $B_1^n \otimes B_1^n$ and $B_2^n \otimes B_2^n$.

n	$M^n \otimes M^n$		$B_1^n \otimes B_1^n$		$B_2^n \otimes B_2^n$	
	λ_{min}	σ_{min}	λ_{min}	σ_{min}	λ_{min}	σ_{min}
3	2.30e - 03	2.19e - 03	8.28e - 04	8.28e - 04	3.23e - 04	3.20e - 04
4	3.43e - 04	3.23e - 04	2.17e - 04	1.97e - 04	1.92e - 05	1.11e - 05
5	5.10e - 05	4.78e - 05	1.04e - 05	1.03e - 05	3.54e - 07	2.77e - 07

Table 2
Infinity condition number k_∞ of $M^n \otimes M^n$, $B_1^n \otimes B_1^n$ and $B_2^n \otimes B_2^n$.

n	$k_\infty(M^n \otimes M^n)$	$k_\infty(B_1^n \otimes B_1^n)$	$k_\infty(B_2^n \otimes B_2^n)$
3	5.1883e+02	1.7361e+03	7.1797e+03
4	3.9690e+03	6.5610e+03	1.6080e+05
5	2.5264e+04	1.3949e+05	6.0028e+06

Table 3
The minimal eigenvalue and singular value of M_T^n , $B_{1,T}^n$ and $B_{3,T}^n$.

n	M_T^n		$B_{1,T}^n$		$B_{3,T}^n$	
	λ_{min}	σ_{min}	λ_{min}	σ_{min}	λ_{min}	σ_{min}
3	1.95e - 03	1.74e - 03	4.06e - 04	3.78e - 04	4.39e - 06	3.82e - 6
4	2.57e - 04	2.05e - 04	1.30e - 04	1.09e - 04	8.86e - 08	2.35e - 08
5	4.75e - 05	4.36e - 05	8.83e - 06	8.66e - 06	2.60e - 10	1.63e - 10

Table 4
Infinity condition number k_∞ of M_T^n , $B_{1,T}^n$, $B_{2,T}^n$ and $B_{3,T}^n$.

n	$k_\infty(M_T^n)$	$k_\infty(B_{1,T}^n)$	$k_\infty(B_{2,T}^n)$	$k_\infty(B_{3,T}^n)$
3	8.1049e+02	5.6308e+03	3.5425e+04	5.8525e+05
4	7.1105e+03	1.3484e+04	2.0327e+06	1.3229e+08
5	3.1318e+04	1.6543e+05	4.0614e+07	1.7440e+10

that s^n , c^n and $m^n = (1, x, \dots, x^n)$ are bases of $\mathcal{P}_n([0, 1])$, then there exist three sequence of weights $(\bar{w}_i^n)_{i=0}^n$, $(\tilde{w}_i^n)_{i=0}^n$ and $(\hat{w}_i^n)_{i=0}^n$ satisfying

$$\sum_{j=0}^n w_j^n b_j^n(x) = \sum_{j=0}^n \bar{w}_j^n s_j^n(x) = \sum_{j=0}^n \tilde{w}_j^n b_j^n(x) = \sum_{j=0}^n \hat{w}_j^n c_j^n(x), \quad x \in [0, 1]. \tag{3}$$

Sequences of positive weights $(w_i^n)_{i=0}^n$ have been randomly generated for $n = 3, 4, 5$, where each w_i^n is an integer in the interval $[1, 1000]$, until we have obtained a sequence such that there exists positive sequences $(\bar{w}_i^n)_{i=0}^n$, $(\tilde{w}_i^n)_{i=0}^n$ and $(\hat{w}_i^n)_{i=0}^n$ satisfying (3). Then we have the normalized B-basis r_B , and the NTP rational bases of $\langle r_B \rangle$ corresponding to the Said–Ball basis, the DP basis and the monomial basis. So, in the second example we have considered the Kronecker products of the collocation matrices of the generated rational Bernstein, Said–Ball, DP and monomial bases for $n = 3, 4, 5$ at $(t_i^n)_{i=1}^{n+1}$ by themselves: $M_T^n = MR^n \otimes MR^n$, $B_{1,T}^n = BR_1^n \otimes BR_1^n$, $B_{2,T}^n = BR_2^n \otimes BR_2^n$ and $B_{3,T}^n = BR_3^n \otimes BR_3^n$, respectively. Then, the computation of the eigenvalues and the singular values of these matrices have been carried out with Mathematica using a precision of 100 digits. We can see the corresponding minimal eigenvalues and singular values of M_T^n , $B_{1,T}^n$ and $B_{3,T}^n$ in Table 3. It can be observed that the minimal eigenvalue, resp. singular value, of M_T^n is higher than the minimal eigenvalue, resp. singular value, of $B_{1,T}^n$ and $B_{3,T}^n$ as Theorem 1 has proved.

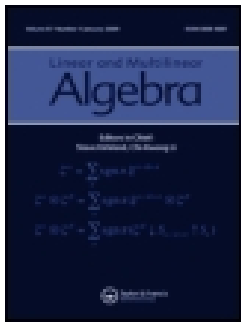
We have also computed $k_\infty(M_T^n)$, $k_\infty(B_{1,T}^n)$, $k_\infty(B_{2,T}^n)$ and $k_\infty(B_{3,T}^n)$ for $n = 3, 4, 5$ with Mathematica. The results can be seen in Table 4. It can be observed that $k_\infty(M_T^n) \leq k_\infty(B_{i,T}^n)$ for $i = 1, 2, 3$ (see Theorem 1).

References

- [1] J.M. Carnicer, J.M. Peña, Totally positive bases for shape preserving curve design and optimality of B-splines, *Comput. Aided Geom. Design* 11 (1994) 633–654.
- [2] J. Delgado, J.M. Peña, Extremal and optimal properties of B-bases collocation matrices, *Numer. Math.* 146 (2020) 105–118.
- [3] R.A. Horn, C.R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.
- [4] T.N.T. Goodman, H.B. Said, Shape preserving properties of the generalised ball basis, *Comput. Aided Geom. Design* 8 (1991) 115–121.

Article 9

- [21] J. Delgado, H. Orera and J. M. Peña. Characterizations and accurate computations for tridiagonal Toeplitz matrices, *Linear and Multilinear Algebra* (2021), Published online, DOI: 10.1080/03081087.2021.1884180.



Characterizations and accurate computations for tridiagonal Toeplitz matrices

Jorge Delgado , Héctor Orera & J. M. Peña

To cite this article: Jorge Delgado , Héctor Orera & J. M. Peña (2021): Characterizations and accurate computations for tridiagonal Toeplitz matrices, Linear and Multilinear Algebra

To link to this article: <https://doi.org/10.1080/03081087.2021.1884180>



Published online: 13 Feb 2021.



Submit your article to this journal [↗](#)




View related articles [↗](#)



View Crossmark data [↗](#)



Characterizations and accurate computations for tridiagonal Toeplitz matrices

Jorge Delgado^a, Héctor Orera ^b and J. M. Peña^b

^aDepartamento de Matemática Aplicada, Escuela de Ingeniería y Arquitectura de Zaragoza, Universidad de Zaragoza, Zaragoza, Spain; ^bDepartamento de Matemática Aplicada, Facultad de Ciencias, Universidad de Zaragoza, Zaragoza, Spain

ABSTRACT

Tridiagonal Toeplitz P -matrices, M -matrices and totally positive matrices are characterized. For some classes of tridiagonal matrices and tridiagonal Toeplitz matrices, it is shown that many algebraic computations can be performed with high relative accuracy.

ARTICLE HISTORY

Received 23 January 2020
Accepted 27 January 2021

COMMUNICATED BY

V. Olshevsky

KEYWORDS

Toeplitz matrices; tridiagonal matrices; high relative accuracy

2010 MATHEMATICS SUBJECT CLASSIFICATIONS

65F05; 65F15; 65G50; 15B05; 15A23

1. Introduction

Toeplitz matrices arise in many important applications, but they provide an example of a structured class of matrices for which it is not possible to perform some elementary algebraic computations with high relative accuracy (HRA). In fact, in [1] it was proved that the determinant of a general square Toeplitz matrix cannot be calculated with HRA. In contrast, for other classes of structured matrices, algorithms with HRA for many algebraic computations, in addition to the determinant, have been found. In this paper, we prove that, for some classes of tridiagonal Toeplitz matrices, many algebraic computations can be performed with HRA. Tridiagonal Toeplitz matrices arise in important applications, such as the solution of ordinary and partial differential equations, time series analysis or as regularization matrices in Tikhonov regularization for the solution of discrete ill-posed problems (see [2–7]). Recent results on the total positivity of some Toeplitz matrices and algorithms for determinants of tridiagonal periodic Toeplitz matrices can be seen in [8,9], respectively.

Let us now recall some concepts and notations used in this paper. Let A be a real matrix. We say that A is a nonnegative (positive) matrix and write $A \geq 0$ ($A > 0$) when all the

CONTACT Héctor Orera  hectororera@unizar.es  Departamento de Matemática Aplicada, Facultad de Ciencias, Universidad de Zaragoza, Zaragoza 50009, Spain

entries of A are nonnegative (positive). A square matrix is a P -matrix if all its principal minors are positive. Let us recall that in a Linear Complementarity Problem, very important in the field of Optimization, there always exists a unique solution if and only if the associated matrix is a P -matrix. Some subclasses of P -matrices are very important in many applications. For instance, nonsingular TP matrices. A matrix A is said to be *totally positive* (TP) if all its minors are nonnegative. If all its minors are positive, then A is called *strictly totally positive* (STP). TP and STP matrices arise in many applications in Approximation Theory, Statistics, Economy, Biology and ComputerAided Geometric Design, among other fields (see [10–12]). A real matrix A is a Z -matrix if all its off-diagonal entries are nonpositive. The matrix A is called an M -matrix if it can be expressed in the form $A = sI - B$, where I is the identity matrix, $B \geq 0$ and $s \geq \rho(B)$, where $\rho(B)$ is the spectral radius of B . If $s > \rho(B)$, then A is a nonsingular M -matrix. Equivalently, a Z -matrix A is a nonsingular M -matrix if and only if its inverse is nonnegative (see characterization (N_{38}) in Theorem (2.3) of [13, Ch. 6]). Nonsingular M -matrices arise in the discretization of partial differential equations and in many applications to Dynamic Systems, Economy and Optimization (see [13]). We call a square real matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ *sign symmetric* (*sign skew-symmetric*, respectively) if $a_{ij}a_{ji} \geq 0$ (≤ 0 , respectively) whenever $i \neq j$ and A is tridiagonal if $a_{ij} = 0$ whenever $|i - j| > 1$. Given a matrix $A = (a_{ij})_{1 \leq i, j \leq n}$, $|A| := (b_{ij})_{1 \leq i, j \leq n}$ denotes the matrix such that $b_{ij} := |a_{ij}|$ for all $1 \leq i, j \leq n$. An algorithm can be performed with HRA (independently of the conditioning of the problem) if all the included subtractions are of initial data, that is, if it only includes products, divisions, sums of numbers of the same sign and subtractions of the initial data (cf. [1,14,15]). A first step to obtain HRA algorithms for a class of matrices is an adequate parametrization of the matrices. Up to now, HRA algorithms for algebraic computations have been obtained for some subclasses of P -matrices, in particular for diagonally dominant M -matrices and for some subclasses of TP matrices (see, for instance, [14,16–22]). This paper shows that some classes of tridiagonal Toeplitz matrices can be added to the previous list.

The paper is organized as follows. Section 2 includes some auxiliary results and presents the Neville elimination and the bidiagonal factorization, which provide the parametrization of nonsingular TP matrices that can be used to apply the HRA algorithms of Koev (see [15,23,24]) for nonsingular TP matrices. With these algorithms and the mentioned parametrization, one can perform the following algebraic calculations with HRA: inverse, all singular values, all eigenvalues and the solution of some linear systems. These algorithms will be used in this paper to obtain HRA computations with some tridiagonal Toeplitz matrices. In Section 3, we introduce Toeplitz matrices and characterize tridiagonal Toeplitz TP matrices, tridiagonal Toeplitz M -matrices and tridiagonal Toeplitz P -matrices. Section 4 deals with sign skew-symmetric tridiagonal matrices with positive diagonal entries. It is shown that their leading principal minors and all minors of their inverses can be computed with HRA. In Section 5, a condition is provided to calculate the bidiagonal decomposition of sign symmetric tridiagonal Toeplitz P -matrices with HRA, and so their eigenvalues and singular values, as also illustrated with numerical experiments in Section 6. As shown in Figure 1, our results outperform those obtained with the usual MATLAB functions. Let us also recall that the eigenvalues of a tridiagonal Toeplitz matrix are already known (cf. page 59 of [7]), in contrast to singular values. Section 5 also provides the bidiagonal factorization of the inverse of a tridiagonal Toeplitz M -matrix.

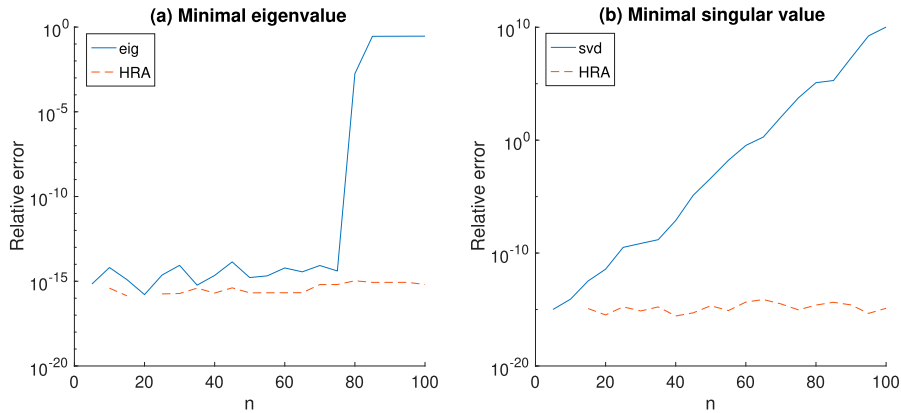


Figure 1. Relative error for the minimal eigenvalues and singular values of $A_5, A_{10}, \dots, A_{100}$.

2. Auxiliary results

Let us denote by $Q_{k,n}$ the set of strictly increasing sequences of k integers chosen from $\{1, \dots, n\}$. Let $\alpha = (\alpha_1, \dots, \alpha_k)$, $\beta = (\beta_1, \dots, \beta_k)$ be two sequences of $Q_{k,n}$. Then $A[\alpha|\beta]$ denotes the $k \times k$ submatrix of A formed using the rows numbered by $\alpha_1, \dots, \alpha_k$ and the columns numbered by β_1, \dots, β_k . Whenever $\alpha = \beta$, the submatrix $A[\alpha|\alpha]$ is called a principal submatrix and it is denoted by $A[\alpha]$, and $\det A[1, \dots, k]$ is called a leading principal minor of A . For each $\alpha \in Q_{k,n}$, the dispersion number $d(\alpha)$ is defined by

$$d(\alpha) := \alpha_k - \alpha_1 - (k - 1). \quad (1)$$

So, α consists of consecutive integers if and only if $d(\alpha) = 0$. Let $D = (d_{ij})_{1 \leq i, j \leq n}$ be a diagonal matrix, which can be denoted by $D = \text{diag}(d_1, \dots, d_n)$, where $d_i := d_{ii}$ for $i = 1, \dots, n$. Let us denote by $E_i(x)$, with $i = 2, \dots, n$, the $n \times n$ lower elementary bidiagonal matrix whose $(i, i - 1)$ entry is x :

$$E_i(x) = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & x & 1 & \\ & & & & \ddots \\ & & & & & 1 \end{pmatrix}. \quad (2)$$

In particular, $E_i(x)$ can be identified by its 2×2 principal submatrix using the rows and columns with indices $i - 1$ and i . This submatrix will be denoted by

$$\bar{E}_i(x) := (E_i(x))[i - 1, i], \quad i = 2, \dots, n. \quad (3)$$

The matrix $E_i^T(x) := (E_i(x))^T$ is called upper elementary bidiagonal matrix.

The following two results will allow us to characterize tridiagonal Toeplitz P -matrices. The next proposition characterizes a P -matrix in terms of the positivity of the real eigenvalues of its principal submatrices.

Proposition 2.1 (cf. 2.5.6.5 in p. 120 of [25]): *An $n \times n$ matrix A is a P -matrix if and only if every real eigenvalue of every principal submatrix of A is positive.*

The following theorem provides a sufficient condition for the total positivity of an $n \times n$ nonnegative tridiagonal matrix using only the positivity of $n-1$ minors.

Theorem 2.2 (Theorem 7 of [26]): *Let A be an $n \times n$ ($n \geq 3$) tridiagonal nonnegative matrix. If $\det A[1, \dots, k] > 0$ for $k \leq n-2$ and $\det A > 0$, then A is TP.*

Neville elimination (NE) has been very useful to characterize TP matrices and for parallel computations (cf. [26,27]). Neville elimination is an alternative procedure to Gaussian elimination that produces zeros in a column of a matrix by adding to each row an appropriate multiple of the previous one. Given a nonsingular matrix $A = (a_{ij})_{1 \leq i, j \leq n}$, the NE procedure consists of $n-1$ steps and leads to the following sequence of matrices:

$$A =: A^{(1)} \rightarrow \tilde{A}^{(1)} \rightarrow A^{(2)} \rightarrow \tilde{A}^{(2)} \rightarrow \dots \rightarrow A^{(n)} = \tilde{A}^{(n)} = U, \quad (4)$$

where U is an upper triangular matrix.

The matrix $\tilde{A}^{(k)} = (\tilde{a}_{ij}^{(k)})_{1 \leq i, j \leq n}$ is obtained from the matrix $A^{(k)} = (a_{ij}^{(k)})_{1 \leq i, j \leq n}$ by a row permutation that moves to the bottom the rows with a zero entry in column k below the main diagonal. For nonsingular TP matrices, it is always possible to perform NE without row exchanges (see [28]). If a row permutation is not necessary at the k th step, we have that $\tilde{A}^{(k)} = A^{(k)}$. The entries of $A^{(k+1)} = (a_{ij}^{(k+1)})_{1 \leq i, j \leq n}$ can be obtained from $\tilde{A}^{(k)} = (\tilde{a}_{ij}^{(k)})_{1 \leq i, j \leq n}$ using the formula:

$$a_{ij}^{(k+1)} = \begin{cases} \tilde{a}_{ij}^{(k)} - \frac{\tilde{a}_{ik}^{(k)}}{\tilde{a}_{i-1,k}^{(k)}} \tilde{a}_{i-1,j}^{(k)}, & \text{if } k \leq j < i \leq n \text{ and } \tilde{a}_{i-1,k}^{(k)} \neq 0, \\ \tilde{a}_{ij}^{(k)}, & \text{otherwise,} \end{cases} \quad (5)$$

for $k = 1, \dots, n-1$. The (i, j) pivot of the NE of A is given by

$$p_{ij} = \tilde{a}_{ij}^{(j)}, \quad 1 \leq j \leq i \leq n.$$

If $i = j$ we say that p_{ii} is a *diagonal pivot*. The (i, j) multiplier of the NE of A , with $1 \leq j \leq i \leq n$, is defined as

$$m_{ij} = \begin{cases} \frac{\tilde{a}_{ij}^{(j)}}{\tilde{a}_{i-1,j}^{(j)}} = \frac{p_{ij}}{p_{i-1,j}}, & \text{if } \tilde{a}_{i-1,j}^{(j)} \neq 0, \\ 0, & \text{if } \tilde{a}_{i-1,j}^{(j)} = 0. \end{cases}$$

The multipliers satisfy that

$$m_{ij} = 0 \Rightarrow m_{hj} = 0 \quad \forall h > i.$$

Nonsingular TP matrices can be expressed as a product of nonnegative bidiagonal matrices. The following theorem (see Theorem 4.2 and p. 120 of [29]) introduces this representation, which is called the *bidiagonal decomposition*.

Theorem 2.3 (cf. Theorem 4.2 of [29]): Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a nonsingular TP matrix. Then A admits the following representation:

$$A = F_{n-1}F_{n-2} \cdots F_1 D G_1 \cdots G_{n-2}G_{n-1}, \quad (6)$$

where D is the diagonal matrix $\text{diag}(p_{11}, \dots, p_{nn})$ with positive diagonal entries and F_i, G_i are the nonnegative bidiagonal matrices given by

$$F_i = \begin{pmatrix} 1 & & & & & & & & & \\ 0 & 1 & & & & & & & & \\ & & \ddots & & & & & & & \\ & & & \ddots & & & & & & \\ & & & & 0 & 1 & & & & \\ & & & & & m_{i+1,1} & 1 & & & \\ & & & & & & & \ddots & & \\ & & & & & & & & \ddots & \\ & & & & & & & & & m_{n,n-i} & 1 \end{pmatrix}, \quad (7)$$

$$G_i = \begin{pmatrix} 1 & 0 & & & & & & & & \\ & 1 & \ddots & & & & & & & \\ & & \ddots & & & & & & & \\ & & & \ddots & & & & & & \\ & & & & 0 & \tilde{m}_{i+1,1} & & & & \\ & & & & 1 & & \ddots & & & \\ & & & & & 1 & & \ddots & & \\ & & & & & & & \ddots & & \\ & & & & & & & & \ddots & \tilde{m}_{n,n-i} \\ & & & & & & & & & 1 \end{pmatrix}, \quad (8)$$

for all $i \in \{1, \dots, n-1\}$. If, in addition, the entries m_{ij} and \tilde{m}_{ij} satisfy

$$\begin{aligned} m_{ij} = 0 &\Rightarrow m_{hj} = 0 \quad \forall h > i, \\ \tilde{m}_{ij} = 0 &\Rightarrow \tilde{m}_{hj} = 0 \quad \forall h > i, \end{aligned} \quad (9)$$

then the decomposition is unique.

In the bidiagonal decomposition given by (6), (7) and (8), the entries m_{ij} and p_{ii} are the multipliers and diagonal pivots, respectively, corresponding to the NE of A (see Theorem 4.2 of [29] and the comment below it) and the entries \tilde{m}_{ij} are the multipliers of the NE of A^T (see p. 116 of [29]). In general, more classes of matrices can be represented as a product of bidiagonal matrices. The following remark shows which hypotheses of Theorem 2.3 are sufficient for the uniqueness of a representation following (6).

Remark 2.4: If we consider the factorization given by (6)–(9) without any further requirement than the nonsingularity of D , by Proposition 2.2 of [30] the uniqueness of (6) holds.

In [15], the following matrix notation $\mathcal{BD}(A)$ was introduced to represent the bidiagonal decomposition of a nonsingular TP matrix

$$(\mathcal{BD}(A))_{ij} = \begin{cases} m_{ij}, & \text{if } i > j, \\ \tilde{m}_{ji}, & \text{if } i < j, \\ p_{ii}, & \text{if } i = j. \end{cases} \quad (10)$$

Throughout this paper, $\mathcal{BD}(A)$ will denote the bidiagonal decomposition of a matrix under the hypotheses of Remark 2.4.

3. Characterizations of tridiagonal Toeplitz P -matrices

An $n \times n$ Toeplitz matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ is a real matrix such that all its diagonals are constant. These matrices can be defined through a sequence of $2n-1$ real numbers $\{\alpha_k\}_{-n+1}^{n-1}$ with

$$a_{ij} := \alpha_{i-j}, \quad 1 \leq i, j \leq n. \quad (11)$$

If an $n \times n$ Toeplitz matrix is also tridiagonal, it can be uniquely represented with 3 parameters,

$$T_n(a, b, c) := \begin{pmatrix} a & c & & & \\ b & a & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & c \\ & & & b & a \end{pmatrix}. \quad (12)$$

Given a positive matrix $A = (a_{ij})_{1 \leq i, j \leq n}$, the following condition is sufficient for its total positivity (see [31] or Section 2.6 of [12]):

$$a_{ij}a_{i+1, j+1} \geq 4 \cos^2 \left(\frac{\pi}{n+1} \right) a_{i, j+1}a_{i+1, j},$$

with $i, j = 1, \dots, n-1$. If all these inequalities are strict, then A is STP. In particular, given an $n \times n$ Toeplitz matrix (11) with $\alpha_i > 0$ for $i = -n+1, \dots, 0, \dots, n-1$, the sufficient condition for a positive matrix A to be TP presents the following form:

$$\alpha_i^2 \geq 4 \cos^2 \left(\frac{\pi}{n+1} \right) \alpha_{i-1} \alpha_{i+1},$$

with $i = -n+2, \dots, 0, \dots, n-2$. This condition requires the positivity of all the entries of the matrix. Nevertheless, we are going to prove that a similar condition (jointly with the nonnegativity of the parameters) is sufficient and also necessary for a tridiagonal Toeplitz matrix to be TP.

Proposition 3.1: *Let $A = T_n(a, b, c)$ be the tridiagonal Toeplitz matrix given by (12). Then A is TP if and only if*

$$a, b, c \geq 0, \quad a \geq 2\sqrt{bc} \cos \left(\frac{\pi}{n+1} \right). \quad (13)$$

Proof: It is known (see p. 59 of [7]) that the eigenvalues of the $n \times n$ tridiagonal Toeplitz matrix $T_n(a, b, c)$ are given by

$$\lambda_k = a + 2\sqrt{bc} \cos\left(\frac{k\pi}{n+1}\right), \quad k = 1, \dots, n. \quad (14)$$

Let us suppose that A is a TP matrix. Then $a, b, c \geq 0$ and its eigenvalues are real and nonnegative (see Corollary 5.5 of [12]). Moreover, since we know that the eigenvalues satisfy (14), it is sufficient to guarantee that the smallest eigenvalue, λ_n , is nonnegative:

$$\lambda_n = a + 2\sqrt{bc} \cos\left(\frac{n\pi}{n+1}\right) = a - 2\sqrt{bc} \cos\left(\frac{\pi}{n+1}\right) \geq 0,$$

or equivalently,

$$a \geq 2\sqrt{bc} \cos\left(\frac{\pi}{n+1}\right),$$

which is precisely (14) for $k = n$.

Let us now suppose that conditions (13) hold. We start with the case where the second inequality of (13) is strict. By Theorem 2.2, in order to prove that A is a TP matrix it is sufficient to check that its leading principal minors of order h are positive for $h = 1, \dots, n-2$ and that its determinant is also positive. Due to the structure of A we have that $A[1, \dots, h] = T_h(a, b, c)$ for $h = 1, \dots, n$. So, let us check the positivity of the minors by studying the positivity of the eigenvalues of the matrices $T_h(a, b, c)$ for $h = 1, \dots, n-2$ and for $h = n$. We can include the case $h = n-1$. Then the set of eigenvalues to check is given by $\lambda_{k,h} := a + 2\sqrt{bc} \cos\left(\frac{k\pi}{h+1}\right)$ with $1 \leq h \leq n$ and $k = 1, \dots, h$, where h represents the size of the $h \times h$ matrix whose eigenvalues are given by $\lambda_{k,h}$.

Since all the eigenvalues are real, it suffices to check that the smallest eigenvalue is positive in order to assure that $\lambda_{k,h} > 0$ for all $h = 1, \dots, n$ and for all $k = 1, \dots, h$:

$$\min_{k,h} \lambda_{k,h} = \lambda_{n,n} = a + 2\sqrt{bc} \cos\left(\frac{n\pi}{n+1}\right) = a - 2\sqrt{bc} \cos\left(\frac{\pi}{n+1}\right) > 0,$$

which is true by hypothesis. The value of the $h \times h$ leading principal minor of A is equal to the product of $\lambda_{1,h}, \dots, \lambda_{h,h}$, and so it is positive. By Theorem 2.2 A is TP, and so the case where the strict inequality holds is proven.

Let us finally consider the case where the second inequality of (13) holds as an equality, $a = 2\sqrt{bc} \cos\left(\frac{\pi}{n+1}\right)$, which corresponds to the singular case. Let us define the set of matrices $T_n(a + \epsilon, b, c)$ with $\epsilon > 0$. These matrices satisfy that $a + \epsilon > 2\sqrt{bc} \cos\left(\frac{\pi}{n+1}\right)$, and so they are TP because of the previous case where the second inequality of (13) was strict. Moreover, this set of matrices satisfies that $\lim_{\epsilon \rightarrow 0} T_n(a + \epsilon, b, c) = T_n(a, b, c)$, and so $T_n(a, b, c)$ is TP because the set of TP matrices is closed (let us recall that this fact is a direct consequence of the continuity of the determinant as a function of the matrix entries). ■

If we consider parameters b and c with nonpositive sign, we can deduce an analogous characterization for M -matrices of the form $T_n(a, b, c)$.

Corollary 3.2: Let $A = T_n(a, b, c)$ be the tridiagonal Toeplitz matrix given by (12). Then A is an M -matrix if and only if $a \geq 2\sqrt{bc} \cos(\frac{\pi}{n+1})$ and $b, c \leq 0$.

Proof: Since M -matrices are Z -matrices, the condition $b, c \leq 0$ is mandatory. Let us recall that a Z -matrix A is an M -matrix if and only if every real eigenvalue of A is nonnegative (see characterization (C_8) of Theorem (4.6) of [13, Ch. 6]). Since A is a tridiagonal Toeplitz matrix we know (see p. 59 of [7]) that its eigenvalues are real, distinct and that they are given by (14). Then we only need to check that the smallest eigenvalue, λ_n , is nonnegative:

$$\lambda_n = a + 2\sqrt{bc} \cos\left(\frac{n\pi}{n+1}\right) = a - 2\sqrt{bc} \cos\left(\frac{\pi}{n+1}\right) \geq 0,$$

which is true if and only if $a \geq 2\sqrt{bc} \cos(\frac{\pi}{n+1})$. ■

We now consider a third case of tridiagonal Toeplitz matrices $T_n(a, b, c)$ where the parameters satisfy $a > 0$ and $bc \leq 0$. This particular case, where the off-diagonal entries have opposite sign, verifies that $T_n(a, b, c)$ is a P -matrix without any further requirement. Moreover, the following result proves that all tridiagonal matrices with positive diagonal and with an analogous sign pattern are P -matrices.

Proposition 3.3: Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a tridiagonal matrix. If $a_{ii} > 0$ for $i = 1, \dots, n$ and $a_{i+1,i}a_{i,i+1} \leq 0$ for $i = 1, \dots, n-1$, then A is a P -matrix.

Proof: Let us first prove by induction that the leading principal minors of A , $\theta_k := \det A[1, \dots, k]$ for $k = 1, \dots, n$, are positive. It is straightforward to see that $\theta_1 = a_{11} > 0$ and that $\theta_2 = a_{11}a_{22} - a_{21}a_{12} > 0$. Let us suppose that $\theta_{k-1}, \theta_{k-2} > 0$ for some $k \in \{3, \dots, n\}$ and let us prove that $\theta_k > 0$. Since A is a tridiagonal matrix, using the Laplace expansion of a determinant we can write θ_k as

$$\theta_k = a_{kk}\theta_{k-1} - a_{k,k-1}a_{k-1,k}\theta_{k-2}, \quad (15)$$

and so $\theta_k > 0$ by the induction hypothesis. Now let us prove that all principal minors using consecutive rows and columns are positive. These minors are of the form $\det A[\alpha]$ with $\alpha = (s, \dots, r)$, $d(\alpha) = 0$ (see (1)) and $1 \leq s < r \leq n$. Given an index $1 \leq s \leq n$ we consider the principal submatrix $A_s := A[s, \dots, n]$. The matrix A_s is a tridiagonal matrix that satisfies the hypotheses of this proposition. Hence, we can apply the previous case to A_s and deduce that its leading principal minors are positive. These minors can be written as $\det A_s[1, \dots, p]$, with $1 \leq p \leq n - s + 1$, and, since A_s is a submatrix of A , these minors satisfy that $\det A_s[1, \dots, p] = \det A[s, \dots, p + s - 1] > 0$. Then we have as a direct consequence the positivity of all the principal minors using consecutive rows and columns. Finally, it only remains to study the principal minors $\det A[\alpha]$ such that $d(\alpha) > 0$. Given $\alpha \in Q_{k,n}$ with $d(\alpha) > 0$, let us consider the decomposition $\alpha = (\beta_1, \dots, \beta_r)$, with $|\beta_i| \geq 1$ and $d(\beta_i) = 0$ for $i = 1, \dots, r$, such that $d(\beta_j, \beta_{j+1}) > 0$ for all $j = 1, \dots, r-1$. Then $A[\alpha]$ is a block diagonal matrix such that the determinant of its i th block $A[\beta_i]$ is a principal minor of A using consecutive rows and columns, and hence, it is positive. So we conclude that $\det A[\alpha] = \det A[\beta_1] \cdots \det A[\beta_r] > 0$. ■

Observe that the previous result can be stated in the following way. A tridiagonal sign skew-symmetric matrix with positive diagonal entries is a P -matrix. We now characterize tridiagonal Toeplitz P -matrices.

Theorem 3.4: *Let $A = T_n(a, b, c)$ be the tridiagonal Toeplitz matrix given by (12). Then A is a P -matrix if and only if one of the following two conditions holds:*

- (i) $bc \leq 0$ and $a > 0$.
- (ii) $bc \geq 0$ and $a > 2\sqrt{bc} \cos(\frac{\pi}{n+1})$.

Proof: If (i) holds, then by Proposition 3.3 A is a P -matrix. Let us now suppose that condition (ii) holds. If $b, c \geq 0$, by Proposition 3.1, A is a nonsingular TP matrix, and hence, a P -matrix because, by Theorem 11.3 of [12], nonsingular TP matrices are P -matrices. If $b, c \leq 0$, by Corollary 3.2, A is a nonsingular M -matrix and so a P -matrix because, by characterization (A_1) of Theorem (2.3) of [13, Ch. 6], nonsingular M -matrices are P -matrices.

Assume now that A is a P -matrix. We have to see that if (i) does not hold, then (ii) holds. Since by definition $A[1, 1] = a > 0$, it is sufficient to consider parameters b, c such that $bc \geq 0$. By Proposition 2.1, the real eigenvalues of all the principal submatrices of A are positive. Given $\alpha \in Q_{h,n}$, $A[\alpha] = T_h(a, b, c)$ whenever $d(\alpha) = 0$. If $d(\alpha) > 0$, then we can consider the decomposition $\alpha = (\beta_1, \dots, \beta_r)$, with $|\beta_i| \geq 1$ and with $d(\beta_i) = 0$ for $i = 1, \dots, r$, such that $d(\beta_j, \beta_{j+1}) > 0$ for all $j = 1, \dots, r - 1$. Then $A[\alpha]$ is a block diagonal matrix such that its i th block $A[\beta_i]$ is the tridiagonal Toeplitz matrix $T_{|\beta_i|}(a, b, c)$. In either case, the eigenvalues of $A[\alpha]$, $\alpha \in Q_{h,n}$, are included in the set $\lambda_{r,h} := a + 2\sqrt{bc} \cos(\frac{r\pi}{h+1})$ with $1 \leq h \leq n$ and with $r = 1, \dots, h$, where h represents the size of the $h \times h$ matrix whose eigenvalues are given by $\lambda_{r,h}$ (see p. 59 of [7]). Therefore, $\lambda_{r,h} > 0$ for all $h = 1, \dots, n$ and $r = 1, \dots, h$. In particular, $\min_{r,h} \lambda_{r,h} = \lambda_{n,n} > 0$, and hence, $a > 2\sqrt{bc} \cos(\frac{\pi}{n+1})$ and the result holds. ■

Remark 3.5: From Proposition 3.1 and Theorem 3.4, we deduce that a tridiagonal Toeplitz matrix $T_n(a, b, c)$ is a nonsingular TP matrix if and only if $a > 2\sqrt{bc} \cos(\frac{\pi}{n+1})$ and $b, c \geq 0$. Analogously, from Corollary 3.2 and Theorem 3.4, we deduce that a tridiagonal Toeplitz matrix $T_n(a, b, c)$ is a nonsingular M -matrix if and only if $a > 2\sqrt{bc} \cos(\frac{\pi}{n+1})$ and $b, c \leq 0$. Then, by Theorem 3.4 a sign symmetric tridiagonal Toeplitz P -matrix is either a nonsingular TP matrix or a nonsingular M -matrix. Besides, taking into account that a tridiagonal Toeplitz matrix is either sign symmetric or sign skew-symmetric, we can reformulate Theorem 3.4 in the following way. A tridiagonal Toeplitz matrix $A = T_n(a, b, c)$ is a P -matrix if and only if $a > 0$ and, if A is sign symmetric, then $a > 2\sqrt{bc} \cos(\frac{\pi}{n+1})$.

In Theorem 3.4 (ii), the condition $a > 2\sqrt{bc} \cos(\frac{\pi}{n+1})$ (or analogously, $a^2 > 4bc \cos^2(\frac{\pi}{n+1})$) has been used to characterize tridiagonal Toeplitz P -matrices. If this condition is satisfied independently of n , we obtain the new condition $a^2 > 4bc$. In fact, this inequality will play a key role in Section 5 since it is used in order to assure HRA for some computations with the matrices $T_n(a, b, c)$. In fact, the positive number $a^2 - 4bc$ will be an additional natural parameter to assure the HRA. The case (i) of Theorem 3.4 will be considered in a more general framework in the following section.

4. Computing with HRA the minors of sign skew-symmetric tridiagonal matrices with positive diagonal entries

Whenever a tridiagonal matrix A satisfies the hypotheses of Proposition 3.3 (sign skew-symmetric with positive diagonal entries), it is possible to compute its bidiagonal decomposition accurately. Moreover, the bidiagonal decomposition allows us to compute all its minors and its inverse with HRA. The following result provides the $\mathcal{BD}(A)$ for such A .

Proposition 4.1: *Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a tridiagonal matrix such that $a_{ii} > 0$ for $i = 1, \dots, n$ and $a_{i+1,i}a_{i,i+1} \leq 0$ for $i = 1, \dots, n-1$. Then*

$$\mathcal{BD}(A) = \begin{pmatrix} \delta_1 & \frac{a_{12}}{\delta_1} & & & & \\ \frac{a_{21}}{\delta_1} & \delta_2 & \ddots & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \frac{a_{n-1,n}}{\delta_{n-1}} & \\ & & & \frac{a_{n,n-1}}{\delta_{n-1}} & \delta_n & \\ & & & & & \delta_n \end{pmatrix}, \quad (16)$$

where δ_i are the diagonal pivots associated to the NE of A . The diagonal pivots satisfy the following recurrence relation:

$$\delta_1 = a_{11}, \quad \delta_i = a_{ii} - \frac{a_{i,i-1}a_{i-1,i}}{\delta_{i-1}} \quad i = 2, \dots, n. \quad (17)$$

If we know the entries of A with HRA, then we can compute $\mathcal{BD}(A)$ (16) to HRA, and hence, the leading principal minors of A to HRA.

Proof: Clearly, for tridiagonal P -matrices, no row exchanges are needed in Neville elimination and Gauss elimination, which coincide. Hence, by Proposition 3.3, $\delta_1, \dots, \delta_n$ are also the pivots of the Gauss elimination of A and it is well known that they satisfy that

$$\delta_k = \frac{\theta_k}{\theta_{k-1}}, \quad k = 1, \dots, n, \quad (18)$$

with $\theta_0 := 1$ and $\theta_k := A[1, \dots, k]$ for $k = 1, \dots, n$. From (15) and (18), we deduce (17). Since $a_{i+1,i}a_{i,i+1} \leq 0$ for $i = 1, \dots, n-1$, the diagonal pivots can be computed by (17) without performing any subtraction. As a consequence, all pivots δ_k are computed to HRA. The leading principal minors can be obtained with HRA through the computation $\theta_k = \delta_1 \cdots \delta_k$, for $k = 1, \dots, n$. \blacksquare

Proposition 4.1 allows us to prove that some computations can be performed with HRA, as the following result shows.

Theorem 4.2: *Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a tridiagonal matrix such that $a_{ii} > 0$ for $i = 1, \dots, n$ and $a_{i+1,i}a_{i,i+1} \leq 0$ for $i = 1, \dots, n-1$. Then all the minors and the inverse of A can be computed to HRA.*

Proof: By Proposition 4.1, we can compute the leading principal minors of A to HRA. Following the proof of Proposition 3.3, it can be deduced that all the principal minors of A can be obtained without subtractions, and so, with HRA. Given $\alpha = (i_1, \dots, i_k), \beta = (j_1, \dots, j_k) \in Q_{k,n}$, if $|i_r - j_r| \geq 2$ for any $r = 1, \dots, k$ then $\det A[i_1, \dots, i_k | j_1, \dots, j_k] = 0$ and if $|i_s - j_s| = 1$ for an index $s = 1, \dots, k$, then $A[\alpha | \beta] = A[i_1, \dots, i_{s-1} | j_1, \dots, j_{s-1}] a_{i_s, j_s} A[i_{s+1}, \dots, i_k | j_{s+1}, \dots, j_k]$. Hence, any nonzero minor of a tridiagonal matrix can be written as a product of off-diagonal entries and principal minors using consecutive rows and columns. Then all the entries of A^{-1} can be computed to HRA as a consequence. For example, by using formula (1.33) of [10], corresponding to the well-known expression of the entries of the inverse in terms of determinants. ■

An alternative HRA method to obtain A^{-1} is presented in the following remark.

Remark 4.3: Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a nonsingular tridiagonal matrix. Then, by (47) of [32], we can give the following explicit expression of the entries of $A^{-1} := (b_{ij})_{1 \leq i, j \leq n}$ in terms of principal minors of A using consecutive rows and columns. In fact,

$$b_{ij} = \begin{cases} \frac{\theta_{i-1} \widehat{\theta}_{n-j}}{\theta_n} \prod_{l=i}^{j-1} -a_{l, l+1}, & \text{for } i < j, \\ \frac{\theta_{i-1} \widehat{\theta}_{n-i}}{\theta_n}, & \text{for } i = j, \\ \frac{\theta_{j-1} \widehat{\theta}_{n-i}}{\theta_n} \prod_{l=j}^{i-1} -a_{l+1, l}, & \text{for } i > j, \end{cases} \quad (19)$$

where $\widehat{\theta}_k := \det A[n-k+1, \dots, n]$ for $k = 1, \dots, n$. If $a_{ii} > 0$ for $i = 1, \dots, n$ and $a_{i+1, i} a_{i, i+1} \leq 0$ for $i = 1, \dots, n-1$, then A^{-1} can also be computed to HRA by (19).

5. Computations with sign symmetric tridiagonal Toeplitz P -matrices with HRA

In this section, we guarantee the HRA for the bidiagonal decomposition, and so for many other algebraic computations, in the case of sign symmetric tridiagonal Toeplitz P -matrices with the additional parameter $a^2 - 4bc$ commented at the end of Section 3. By Theorem 3.4 and Remark 3.5, the P -matrices corresponding to this case are either nonsingular M -matrices or nonsingular TP matrices.

From now on, we assume that the parameters a, b, c are always positive:

$$a, b, c > 0.$$

Let us recall that the inverse of a nonsingular tridiagonal M -matrix is TP (see [33]). We are going to obtain the bidiagonal decomposition of an M -matrix $A = T_n(a, -b, -c)$. From the $\mathcal{BD}(A)$ obtained in Theorem 5.1, in Theorem 5.5 we shall deduce $\mathcal{BD}(A^{-1})$. Besides, by Remark 5.4, if we know $\mathcal{BD}(A)$ to HRA, then we can also perform many algebraic computations with A to HRA.

It is well known (see p. 99 of [12]) that the principal minors of a tridiagonal matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ satisfy:

$$\begin{aligned} \det A &= \det A[1, \dots, i] \det A[i+1, \dots, n] \\ &\quad - a_{i,i+1} a_{i+1,i} \det A[1, \dots, i-1] \det A[i+2, \dots, n]. \end{aligned} \quad (20)$$

From (20), we deduce that the leading principal minors of a tridiagonal Toeplitz matrix A , $\theta_j := \det A[1, \dots, j]$ with $j = 1, \dots, n$, satisfy the following relation:

$$\theta_n = \theta_j \theta_{n-j} - bc \theta_{j-1} \theta_{n-j-1}, \quad \text{with } \theta_{-1} = 0, \theta_0 = 1, j = 1, \dots, n. \quad (21)$$

Theorem 5.1: Let $A = T_n(a, -b, -c)$ be a nonsingular M -matrix given by (12). Then

$$\mathcal{BD}(A) = \begin{pmatrix} \delta_1 & -\frac{c}{\delta_1} & & & & \\ -\frac{b}{\delta_1} & \delta_2 & \ddots & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & -\frac{c}{\delta_{n-1}} & \\ & & & -\frac{b}{\delta_{n-1}} & \delta_n & \end{pmatrix}, \quad (22)$$

where δ_i are the diagonal pivots associated to the NE of A and are given by:

$$\delta_1 = a, \quad \delta_i = a - \frac{bc}{\delta_{i-1}} \quad \text{with } i = 2, \dots, n. \quad (23)$$

Moreover, if we know a, b, c with HRA and $a^2 - 4bc$ is a positive number known with HRA, then we can compute $\mathcal{BD}(A)$ (22) to HRA.

Proof: Since nonsingular M -matrices are P -matrices (see characterization (A_1) of Theorem 2.3 of [13, Ch. 6]), the principal minors of A are positive, and so, $\theta_i > 0$ for $i = 1, \dots, n$. Since A is a tridiagonal Toeplitz matrix, its leading principal minors satisfy

$$\theta_i = a\theta_{i-1} - bc\theta_{i-2}, \quad \text{with } \theta_{-1} = 0, \theta_0 = 1, i = 1, \dots, n \quad (24)$$

by (15). Moreover, (23) is a consequence of (24) and (18). There is an explicit expression for the leading principal minors of A (see p. 15 of [34]):

$$\theta_i = (\sqrt{bc})^i U_i \left(\frac{a}{2\sqrt{bc}} \right), \quad (25)$$

where $U_i(x)$ is the i th Chebyshev polynomial of the second kind. We can evaluate $U_i(x)$ through (see Section 3 of [34]):

$$\begin{aligned} U_i(x) &= \frac{r_+^{i+1}(x) - r_-^{i+1}(x)}{r_+(x) - r_-(x)}, \\ &\quad \text{with } r_+(x) := x + \sqrt{x^2 - 1} \text{ and } r_-(x) := x - \sqrt{x^2 - 1}. \end{aligned} \quad (26)$$

Let us denote $s_+ := r_+(\frac{a}{2\sqrt{bc}})$ and $s_- := r_-(\frac{a}{2\sqrt{bc}})$. By (25) and (27), we can write the pivots δ_i as:

$$\delta_i = \frac{\theta_i}{\theta_{i-1}} = \sqrt{bc} \frac{s_+^{i+1} - s_-^{i+1}}{s_+^i - s_-^i} = \sqrt{bc} s_+ \frac{1 + \frac{s_-}{s_+} + \dots + \frac{s_-^i}{s_+^i}}{1 + \frac{s_-}{s_+} + \dots + \frac{s_-^{i-1}}{s_+^{i-1}}}.$$

If we obtain s_+ and $\frac{s_-}{s_+}$ with HRA, then we can compute δ_i for $i = 1, \dots, n$ to HRA, and as a direct consequence, $\mathcal{BD}(A)$ to HRA. We can compute s_+ by

$$s_+ = \frac{a}{2\sqrt{bc}} + \sqrt{\frac{a^2}{4bc} - 1} = \frac{a + \sqrt{a^2 - 4bc}}{2\sqrt{bc}},$$

and the quotient $\frac{s_-}{s_+}$ by

$$\frac{s_-}{s_+} = \frac{s_- s_+}{s_+^2} = \frac{4bc}{2a^2 - 4bc + 2a\sqrt{a^2 - 4bc}} = \frac{4bc}{a^2 + (a^2 - 4bc) + 2a\sqrt{a^2 - 4bc}}.$$

Since $a^2 - 4bc$ is known with HRA by hypothesis, $\mathcal{BD}(A)$ can be obtained with HRA. \blacksquare

Remark 5.2: The computational cost of obtaining $\mathcal{BD}(A)$ following Theorem 5.1 is of $6n$ elementary operations, as can be checked from its proof.

Corollary 5.3: Let $A := T_n(a, -b, -c)$ be a nonsingular M -matrix. If we know a, b, c and $a - 2 \max\{b, c\}$ with HRA and $a - 2 \max\{b, c\} \geq 0$, then we can compute $\mathcal{BD}(A)$ (22) to HRA.

Proof: Without loss of generality, let us suppose that $b \geq c$. Then we can write the quantity $a^2 - 4bc$ as

$$a^2 - 4bc = (a - 2b)(a + 2c) + 2a(b - c). \quad (27)$$

Taking into account that, by hypothesis, $a - 2b$ is known to HRA and that $b - c$ is a subtraction of initial data, $a^2 - 4bc$ can also be computed to HRA. As a consequence, s_+ and $\frac{s_-}{s_+}$ can be obtained with HRA. Finally, following the proof of Theorem 5.1 we can compute $\mathcal{BD}(A)$ to HRA. \blacksquare

The bidiagonal decomposition of a nonsingular M -matrix is unique by Remark 2.4. If A is a tridiagonal M -matrix, then $\mathcal{BD}(A)$ allows us to perform some algebraic computations with A to HRA.

Remark 5.4: Let A be a tridiagonal Toeplitz M -matrix such that we know $\mathcal{BD}(A)$ to HRA. In this case, we also know the bidiagonal decomposition to HRA of $|A| = J_n A J_n$, where $J_n = \text{diag}(1, -1, \dots, (-1)^{n-1})$. Since $|A|$ is TP by Proposition 3.1, we can apply the HRA algorithms for TP matrices to $\mathcal{BD}(|A|) = |\mathcal{BD}(A)|$. For instance, in Section 6, we comment how to compute the singular values and eigenvalues of A to HRA.

The following result provides the bidiagonal decomposition of the inverse of a nonsingular tridiagonal Toeplitz M -matrix.

Theorem 5.5: Let $A = T_n(a, -b, -c)$ be a nonsingular M -matrix. Then A^{-1} is a TP matrix and

$$\mathcal{BD}(A^{-1}) = \begin{pmatrix} 1/\delta_n & c/\delta_{n-1} & c/\delta_{n-2} & \cdots & c/\delta_1 \\ b/\delta_{n-1} & 1/\delta_{n-1} & 0 & \cdots & 0 \\ b/\delta_{n-2} & 0 & 1/\delta_{n-2} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ b/\delta_1 & 0 & \cdots & 0 & 1/\delta_1 \end{pmatrix}, \quad (28)$$

where δ_i are the diagonal pivots associated to the NE of A for $i = 1, \dots, n$.

Proof: A^{-1} is TP because it is the inverse of a tridiagonal M -matrix (see Theorem 2.2 of [33]). Let us define $D := \text{diag}(\delta_1, \dots, \delta_n)$. By Theorem 5.1, we can write A as

$$A = E_2 \begin{pmatrix} -b \\ \delta_1 \end{pmatrix} \cdots E_n \begin{pmatrix} -b \\ \delta_{n-1} \end{pmatrix} D E_n^T \begin{pmatrix} -c \\ \delta_{n-1} \end{pmatrix} \cdots E_2^T \begin{pmatrix} -c \\ \delta_1 \end{pmatrix},$$

and so

$$A^{-1} = E_2^T \begin{pmatrix} c \\ \delta_1 \end{pmatrix} \cdots E_n^T \begin{pmatrix} c \\ \delta_{n-1} \end{pmatrix} D^{-1} E_n \begin{pmatrix} b \\ \delta_{n-1} \end{pmatrix} \cdots E_2 \begin{pmatrix} b \\ \delta_1 \end{pmatrix}. \quad (29)$$

The factorization (29) is different from the bidiagonal decomposition (6). In order to obtain $\mathcal{BD}(A^{-1})$ from (29), we first need to rewrite $E_n^T \begin{pmatrix} c \\ \delta_{n-1} \end{pmatrix} D^{-1} E_n \begin{pmatrix} b \\ \delta_{n-1} \end{pmatrix}$ as the product of a lower elementary bidiagonal matrix $E_n(\alpha)$, a diagonal matrix and an upper elementary bidiagonal matrix $E_n^T(\beta)$, with $\alpha, \beta \in \mathbb{R}$.

Let us start by computing the following product:

$$E_n^T \begin{pmatrix} c \\ \delta_{n-1} \end{pmatrix} D^{-1} E_n \begin{pmatrix} b \\ \delta_{n-1} \end{pmatrix} = \begin{pmatrix} \frac{1}{\delta_1} & & & & \\ & \ddots & & & \\ & & \frac{1}{\delta_{n-2}} & & \\ & & & \frac{1}{\delta_{n-1}} + \frac{bc}{\delta_{n-1}^2 \delta_n} & \frac{c}{\delta_{n-1} \delta_n} \\ & & & \frac{b}{\delta_{n-1} \delta_n} & \frac{1}{\delta_n} \end{pmatrix}. \quad (30)$$

By (23) for $i = 1, n$, the $(n-1, n-1)$ entry of (30) can be written as

$$\frac{1}{\delta_{n-1}} + \frac{bc}{\delta_{n-1}^2 \delta_n} = \frac{1}{\delta_{n-1} \delta_n} \left(\delta_n + \frac{bc}{\delta_{n-1}} \right) = \frac{1}{\delta_{n-1} \delta_n} \left(a - \frac{bc}{\delta_{n-1}} + \frac{bc}{\delta_{n-1}} \right) = \frac{\delta_1}{\delta_{n-1} \delta_n}.$$

The effect of the matrices $E_n(\alpha)$, $E_n^T(\beta)$ over D is restricted to the submatrix $D^{-1}[n-1, n]$. Hence, using the notation (3), we can decompose the principal submatrix of (30) using the

$n-1$, n rows as

$$\begin{pmatrix} \frac{\delta_1}{\delta_{n-1}\delta_n} & \frac{c}{\delta_{n-1}\delta_n} \\ \frac{\delta_{n-1}\delta_n}{b} & \frac{1}{\delta_n} \end{pmatrix} = \bar{E}_n \begin{pmatrix} b \\ \delta_1 \end{pmatrix} \begin{pmatrix} \frac{\delta_1}{\delta_{n-1}\delta_n} & \\ & \frac{1}{\delta_n} - \frac{cb}{\delta_{n-1}\delta_n\delta_1} \end{pmatrix} \bar{E}_n^T \begin{pmatrix} c \\ \delta_1 \end{pmatrix}. \quad (31)$$

Then by (31) we have that the required elementary matrices are $E_n(\frac{b}{\delta_1})$ and $E_n^T(\frac{c}{\delta_1})$. Moreover, using again (23) we can write the last entry of the diagonal matrix in (31) as

$$\frac{1}{\delta_n} - \frac{cb}{\delta_{n-1}\delta_n\delta_1} = \frac{1}{\delta_1\delta_n} \left(\delta_1 - \frac{bc}{\delta_{n-1}} \right) = \frac{\delta_n}{\delta_1\delta_n} = \frac{1}{\delta_1}.$$

If we denote by $D^{(2)} := \text{diag}(\delta_1^{-1}, \dots, \delta_{n-2}^{-1}, \frac{\delta_1}{\delta_n\delta_{n-1}}, \delta_1^{-1})$, then, by (31), we have that

$$E_n \begin{pmatrix} b \\ \delta_1 \end{pmatrix} D^{(2)} E_n^T \begin{pmatrix} c \\ \delta_1 \end{pmatrix} = E_n^T \begin{pmatrix} c \\ \delta_{n-1} \end{pmatrix} D^{-1} E_n \begin{pmatrix} b \\ \delta_{n-1} \end{pmatrix},$$

and so we have achieved our first goal. Let us now express A^{-1} as the following matrix product:

$$A^{-1} = E_2^T \begin{pmatrix} c \\ \delta_1 \end{pmatrix} \cdots E_{n-1}^T \begin{pmatrix} c \\ \delta_{n-2} \end{pmatrix} E_n \begin{pmatrix} b \\ \delta_1 \end{pmatrix} D^{(2)} E_n^T \begin{pmatrix} c \\ \delta_1 \end{pmatrix} E_{n-1} \begin{pmatrix} b \\ \delta_{n-2} \end{pmatrix} \cdots E_2 \begin{pmatrix} b \\ \delta_1 \end{pmatrix}. \quad (32)$$

Since the elementary bidiagonal matrices satisfy that $E_j(\alpha_j)E_n^T(\alpha_n) = E_n^T(\alpha_n)E_j(\alpha_j)$ whenever $j < n$, we can reorder the matrices in (32) and deduce that

$$A^{-1} = E_n \begin{pmatrix} b \\ \delta_1 \end{pmatrix} E_2^T \begin{pmatrix} c \\ \delta_1 \end{pmatrix} \cdots E_{n-1}^T \begin{pmatrix} c \\ \delta_{n-2} \end{pmatrix} D^{(2)} E_{n-1} \begin{pmatrix} b \\ \delta_{n-2} \end{pmatrix} \cdots E_2 \begin{pmatrix} b \\ \delta_1 \end{pmatrix} E_n^T \begin{pmatrix} c \\ \delta_1 \end{pmatrix}. \quad (33)$$

After rearranging the matrices we arrive at an analogous problem to (30). Hence, our aim is now expressing $E_{n-1}^T(\frac{c}{\delta_{n-2}})D^{(2)}E_{n-1}(\frac{b}{\delta_{n-2}})$ as the product of a matrix $E_{n-1}(\alpha)$, a diagonal matrix that will be denoted by $D^{(3)}$ and a matrix $E_{n-1}^T(\beta)$. Then we could rearrange again the elementary bidiagonal matrices as we did in (33). In general, after performing this procedure $k-1$ times we would obtain the following factorization:

$$\begin{aligned} A^{-1} &= E_n \begin{pmatrix} b \\ \delta_1 \end{pmatrix} \cdots E_{n-k+2} \begin{pmatrix} b \\ \delta_{k-1} \end{pmatrix} E_2^T \begin{pmatrix} c \\ \delta_1 \end{pmatrix} \cdots E_{n-k+1}^T \begin{pmatrix} c \\ \delta_{n-k} \end{pmatrix} D^{(k)} \\ &\quad \cdot E_{n-k+1} \begin{pmatrix} b \\ \delta_{n-k} \end{pmatrix} \cdots E_2 \begin{pmatrix} b \\ \delta_1 \end{pmatrix} E_{n-k+2}^T \begin{pmatrix} c \\ \delta_{k-1} \end{pmatrix} \cdots E_n^T \begin{pmatrix} c \\ \delta_1 \end{pmatrix}, \end{aligned} \quad (34)$$

where $D^{(k)} = \text{diag}(\delta_1^{-1}, \dots, \delta_{n-k}^{-1}, \frac{\theta_{k-1}\theta_{n-k}}{\theta_n}, \delta_{k-1}^{-1}, \dots, \delta_1^{-1})$.

When $k = n$, (34) coincides with the decomposition (28). Therefore, let us prove that (34) holds by induction on $k \in \{2, \dots, n\}$. We have already checked the first step, $k = 2$. So, let us assume that (34) holds for $k \in \{2, \dots, n-1\}$ and let us prove that it also

holds for $k + 1$. Let us first see that

$$E_{n-k+1}^T \begin{pmatrix} c \\ \delta_{n-k} \end{pmatrix} D^{(k)} E_{n-k+1} \begin{pmatrix} b \\ \delta_{n-k} \end{pmatrix} = E_{n-k+1} \begin{pmatrix} b \\ \delta_k \end{pmatrix} D^{(k+1)} E_{n-k+1}^T \begin{pmatrix} c \\ \delta_{n-k} \end{pmatrix}.$$

In general, the effect of the matrices $E_{n-k+1}(\alpha)$, $E_{n-k+1}^T(\beta)$ is restricted to the submatrix $D^{(k)}[n - k, n - k + 1]$. Using the notation (3), we deduce that

$$\begin{aligned} & \bar{E}_{n-k+1}^T \begin{pmatrix} c \\ \delta_{n-k} \end{pmatrix} \begin{pmatrix} \frac{1}{\delta_{n-k}} & \\ & \frac{\theta_{k-1}\theta_{n-k}}{\theta_n} \end{pmatrix} \bar{E}_{n-k+1} \begin{pmatrix} b \\ \delta_{n-k} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\theta_{n-k-1}}{\theta_{n-k}} + \frac{bc\theta_{k-1}\theta_{n-k-1}^2}{\theta_n\theta_{n-k}} & \frac{\theta_{k-1}\theta_{n-k-1}}{\theta_n} c \\ \frac{\theta_{k-1}\theta_{n-k-1}}{\theta_n} b & \frac{\theta_{k-1}\theta_{n-k}}{\theta_n} \end{pmatrix}. \end{aligned} \quad (35)$$

By (21) with $j = k$, the first diagonal entry of (36) can be written as

$$\begin{aligned} & \frac{\theta_{n-k-1}\theta_n + bc\theta_{k-1}\theta_{n-k-1}^2}{\theta_n\theta_{n-k}} = \frac{\theta_{n-k-1}^2}{\theta_n\theta_{n-k}} \left(\frac{\theta_n}{\theta_{n-k+1}} + bc\theta_{k-1} \right) \\ &= \frac{\theta_{n-k-1}^2}{\theta_n\theta_{n-k}} \left(\theta_k \frac{\theta_{n-k}}{\theta_{n-k-1}} - bc\theta_{k-1} + bc\theta_{k-1} \right) = \frac{\theta_k\theta_{n-k-1}}{\theta_n}. \end{aligned}$$

Applying Gauss elimination to the submatrix (35) we obtain, by (18), the following multiplier

$$\frac{\theta_{k-1}\theta_{n-k-1}\theta_n}{\theta_{n-k-1}\theta_k\theta_n} b = \frac{\theta_{k-1}}{\theta_k} b = \frac{b}{\delta_k}.$$

Analogously, applying Gauss elimination to the transpose of that submatrix we obtain the multiplier $\frac{c}{\delta_k}$. Hence, we can decompose (35) as

$$\bar{E}_{n-k+1} \begin{pmatrix} b \\ \delta_k \end{pmatrix} \begin{pmatrix} \frac{\theta_k\theta_{n-k-1}}{\theta_n} & \\ & \frac{\theta_{k-1}\theta_{n-k}}{\theta_n} - bc \frac{\theta_{k-1}\theta_{n-k-1}}{\theta_n\delta_k} \end{pmatrix} \bar{E}_{n-k+1}^T \begin{pmatrix} c \\ \delta_k \end{pmatrix}. \quad (36)$$

Using (21), we express the last entry of the diagonal matrix in (36) in terms of the diagonal pivots

$$\frac{\theta_{k-1}}{\theta_n\delta_k} \left(\theta_{n-k} \frac{\theta_k}{\theta_{k-1}} - bc\theta_{n-k-1} \right) = \frac{\theta_n}{\theta_n\delta_k} = \frac{1}{\delta_k}.$$

Then we have deduced that $D^{(k+1)} = \text{diag}(\delta_1^{-1}, \dots, \delta_{n-k-1}^{-1}, \frac{\theta_k\theta_{n-k-1}}{\theta_n}, \delta_k^{-1}, \dots, \delta_1^{-1})$, and so we can factorize A^{-1} as

$$A^{-1} = E_n \begin{pmatrix} b \\ \delta_1 \end{pmatrix} \cdots E_{n-k+2} \begin{pmatrix} b \\ \delta_{k-1} \end{pmatrix} E_2^T \begin{pmatrix} c \\ \delta_1 \end{pmatrix} \cdots E_{n-k}^T \begin{pmatrix} c \\ \delta_{n-k-1} \end{pmatrix} E_{n-k+1} \begin{pmatrix} b \\ \delta_k \end{pmatrix}$$

$$\cdot D^{(k+1)} E_{n-k+1}^T \begin{pmatrix} c \\ \delta_k \end{pmatrix} E_{n-k} \begin{pmatrix} b \\ \delta_{n-k-1} \end{pmatrix} \cdots E_2 \begin{pmatrix} b \\ \delta_1 \end{pmatrix} E_{n-k+2}^T \begin{pmatrix} c \\ \delta_{k-1} \end{pmatrix} \cdots E_n^T \begin{pmatrix} c \\ \delta_1 \end{pmatrix}. \quad (37)$$

Finally, reordering the elementary bidiagonal matrices of (37) we deduce (34) for $k+1$. Therefore, (34) holds for $k=2, \dots, n$, and, taking $k=n$ in (34), we deduce that

$$A^{-1} = E_n \begin{pmatrix} b \\ \delta_1 \end{pmatrix} \cdots E_2 \begin{pmatrix} b \\ \delta_{n-1} \end{pmatrix} D^{(n)} E_2^T \begin{pmatrix} c \\ \delta_{n-1} \end{pmatrix} \cdots E_n^T \begin{pmatrix} c \\ \delta_1 \end{pmatrix}$$

with $D^{(n)} = \text{diag}(\delta_n^{-1}, \dots, \delta_1^{-1})$, which is precisely $\mathcal{BD}(A^{-1})$. ■

6. Numerical experiments

In [15,23], assuming that the parameterization $\mathcal{BD}(A)$ of an square TP matrix A is known with HRA, Plamen Koev presented algorithms to solve some algebraic problems for A to HRA. Let us focus on the computation of the eigenvalues and the singular values. Koev implemented these algorithms in order to be used with Matlab and Octave in the software library *TNTool* available in [24]. The corresponding functions are `TNEigenValues` and `TNSingularValues`, respectively. The functions require as input argument the data determining the bidiagonal decomposition (6) of A , $\mathcal{BD}(A)$ given by (10), to HRA.

Let

$$A = T_n(a, -b, -c), \quad a, b, c > 0,$$

be a tridiagonal Toeplitz matrix satisfying $a^2 \geq 4bc \cos^2(\frac{\pi}{n+1})$. Let us denote by J_n the $n \times n$ matrix $\text{diag}(1, -1, \dots, (-1)^{n-1})$. Then, by Proposition 3.1, the matrix $J_n A J_n = |A|$ is TP. In addition, taking into account that $J_n^{-1} = J_n$, the matrix A is similar to the TP matrix $|A| = J_n A J_n$. Thus, A and $|A|$ have the same eigenvalues and, since J_n is unitary, also the same singular values. In Algorithm 1, the pseudocode for the computation of $\mathcal{BD}(A)$ to HRA can be seen. Taking into account that $\mathcal{BD}(|A|) = |\mathcal{BD}(A)|$, the eigenvalues and singular values of A can be computed to HRA by using Koev's algorithms and Algorithm 1 if $a^2 - 4bc$ is known to HRA.

In order to illustrate the accuracy of `TNEigenValues` and `TNSingularValues` with Algorithm 1, the sequence of matrices $A_5, A_{10}, \dots, A_{100}$, given by $A_n = T_n(4, -1/4, -15)$, has been considered. First, we have computed the eigenvalues and the singular values of these matrices with Mathematica using a precision of 100 digits. We have also computed approximations to the eigenvalues of those matrices in Matlab with `eig` and also with `TNEigenValues` using the absolute value of the bidiagonal decomposition provided by Algorithm 1. Then we have computed the relative errors of the approximations obtained considering the eigenvalues obtained with Mathematica as exact computations.

In Figure 1 (a), we can see the relative error for the minimal eigenvalue of each matrix $A_5, A_{10}, \dots, A_{100}$ for both `eig` and `TNEigenValues`.

We have also computed approximations to the singular values of the matrices A_5, \dots, A_{100} in Matlab with `svd` and also with `TNSingularValues` using the absolute value of the bidiagonal decomposition provided by Algorithm 1. Then we have computed the relative errors of the approximations obtained considering the singular values obtained with Mathematica as exact computations. In Figure 1 (b), we can see the relative

Algorithm 1 Computation of the bidiagonal decomposition of A to HRA.

Require: $n, a > 0, b, c < 0$ such that $m = a^2 - 4bc > 0$, and m known to HRA

Ensure: The $n \times n$ $\mathcal{BD}(A)$ of $A = T_n(a, -b, -c)$ to HRA

$$\frac{s_-}{s_+} = \frac{4bc}{a^2 + m + 2a\sqrt{m}}$$

$$s_+ = \frac{a + \sqrt{m}}{2\sqrt{bc}}$$

$$\text{num} = 1 + \frac{s_-}{s_+}$$

$$\text{den} = 1$$

for $i = 0 : n$ **do**

$$\delta_i = \sqrt{bc} s_+ \frac{\text{num}}{\text{den}}$$

$$\text{den} = \text{num}$$

$$\text{num} = \text{num} \frac{s_-}{s_+} + 1$$

end for

$$(\mathcal{BD}(A))_{ij} = 0 \text{ for } 1 \leq i, j \leq n$$

$$(\mathcal{BD}(A))_{11} = \delta_1$$

$$(\mathcal{BD}(A))_{12} = -\frac{c}{\delta_1}$$

for $i=2:n-1$ **do**

$$(\mathcal{BD}(A))_{i,i-1} = -\frac{b}{\delta_{i-1}}$$

$$(\mathcal{BD}(A))_{ii} = \delta_i$$

$$(\mathcal{BD}(A))_{i+1,i} = -\frac{c}{\delta_i}$$

end for

$$(\mathcal{BD}(A))_{n,n-1} = -\frac{b}{\delta_{n-1}}$$

$$(\mathcal{BD}(A))_{nn} = \delta_n$$

error for the minimal singular value of each matrix $A_5, A_{10}, \dots, A_{100}$ for both `svd` and `TNSingularValues`.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Gobierno de Aragón [E41-17R] and Ministerio de Ciencia, Innovación y Universidades [PGC2018-096321-B-I00].

ORCID

Héctor Orera  <http://orcid.org/0000-0002-4794-5875>

References

- [1] Demmel J, Dumitriu I, Holtz O, et al. Accurate and efficient expression evaluation and linear algebra. *Acta Numer.* 2008;17:87–145.
- [2] Elouafi M. Explicit inversion of band Toeplitz matrices by discrete Fourier transform. *Linear Multilinear Algebra.* 2018;66:1767–1782.
- [3] Luati A, Proietti T. On the spectral properties of matrices associated with trend filters. *Econ Theory.* 2010;26:1247–1261.

- [4] Noschese S, Pasquini L, Reichel L. Tridiagonal Toeplitz matrices: properties and novel applications. *Numer Linear Algebra Appl.* **2013**;20:302–326.
- [5] Noschese S, Reichel L. Eigenvector sensitivity under general and structured perturbations of tridiagonal Toeplitz-type matrices. *Numer. Linear Algebra Appl.* **2019**;26:e2232.
- [6] Reichel L, Ye Q. Simple square smoothing regularization operators. *Electron Trans Numer Anal.* **2009**;33:63–83.
- [7] Smith GD. Numerical solution of partial differential equations: finite difference methods. 3rd ed. New York (NY): Oxford University Press; **1985**.
- [8] Došlić T, Martinjak I, Škrekovski R. Total positivity of Toeplitz matrices of recursive hypersequences. *Ars Math Contemp.* **2019**;17:126–139.
- [9] Jia J. A breakdown-free algorithm for computing the determinants of periodic tridiagonal matrices. *Numer Algorithms.* **2020**;83:149–163.
- [10] Ando T. Totally positive matrices. *Linear Algebra Appl.* **1987**;90:165–219.
- [11] Gasca M, Micchelli CA. Total positivity and its applications: mathematics and its applications. Dordrecht (Amsterdam): Kluwer Academic Publishers Group; **1996**.
- [12] Pinkus A. Totally positive matrices. Cambridge (UK): Cambridge University Press; **2010**.
- [13] Berman A, Plemmons RJ. Nonnegative matrices in the mathematical sciences. Philadelphia (PA): Society for Industrial and Applied Mathematics (SIAM); **1994**. (Classics in Applied Mathematics; vol. 9).
- [14] Demmel J, Koev P. The accurate and efficient solution of a totally positive generalized Vandermonde linear system. *SIAM J Matrix Anal Appl.* **2005**;27:142–152.
- [15] Koev P. Accurate computations with totally nonnegative matrices. *SIAM J Matrix Anal Appl.* **2007**;29:731–751.
- [16] Delgado J, Orera H, Peña JM. Accurate computations with Laguerre matrices. *Numer Linear Algebra Appl.* **2019**;26:e2217.
- [17] Delgado J, Orera H, Peña JM. Accurate algorithms for Bessel matrices. *J Sci Comput.* **2019**;80:1264–1278.
- [18] Delgado J, Peña JM. Accurate computations with collocation matrices of q-Bernstein polynomials. *SIAM J Matrix Anal Appl.* **2015**;36:880–893.
- [19] Delgado J, Peña JM. Accurate computations with Lupaş matrices. *Appl Math Comput.* **2017**;303:171–177.
- [20] Marco A, Martínez J-J. A total positivity property of the Marchenko-Pastur law. *Electron J Linear Algebra.* **2015**;30:106–117.
- [21] Marco A, Martínez J-J. Bidiagonal decomposition of rectangular totally positive Said-Ball-Vandermonde matrices: error analysis, perturbation theory and applications. *Linear Algebra Appl.* **2016**;495:90–107.
- [22] Marco A, Martínez J-J, Viaña R. Accurate bidiagonal decomposition of totally positive h-Bernstein-Vandermonde matrices and applications. *Linear Algebra Appl.* **2019**;579:320–335.
- [23] Koev P. Accurate eigenvalues and SVDs of totally nonnegative matrices. *SIAM J Matrix Anal Appl.* **2005**;27:1–23.
- [24] Koev P. [cited 2020 Jan 16]. Available From: <http://www.math.sjsu.edu/koev/software/TNTool.html>
- [25] Horn RA, Johnson CR. Topics in matrix analysis. Cambridge (UK): Cambridge University Press; **1991**.
- [26] Barreras A, Peña JM. On tridiagonal sign regular matrices and generalizations. In: Casas F, Martínez V, editors. *Advances in differential equations and applications*. Cham (Switzerland): Springer International Publishing Switzerland; **2014**. p. 239–248.
- [27] Alonso P, Cortina R, Ranilla J, et al. An efficient and scalable block parallel algorithm of Neville elimination as a tool for the CMB maps problem. *J Math Chem.* **2012**;50:345–358.
- [28] Gasca M, Peña JM. Total positivity and Neville elimination. *Linear Algebra Appl.* **1992**;165:25–44.
- [29] Gasca M, Peña JM. On factorizations of totally positive matrices. In: Gasca M, Micchelli CA, editors. *Total positivity and its applications*. Dordrecht (Amsterdam): Kluwer Academic Publishers Group; **1996**. p. 109–130.

- [30] Barreras A, Peña JM. Accurate computations of matrices with bidiagonal decomposition using methods for totally positive matrices. *Numer Linear Algebra Appl.* [2013](#);20:413–424.
- [31] Katkova O, Vishnyakova A. On sufficient conditions for the total positivity and for the multiple positivity of matrices. *Linear Algebra Appl.* [2006](#);416:1083–1097.
- [32] Kavčić A, Moura J. Matrices with banded inverses: inversion algorithms and factorization of Gauss–Markov processes. *IEEE Trans Inf Theory.* [2000](#);46:1495–1509.
- [33] Peña JM. *M*-matrices whose inverses are totally positive. *Linear Algebra Appl.* [1995](#);221:189–193.
- [34] da Fonseca CM, Petronilho J. Explicit inverses of some tridiagonal matrices. *Linear Algebra Appl.* [2001](#);325:7–21.

Article 10

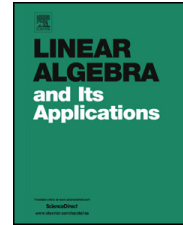
- [82] H. Orera and J. M. Peña. Accurate determinants of some classes of matrices. *Linear Algebra Appl.* 630 (2021), 1-14.



Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa



Accurate determinants of some classes of matrices [☆]



H. Orera ^{*}, J.M. Peña

Departamento de Matemática Aplicada/IUMA, Universidad de Zaragoza, Spain

ARTICLE INFO

Article history:

Received 8 February 2021
Accepted 22 July 2021
Available online 30 July 2021
Submitted by Y. Nakatsukasa

MSC:

15A15
15B35
15B48
65F40

Keywords:

B-matrix
Nekrasov matrix
B-Nekrasov matrix
Determinant
High relative accuracy

ABSTRACT

We present parametrizations of *B*-matrices, Nekrasov matrices and *B*-Nekrasov matrices that allow us to compute their determinants with high relative accuracy. Numerical examples confirm the accuracy of the method.

© 2021 Elsevier Inc. All rights reserved.

1. Introduction

The accurate computation with structured classes of matrices is receiving increasing attention in the recent years. For this purpose, a first step consists of a parametrization adapted to the structure of the considered matrices. It is known that an algorithm can be

[☆] This work was partially supported through the Spanish research grant PGC2018-096321-B-I00 (MCIU/AEI) and by Gobierno de Aragón (E41_20R).

^{*} Corresponding author.

E-mail addresses: hectororera@unizar.es (H. Orera), jmpena@unizar.es (J.M. Peña).

performed with high relative accuracy (HRA) if it does not include subtractions (except of the initial data), that is, if it only includes products, divisions, sums of numbers of the same sign, and subtractions of the initial data (cf. [6]). So, in particular, a subtraction-free algorithm can be carried out with HRA. Performing an algorithm with HRA is a very desirable goal because it implies that the relative errors of the computations are of the order of the machine precision even for ill-conditioned matrices. Up to now, only for a few classes of matrices HRA algorithms for their algebraic computations have been found. This paper contributes to this field providing adequate parametrizations and the corresponding HRA algorithms to compute with HRA the determinant of matrices belonging to several classes of matrices.

One of the classes of matrices considered in this paper is formed by Nekrasov matrices (see Section 3), which generalizes the class of strictly diagonally dominant matrices. On recent applications of Nekrasov matrices, see [8,15,17,22]. A second class of matrices considered is formed by B -matrices (see Section 2). In contrast to Nekrasov matrices, B -matrices can be very far from diagonal dominance (see the example given by (15)). However, they can also be applied to the localization of eigenvalues (see [18]). Finally, it is also considered the class of B -Nekrasov matrices (see Section 4), which contains B -matrices and Nekrasov Z -matrices with positive diagonal entries. Applications of B -Nekrasov matrices to linear complementarity problems can be seen in [3,4,9–11,13].

Let us now recall some related classes of matrices. A matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ is *strictly diagonally dominant* (*diagonally dominant*, respectively) if $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$ ($|a_{ii}| \geq \sum_{j \neq i} |a_{ij}|$, respectively) for all $i = 1, \dots, n$. A real matrix A is a Z -matrix if all its off-diagonal entries are nonpositive. A Z -matrix A is a nonsingular M -matrix if its inverse is nonnegative. Given a complex matrix $A = (a_{ij})_{1 \leq i, j \leq n}$, its *comparison matrix* $\mathcal{M}(A) = (\tilde{a}_{ij})_{1 \leq i, j \leq n}$ satisfies that $\tilde{a}_{ii} := |a_{ii}|$ and $\tilde{a}_{ij} := -|a_{ij}|$ for all $j \neq i$ and $i, j = 1, \dots, n$. Finally, we say that a complex matrix is an H -matrix if its comparison matrix is a nonsingular M -matrix. A more general definition of H -matrix can be found in [2].

The paper is organized as follows. In Section 2, we provide a method of $\mathcal{O}(n^3)$ elementary operations to compute with HRA the determinant of an $n \times n$ B -matrix from an adequate parametrization of the B -matrix. Section 3 uses the parametrization given in [16] to obtain a subtraction-free (and so with HRA) method of $\mathcal{O}(n^3)$ elementary operations to compute the determinant of an $n \times n$ Nekrasov Z -matrix with positive diagonal entries. Let us recall that a method to construct the inverse of such a matrix was presented in [16]. Section 4 provides a method of $\mathcal{O}(n^3)$ elementary operations to compute with HRA the determinant of an $n \times n$ B -Nekrasov matrix. Section 5 includes algorithms used in our methods and presents numerical examples that illustrate their great accuracy. Finally, Section 6 summarizes the main conclusions of the paper and comments some related problems and their difficulties.

2. Accurate determinants of B -matrices

Let us start by recalling the definition of a B -matrix [18].

Definition 2.1. A square real matrix $A := (a_{ij})_{1 \leq i, j \leq n}$ with positive row sums is a B -matrix if all its off-diagonal elements are bounded above by the corresponding row means, i.e., for all $i = 1, \dots, n$,

$$\sum_{j=1}^n a_{ij} > 0, \quad \frac{1}{n} \left(\sum_{k=1}^n a_{ik} \right) > a_{ij} \quad \forall j \neq i. \tag{1}$$

Let us now recall a useful decomposition of B -matrices. For this purpose, we first introduce the following notation. Given a real matrix $B = (b_{ij})_{1 \leq i, j \leq n}$, let us define for each $i = 1, \dots, n$, $r_i^+ := \max_{j \neq i} \{0, b_{ij}\}$. Then B can be decomposed in the form

$$B = B^+ + C, \tag{2}$$

$$B^+ = \begin{pmatrix} b_{11} - r_1^+ & \dots & b_{1n} - r_1^+ \\ \vdots & & \vdots \\ b_{n1} - r_n^+ & \dots & b_{nn} - r_n^+ \end{pmatrix}, \quad C = \begin{pmatrix} r_1^+ & \dots & r_1^+ \\ \vdots & & \vdots \\ r_n^+ & \dots & r_n^+ \end{pmatrix}. \tag{3}$$

Observe that, if B is a B -matrix, then B^+ is a strictly diagonally dominant Z -matrix (cf. Prop 2.3 of [18]). Therefore, for each $i = 1, \dots, n$,

$$d_{ii} = \sum_{j=1}^n (b_{ij} - r_i^+) > 0. \tag{4}$$

Let us recall that, given a diagonally dominant Z -matrix $A = (a_{ij})_{1 \leq i, j \leq n}$, the n^2 parameters given to assure that many algebraic computations can be performed with HRA are the off-diagonal entries of A and the n (nonnegative) row sums of A . These n^2 parameters were called the DD-parameters in [16]

$$\begin{cases} a_{ij}, & i \neq j, \\ s_i := \sum_{j=1}^n a_{ij}, & i = j. \end{cases} \tag{5}$$

The n^2 parameters of a B -matrix $B = (b_{ij})_{1 \leq i, j \leq n}$ that will be used to compute its determinant with HRA will be again its off-diagonal entries (as with diagonally dominant Z -matrices and with Nekrasov Z -matrices, see [16]) and the n positive parameters given by (4). We call these n^2 parameters of a B -matrix its B -parameters

$$\begin{cases} b_{ij}, & i \neq j, \\ d_{ii}, & i = j. \end{cases} \tag{6}$$

Theorem 2.2. *Let $B = (b_{ij})_{1 \leq i, j \leq n}$ be a B -matrix. Given its B -parameters (see (6)) we can compute $\det B$ with HRA.*

Proof. As commented earlier, the decomposition (2) of the matrix $B = B^+ + C$ satisfies that $A := B^+$ is a strictly diagonally dominant Z -matrix and that $C = re^T$, where $r := (r_1^+, \dots, r_n^+)^T$ and $e := (1, \dots, 1)^T$.

Let us first check that the DD -parameters of the diagonally dominant Z -matrix A can be computed with HRA. The n row sums of A coincide with the n B -parameters given by (4). Besides, the $n^2 - n$ remaining DD -parameters of A are $b_{ij} - r_i^+$ ($i \neq j$), which can be computed with HRA because they are subtractions of initial data, in fact of off-diagonal entries of B , which belong to the B -parameters given by (6). Since we can compute the DD -parameters of a diagonally dominant Z -matrix A with HRA and, by Proposition 2.4 of [16], we can use them to compute A^{-1} through a subtraction-free algorithm (and so with HRA), we conclude that we can compute A^{-1} with HRA. Since A is a strictly diagonally dominant Z -matrix with positive row sums, by the characterization of Theorem 2.3 of Chapter 6 of [1] it is a nonsingular M -matrix and so A^{-1} is a nonnegative matrix. Hence $A^{-1}r$ is a vector with all components nonnegative and so $e^T A^{-1}r$ is nonnegative. Therefore $1 + e^T A^{-1}r \neq 0$ and we can apply the matrix determinant lemma (see formula (13) of [20]) to $B = B^+ + C = A + re^T$ to derive

$$\det B = \det(A + re^T) = (1 + e^T A^{-1}r) \det A. \quad (7)$$

Since we can compute A^{-1} to HRA and e , A^{-1} and r are nonnegative, we can compute $1 + e^T A^{-1}r$ with HRA. Since we know the DD -parameters of A with HRA and $\det A$ can be computed from the DD -parameters of A with a subtraction-free algorithm, we conclude that we can compute $\det A$ with HRA. For instance, it can be obtained with the proof of Proposition 2.4 of [16] (alternative procedures using pivoting strategies can be found in [7] or in [19]). In fact, it is sufficient to compute the upper triangular matrix U in the proof of Proposition 2.4 of [16] and $\det A$ coincides with the product of the diagonal entries of U . In conclusion, we can compute $\det B$ by (7) with HRA. \square

Remark 2.3. *The computational cost of the algorithm suggested by the proof of Theorem 2.2 to compute $\det B$ with HRA for an $n \times n$ B -matrix B has $\mathcal{O}(n^3)$ elementary operations. In fact, calculating $A = B^+$ requires n^2 subtractions, the computation of $\det A$ and A^{-1} using the procedure of Proposition 2.4 of [16] requires $\mathcal{O}(n^3)$ elementary operations and, finally, the computations of (7) require $\mathcal{O}(n^2)$ elementary operations.*

3. Accurate determinants of Nekrasov Z -matrices with positive diagonal entries

Let $N := \{1, \dots, n\}$. Given a complex matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ with $a_{ii} \neq 0$ for all $i \in N$, let us define

$$h_1(A) := \sum_{j \neq 1} |a_{1j}|, \quad h_i(A) := \sum_{j=1}^{i-1} |a_{ij}| \frac{h_j(A)}{|a_{jj}|} + \sum_{j=i+1}^n |a_{ij}|, \quad i = 2, \dots, n. \quad (8)$$

The matrix A is called a *Nekrasov matrix* if $|a_{ii}| > h_i(A)$ for all $i \in N$ (see [21]). Nekrasov matrices are nonsingular H -matrices. In particular, a Nekrasov Z -matrix with positive diagonal entries is a nonsingular M -matrix.

The parametrization that we consider for an $n \times n$ Nekrasov Z -matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ with positive diagonal is given by the following n^2 parameters, which were called N -parameters in [16]:

$$\begin{cases} a_{ij}, & i \neq j, \\ \Delta_j(A) := a_{jj} - h_j(A), & j \in N. \end{cases} \quad (9)$$

Let us also recall that the diagonal matrix

$$S = \begin{pmatrix} \frac{h_1(A)}{a_{11}} & & & \\ & \frac{h_2(A)}{a_{22}} & & \\ & & \ddots & \\ & & & \frac{h_n(A)}{a_{nn}} \end{pmatrix} \quad (10)$$

holds that AS is diagonally dominant (see Lemma 2.2 of [16]).

Lemma 3.1 (Lemma 3.1 of [16]). *Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a Nekrasov matrix, and let $J = \{i_1, \dots, i_k\} \subseteq N$ ($i_1 < i_2 < \dots < i_k$) be the ordered set of indices such that $h_{i_j}(A) = 0$. Then at least $n - j$ off-diagonal entries of the row i_j are zero for all $j = 1, \dots, k$.*

In the following result we denote by $J := \{i_1, \dots, i_k\} \subseteq N$ ($i_1 < i_2 < \dots < i_k$) the ordered set of indices such that $h_{i_j}(A) = 0$. Given a set of indices $\alpha \subseteq N$, $A(\alpha) := A[\alpha^c]$, where α^c is the complement set of α .

Theorem 3.2. *Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a Nekrasov Z -matrix with positive diagonal entries. If we know its n^2 N -parameters (9), then we can compute its determinant to HRA using a subtraction-free algorithm of $\mathcal{O}(n^3)$ elementary operations.*

Proof. Let us start by computing $h_1(A), a_{11}, \dots, h_n(A), a_{nn}$ by (8) and (9) using the N -parameters of A . These computations require $\mathcal{O}(n^2)$ elementary operations, but they do not require any subtraction. Then we build the ordered set $I \subseteq N$ given by the increasing sequence of indices such that $h_i(A) \neq 0$. Let us first consider the case $I \neq N$. By Lemma 3.1 we know that the off-diagonal entries of the row $i_1 \notin I$ are zero. Using the cofactor expansion of the determinant we see that

$$\det A = a_{i_1 i_1} \det A(i_1). \quad (11)$$

If we apply again Lemma 3.1, we see that the row $i_2 \notin I$ has $n - 2$ zeros. In fact, those are the off-diagonal entries of the row $i_2 - 1$ of the matrix $A(i_1)$. Therefore,

$$\det A = a_{i_1 i_1} a_{i_2 i_2} \det A(i_1, i_2) \quad (12)$$

Following this argumentation with all the indices of $J = N \setminus I$ we deduce that

$$\det A = \det A[I] \prod_{k=1}^{n-|I|} a_{i_k i_k}. \quad (13)$$

Hence, computing $\det A[I]$ to HRA, gives $\det A$ also with HRA through (13).

Let S be the diagonal matrix given by (10) and let us define the submatrix $B := (AS)[I]$. Since B is a principal submatrix of AS it is a diagonally dominant matrix. Let us prove that we can compute its determinant with $\mathcal{O}(n^3)$ elementary operations without performing any subtraction. The first step consists on obtaining an adequate parametrization of B with a subtraction-free algorithm. The required parameters are its DD -parameters (5), i.e., its off-diagonal entries, $a_{ij} \frac{h_j(A)}{a_{jj}}$, and its row sums. By the choice of I , formulae (8), (9) and the sign pattern of a Z -matrix the row sums can be written as:

$$\begin{aligned} s_i &= \sum_{j \in I, j \neq i} a_{ij} \frac{h_j(A)}{a_{jj}} + h_i(A) = \sum_{j=1}^{i-1} a_{ij} \frac{h_j(A)}{a_{jj}} + h_i(A) + \sum_{j=i+1}^n a_{ij} \frac{h_j(A)}{a_{jj}} \\ &= \sum_{j=i+1}^n (-a_{ij}) \left(1 - \frac{h_j(A)}{a_{jj}}\right) = \sum_{j=i+1}^n |a_{ij}| \frac{a_{jj} - h_j(A)}{a_{jj}} = \sum_{j=i+1}^n |a_{ij}| \frac{\Delta_j(A)}{a_{jj}}. \end{aligned}$$

Therefore, we can obtain the DD -parametrization of B from (9) by a subtraction-free procedure that requires $\mathcal{O}(n^2)$ elementary operations. With these DD -parameters, we can obtain $\det B$ using a subtraction-free algorithm of $\mathcal{O}(|I|^3)$ elementary operations. Then it only remains to compute accurately $\det A[I] = \det B (\det S[I])^{-1}$ and use (13) to compute $\det A$ to HRA.

If $I = N$, the matrix $B = AS$ is a diagonally dominant M -matrix and we can compute its off-diagonal entries and its row sums with HRA by Theorem 2.3 of [16]. In fact, we can obtain its DD -parametrization (5) by a subtraction-free procedure and $\mathcal{O}(n^2)$ elementary operations. Analogously to the previous case, we can use these parameters to compute $\det B$ with a subtraction-free algorithm and $\mathcal{O}(n^3)$ elementary operations following the proof of Proposition 2.4 of [16]. Finally, we can use $\det B$ to obtain $\det A = \det B (\det S)^{-1}$. \square

4. Accurate determinants of B -Nekrasov matrices

Given a real matrix $B = (b_{ij})_{1 \leq i, j \leq n}$, let us consider the decomposition given by (2) and (3). Then we say that B is a B -Nekrasov matrix if the matrix B^+ given by (3) is a Nekrasov Z -matrix with positive diagonal entries.

Let us recall that the classes of matrices studied in the previous sections are also B -Nekrasov matrices. It is straightforward to see that Nekrasov Z -matrices admit the decomposition (2) with $C = 0$ and so Nekrasov Z -matrices with positive diagonal entries are B -Nekrasov matrices. On the other hand, given the decomposition (2) of a B -matrix, since B^+ is a strictly diagonally dominant Z -matrix with positive diagonal entries, it is also a Nekrasov Z -matrix with positive diagonal entries. Hence, a B -matrix is also a B -Nekrasov matrix.

We have seen that it is possible to compute the determinants of these two classes of matrices to HRA whenever the adequate parametrization is known with HRA. In this section, our aim is to extend these results to the wider class of B -Nekrasov matrices. Hence, we first introduce the following parametrization for this class. We call these n^2 parameters of a B -Nekrasov matrix its BN -parameters

$$\begin{cases} b_{ij}, & i \neq j, \\ \Delta_j(B^+), & j \in N. \end{cases} \quad (14)$$

Theorem 4.1. *Let $B = (b_{ij})_{1 \leq i, j \leq n}$ be a B -Nekrasov matrix. Given its BN -parameters (see (14)) we can compute $\det B$ with HRA.*

Proof. The decomposition (2) of the matrix $B = B^+ + C$ satisfies that $A := B^+$ is a Nekrasov Z -matrix with positive diagonal entries and that $C = re^T$, where $r := (r_1^+, \dots, r_n^+)^T$ and $e := (1, \dots, 1)^T$.

Let us first check that the N -parameters (9) of A can be computed with HRA. The N -parameters $\Delta_i(A)$ are also BN -parameters of B and the $n^2 - n$ remaining N -parameters are the off-diagonal entries of A , $b_{ij} - r_i^+$ ($i \neq j$). All the N -parameters of A are subtractions of the initial data provided by the BN -parameters (14) of B , and hence, they can be computed with HRA. Since we can compute the N -parameters of A with HRA, by Theorem 3.2 of [16] we can use these parameters to compute A^{-1} through a subtraction-free algorithm (and so with HRA). Moreover, since Nekrasov matrices are nonsingular H -matrices (see [21]) and A is a Nekrasov Z -matrix with positive diagonal entries, we deduce that A is a nonsingular M -matrix and that A^{-1} is a nonnegative matrix. Hence, $A^{-1}r$ is a vector with all components nonnegative and $e^T A^{-1}r$ is nonnegative. Therefore $1 + e^T A^{-1}r \neq 0$ and we can apply the matrix determinant lemma (see formula (13) of [20]) to $B = B^+ + C = A + re^T$ to derive (7). Since we can compute A^{-1} to HRA and e , A^{-1} and r are nonnegative, we can compute $1 + e^T A^{-1}r$ with HRA. By Theorem 3.2 we can also compute $\det A$ with a SF algorithm (and so, with HRA). In conclusion, we can compute $\det B$ by (7) with HRA. \square

In subsection 5.2 we introduce an implementation of the algorithm suggested by the proof of Theorem 4.1 to compute the determinant of a B -Nekrasov matrix to HRA.

Remark 4.2. *The computational cost to compute $\det B$ with HRA for an $n \times n$ B -Nekrasov matrix B following the proof of Theorem 4.1 is of $\mathcal{O}(n^3)$ elementary operations. In fact, calculating $A = B^+$ requires n^2 subtractions and the computation of $\det A$ and A^{-1} following the proofs of Theorem 3.2 and Proposition 2.4 of [16] requires $\mathcal{O}(n^3)$ elementary operations. Finally, the computation of (7) requires $\mathcal{O}(n^2)$ elementary operations.*

5. Algorithms and numerical experiments

In the previous sections we have seen that it is possible to compute the determinant of a B -matrix, a Nekrasov Z -matrix with positive diagonal entries or a B -Nekrasov matrix to HRA, whenever an adequate parametrization is known with HRA. In fact, we can build efficient algorithms to compute these determinants to HRA following the argumentation given by the proofs of Theorems 2.2, 3.2 and 4.1.

First, let us recall that, if we know the DD -parameters (5) of a diagonally dominant Z -matrix with positive diagonal entries with HRA, then we can compute its inverse also to HRA using a modified version of Gauss-Jordan elimination that is subtraction-free (see Proposition 2.4 of [16]). Moreover, with this method we can obtain the determinant of this matrix to HRA. In Algorithm 1, we introduce an implementation of Gauss-Jordan elimination without pivoting that computes the inverse and the determinant of a diagonally dominant Z -matrix to HRA from its DD -parameters.

Algorithm 1 Adapted G-J elimination.

Require: The DD -parameters: $A = (a_{ij})(i \neq j)$, $\mathbf{s} = (s_i)_{i=1}^n$

Ensure: $\det A$ and $A^{-1} = P$ to HRA

```

 $P = I_n$ 
for  $k = 1 : n - 1$  do
   $a_{kk} = s_k - \sum_{j=k+1}^n a_{kj}$ 
  for  $i = k + 1 : n$  do
     $p = a_{ik}/a_{kk}$ 
     $s_i = s_i - p * s_k$ 
     $a_{ik} = 0$ 
    for  $j = k + 1 : n$  do
      If  $i \neq j$  then  $a_{ij} = a_{ij} - p * a_{kj}$ 
    end for
     $P(i, :) = P(i, :) - p * P(k, :)$ 
  end for
end for
 $a_{nn} = s_n$ 
 $\det A = \prod_{i=1}^n a_{ii}$ 
for  $k = n : -1 : 2$  do
  for  $i = k - 1 : -1 : 1$  do
     $p = a_{ik}/a_{kk}$ 
     $P(i, :) = P(i, :) - p * P(k, :)$ 
  end for
end for
for  $i = 1 : n$  do
   $P(i, :) = P(i, :)/a_{ii}$ 
end for

```

Table 1
Relative error of $\det B_{30}$.

	$\varepsilon = 10^{-3}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-9}$
$\kappa_\infty(B_n)$	58001	$5.8 * 10^7$	$5.8 * 10^{10}$
HRA	$8.33401 * 10^{-16}$	$1.07732 * 10^{-15}$	$1.20346 * 10^{-15}$
MATLAB	$3.22764 * 10^{-12}$	$2.9085 * 10^{-9}$	$2.40044 * 10^{-6}$

5.1. B-matrices

In Algorithm 2 we have implemented a method to compute the determinant of a B -matrix to HRA following the argumentation given by the proof of Theorem 2.2. This algorithm takes the B -parameters (6) as input. It starts by computing the DD -parameters of the matrix B^+ given by the decomposition (2), then computes $(B^+)^{-1}$ and $\det B^+$ using the adapted version of Gauss-Jordan elimination given by Algorithm 1 and finally it calculates the determinant through formula (7).

Algorithm 2 Computation of the determinant of a B -matrix to HRA.

Require: The B -parameters (6): $B = (b_{ij})(i \neq j)$, $\mathbf{d} = (d_{ii})_{i=1}^n$

Ensure: $\det B$ to HRA

```

for  $i = 1 : n$  do
     $r_i = \max_{i \neq j} \{b_{ij}, 0\}$ 
    for  $j = 1 : i - 1$  do
         $a_{ij} = b_{ij} - r_i$ 
    end for
    for  $j = i + 1 : n$  do
         $a_{ij} = b_{ij} - r_i$ 
    end for
end for
Build the vector  $r = (r_i)_{1 \leq i \leq n}$ 
 $[\det A, A^{-1}] = \text{G-J}(A, \mathbf{d})$ 
 $\det B = (1 + e^T A^{-1} r) \det A$ 
    
```

▷ Algorithm 1

In order to test Algorithm 2 we have defined the family of B -matrices $B_n = (b_{ij})_{1 \leq i, j \leq n}$ with

$$b_{ij} = \begin{cases} 1, & i \neq j, \\ 1 + \varepsilon, & i = j. \end{cases} \tag{15}$$

Observe that the B -parameters (6) of B are provided by 1, corresponding to the off-diagonal entries of B_n , and ε .

We have implemented Algorithm 2 in Matlab to compute $\det B_{30}$ for different values of ε . We have also computed these determinants with the usual Matlab function `det` and in Mathematica using exact arithmetic. In Table 1, we compare the relative error of our approximations with those obtained using `det`.

5.2. B -Nekrasov matrices

In this subsection we introduce Algorithm 5 to compute the determinant of a B -Nekrasov matrix from its BN -parameters (14). Let us consider the decomposition of a B -Nekrasov matrix $B = B^+ + C$ given by (2). Algorithm 5 starts by computing the parametrization of B^+ using the function BNtoDD given by Algorithm 3. Then this function identifies the indices $i \in N$ such that $h_i(B) \neq 0$ and computes the parametrization of the diagonally dominant Z -matrix $B^+[I]S$ as well as the vector r (see the proofs of Theorems 3.2 and 4.1). Then Gauss-Jordan elimination is applied to this submatrix

Algorithm 3 BNtoDD.

Require: The BN -parameters (14): $B = (b_{ij})(i \neq j)$, $\Delta_j(B^+)$

Ensure: Parametrization of $A := B^+$, S , I , r

```

for  $i = 1 : n$  do
   $r_i = \max_{i \neq j} \{b_{ij}, 0\}$ 
  for  $j = 1 : i - 1$  do
     $a_{ij} = b_{ij} - r_i$ 
  end for
  for  $j = i + 1 : n$  do
     $a_{ij} = b_{ij} - r_i$ 
  end for
end for
Build the vector  $r = (r_i)_{1 \leq i \leq n}$ 
for  $i = 1 : n$  do
   $h_i = -\sum_{j=1}^{i-1} a_{ij}k_j - \sum_{j=i+1}^n a_{ij}$ 
   $a_{ii} = \Delta_i + h_i$ 
   $k_i = h_i/a_{ii}$ 
end for
Build  $I$ , the set of indices such that  $h_i(A) \neq 0$ .
if  $|I| > 1$  then
  for  $i \in I$  do
     $a_{ii} = -\sum_{j=i+1}^n a_{ij}\Delta_j/a_{jj}$ 
    for  $j \in I \setminus \{i\}$  do
       $a_{ij} = a_{ij}k_j$ 
    end for
  end for
  Build  $S$ , the  $|I| \times |I|$  diagonal matrix whose diagonal entries are  $k_j$ ,  $j \in I$ .
else if  $|I| = 1$  then
   $a_{II} = 1/a_{II}$ 
else
   $a_{nn} = 1/a_{nn}$ 
   $I = [n]$ 
end if

```

in order to compute its inverse and determinant. After multiplying this inverse by an appropriate scaling matrix, Algorithm 4 is used to compute the inverse and determinant of the whole matrix. If we consider only the case $I = N$, then Algorithm 4 would not be necessary. Finally, the determinant is computed through (7).

Remark 5.1. We have omitted an implementation specific for Z -Nekrasov matrices with positive diagonal entries because Algorithm 5 can be used to compute the determinant of this class of matrices whenever its N -parameters (9) are known to HRA. In that case, the decomposition $B = B^+ + C$ given by (2) would be trivial since $C = 0$ and $B = B^+$.

Algorithm 4 buildnekinv.

Require: A, I ▷ $A[I]$ contains $A[I]^{-1}$
Ensure: A^{-1}, t
 Build the set of ordered indices $J := I^c = \{j_1, \dots, j_k\}$ such that $j_1 > j_2 > \dots > j_k$.
 $t = 1$
for $i = J$ **do**
 $t = t * a_{ii}$
 $a_{ii} = 1/a_{ii}$
 $A[I|i] = -A[I](A[I|i]. * a_{ii})$ ▷ $*$ means component-wise multiplication
 $I = I \cup \{i\}$ (ordered)
end for

Algorithm 5 Computation of the determinant of a BN -matrix to HRA.

Require: The BN -parameters (14): $B = (b_{ij})(i \neq j)$, $\Delta_j(B^+)$
Ensure: $\det B$ to HRA
 $[A, S, I, r] = \text{BNtoDD}((b_{ij})(i \neq j), \Delta_j(B^+))$ ▷ Algorithm 3
if $|I| > 1$ **then**
 $[\det(A[I]), (A[I])^{-1}] = \text{G-J}(A[I])$ ▷ Algorithm 1
 $d = \det(A[I]) / (\det S)$
 $A[I] = S * (A[I])^{-1}$
else if $|I| = 1$ **then**
 $d = 1/a_{II}$
end if
 $[A^{-1}, t] = \text{buildnekinv}(A, I)$ ▷ Algorithm 4
 $\det A = d * t$
 $\det B = (1 + e^T A^{-1} r) \det A$

Moreover, the BN -parameters of a Z -Nekrasov matrix are its N -parameters. This fact implies that $r = 0$ and that the last line of Algorithm 5 would be redundant for these matrices.

The next example provides a family of B -Nekrasov matrices. Given a parameter $M > 5$, let us consider the matrices $C_n = (c_{ij})_{1 \leq i, j \leq n}$ with

$$c_{11} = 2M - 4, \quad c_{ii} = M + 2 + \frac{1}{M} \text{ for } 2 \leq i \leq n - 1, \quad c_{1j} = 4 \text{ for } 2 \leq j \leq n, \\ c_{nn} = M + 1 + \frac{1}{M}, \quad c_{i,i+1} = M - 1 \text{ for } 1 \leq i \leq n - 1, \quad c_{ij} = M \text{ elsewhere,}$$

$$C_n = \begin{pmatrix} 2M - 4 & M - 1 & M & M & \dots & M \\ 4 & M + 2 + \frac{1}{M} & M - 1 & M & \dots & M \\ 4 & M & M + 2 + \frac{1}{M} & M - 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & M \\ \vdots & \vdots & \ddots & M & M + 2 + \frac{1}{M} & M - 1 \\ 4 & M & \dots & M & M & M + 1 + \frac{1}{M} \end{pmatrix}.$$

The BN -parameters (14) of C_n are given by

$$\begin{cases} c_{ij}, & i \neq j, \\ M - 5, & i = j = 1, \\ \frac{1}{M}, & 2 \leq i \leq n. \end{cases} \tag{16}$$

Table 2Relative error of $\det C_{30}$.

	$M = 10^3$	$M = 10^6$	$M = 10^9$	$M = 10^{12}$
$\kappa_\infty(C_n)$	51522.9	$5.15776 * 10^7$	$5.15776 * 10^{10}$	$5.15776 * 10^{13}$
HRA	$1.01742 * 10^{-15}$	$3.63682 * 10^{-15}$	$2.96607 * 10^{-15}$	$3.70256 * 10^{-15}$
MATLAB	$2.77755 * 10^{-13}$	$3.21688 * 10^{-10}$	$1.69452 * 10^{-7}$	0.000261219

Table 3Relative error of $\det C_{200}$.

	$M = 10^3$	$M = 10^6$	$M = 10^9$	$M = 10^{12}$
$\kappa_\infty(C_n)$	397553	$3.97959 * 10^8$	$3.9796 * 10^{11}$	$3.9796 * 10^{14}$
HRA	$4.08576 * 10^{-15}$	$5.80975 * 10^{-14}$	$5.94435 * 10^{-14}$	$3.31511 * 10^{-14}$
MATLAB	$8.77075 * 10^{-13}$	$7.72472 * 10^{-10}$	$1.44196 * 10^{-6}$	0.00638139

We have computed $\det C_{30}$ from its BN -parameters using the HRA Algorithm 5 for different values of M . In Table 2 we compare our results with the determinants obtained using the Matlab function `det`. The relative error has been computed considering the determinant obtained with Mathematica using exact arithmetic.

In Table 3 we also show the results obtained computing $\det C_{200}$ from the BN -parameters with the HRA Algorithm 5 and we compare the results with the determinants computed with the Matlab function `det`.

6. Conclusions and some related open problems

HRA algorithms to compute the determinants of B -matrices, Nekrasov Z -matrices with positive diagonal entries and B -Nekrasov matrices, from adequate parametrizations, are provided. These algorithms have a computational cost of $\mathcal{O}(n^3)$ elementary operations for $n \times n$ matrices. In contrast to the high relative accuracy of our algorithms, our numerical experiments show that the usual Matlab function `det` can be very inaccurate for some ill-conditioned examples of these matrices.

We now comment some open problems related with the results and techniques included in this manuscript. First, we comment in the next two paragraphs two possible ways of extending the computations of determinants with HRA to other classes of matrices. Finally, we comment the problem of extending, in a natural way, our techniques for computing the determinants with HRA to the problem of computing inverses with HRA.

As it was pointed out in Remark 2.5 of [16], the version of Gauss-Jordan elimination given by Algorithm 1 can be adapted to produce the determinant and the inverse of a nonsingular M -matrix A with HRA whenever the following parameters are known to HRA: the $n^2 - n$ off-diagonal entries of A , the n entries of a vector $z > 0$ such that $s := Az > 0$ and the n entries of the vector s . For diagonally dominant M -matrices and Nekrasov Z -matrices with positive diagonal entries, it has been possible to find a representation of n^2 parameters that can be used to achieve HRA for these computations. The problem of finding suitable parametrizations for HRA of other subclasses of nonsingular

M -matrices remains open. For example, for the class of QN -matrices (quasi-Nekrasov matrices) introduced in [12], it is known a possible vector z such that $Az > 0$ under some additional hypothesis for the matrix entries, as it is shown in Theorem 2.2 of [5]. For a general QN -matrix, how to achieve the accurate computation of the determinant remains as an open problem.

Let us now consider the class of matrices such that the matrix B^+ from the decomposition $B = B^+ + C$ given by (2) and (3) is a nonsingular M -matrix. These matrices were introduced in [14] and they were called MB -matrices. Let us suppose that B^+ belongs to a subclass of nonsingular M -matrices with a known vector $z > 0$ such that $Az > 0$. If we know the $n^2 - n$ off-diagonal entries of B , the n entries of the vector $z > 0$ such that $s := B^+z > 0$ and the n entries of s , then we can compute the determinant of B to HRA. We can compute the off-diagonal entries of B^+ with subtractions of initial data. Using an adapted version of Gauss-Jordan elimination (as commented in the previous paragraph), we can compute both the inverse and the determinant of B^+ to HRA. Since B^+ is a nonsingular M -matrix, its inverse is nonnegative and therefore we can obtain $\det B$ with HRA using (7). Hence, it could be possible looking for parametrizations for other subclasses of MB -matrices to compute their determinants to HRA.

Finally, another open problem would be finding a method to compute the inverses of B -matrices and B -Nekrasov matrices to HRA. Based on the decomposition $B = B^+ + C$ given by (2) and (3), a natural choice would be using the Sherman-Morrison formula to compute the inverse,

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^tA^{-1}}{1 + v^tA^{-1}u},$$

instead of the matrix determinant formula (7) that we have used to compute the determinants with HRA. However, this method implies subtractions and we cannot assure the computation of the inverse to HRA. So it seems that new approaches are convenient for this problem.

Declaration of competing interest

The authors declare that there is no competing interest.

References

- [1] A. Berman, R.J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Classics in Applied Mathematics, vol. 9, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1994.
- [2] R. Bru, I. Giménez, A. Hadjidimos, Is $A \in \mathbf{C}^{n \times n}$ a general H -matrix?, *Linear Algebra Appl.* 436 (2012) 364–380.
- [3] P.F. Dai, J. Li, J. Bai, L. Dong, New error bounds for linear complementarity problems of S-Nekrasov matrices and B-S-Nekrasov matrices, *Comput. Appl. Math.* 38 (2) (2019), Paper No. 61, 14 pp.
- [4] P.F. Dai, J. Li, J. Bai, L. Dong, Notes on new error bounds for linear complementarity problems of Nekrasov matrices, B-Nekrasov matrices and QN -matrices, *Numer. Math., Theory Methods Appl.* 12 (2019) 1191–1212.

- [5] P.F. Dai, J.C. Li, Y.T. Li, C.Y. Zhang, Error bounds to linear complementarity problem of QN-matrices, *Calcolo* 53 (2016) 647–657.
- [6] J. Demmel, I. Dumitriu, O. Holtz, P. Koev, Accurate and efficient expression evaluation and linear algebra, *Acta Numer.* 17 (2008) 87–145.
- [7] J. Demmel, P. Koev, Accurate SVDs of weakly diagonally dominant m-matrices, *Numer. Math.* 98 (2004) 99–104.
- [8] L. Gao, Q. Liu, New upper bounds for the infinity norm of Nekrasov matrices, *J. Math. Inequal.* 14 (2020) 723–733.
- [9] L. Gao, Y. Wang, C. Li, Y. Li, Error bounds for linear complementarity problems of S-Nekrasov matrices and B-S-Nekrasov matrices, *J. Comput. Appl. Math.* 336 (2018) 147–159.
- [10] M. García-Esnaola, J.M. Peña, B-Nekrasov matrices and error bounds for linear complementarity problems, *Numer. Algorithms* 72 (2016) 435–445.
- [11] M. García-Esnaola, J.M. Peña, On the asymptotic optimality of error bounds for some linear complementarity problems, *Numer. Algorithms* 80 (2019) 521–532.
- [12] L.Y. Kolotilina, Bounds for the inverses of generalized Nekrasov matrices, *J. Math. Sci. (N.Y.)* 207 (2015) 786–794.
- [13] C. Li, P. Dai, Y. Li, New error bounds for linear complementarity problems of Nekrasov matrices and B-Nekrasov matrices, *Numer. Algorithms* 74 (2017) 997–1009.
- [14] H.B. Li, T.Z. Huang, H. Li, On some subclasses of P -matrices, *Numer. Linear Algebra Appl.* 14 (2007) 391–405.
- [15] C. Li, S. Yang, H. Huang, Y. Li, Y. Wei, Note on error bounds for linear complementarity problems of Nekrasov matrices, *Numer. Algorithms* 83 (2020) 355–372.
- [16] H. Orera, J.M. Peña, Accurate inverses of Nekrasov Z-matrices, *Linear Algebra Appl.* 574 (2019) 46–59.
- [17] H. Orera, J.M. Peña, Infinity norm bounds for the inverse of Nekrasov matrices using scaling matrices, *Appl. Math. Comput.* 358 (2019) 119–127.
- [18] J.M. Peña, A class of P -matrices with applications to the localization of the eigenvalues of a real matrix, *SIAM J. Matrix Anal. Appl.* 22 (2001) 1027–1037.
- [19] J.M. Peña, LDU decompositions with L and U well conditioned, *Electron. Trans. Numer. Anal.* 18 (2004) 198–208.
- [20] S.M. Rump, Ill-conditioned matrices are componentwise near to singularity, *SIAM Rev.* 41 (1999) 102–112.
- [21] T. Szulc, Some remarks on a theorem of Gudkov, *Linear Algebra Appl.* 225 (1995) 221–235.
- [22] J. Zhang, C. Bu, Nekrasov tensors and nonsingular H-tensors, *Comput. Appl. Math.* 37 (2018) 4917–4930.

Article 11

- [83] H. Orera and J. M. Peña. Error bounds for linear complementarity problems of B_{π}^R -matrices. *Comput. Appl. Math.* 40 (2021), Paper No. 94, 13 pp.



Error bounds for linear complementarity problems of B_{π}^R -matrices

Héctor Orera¹ · Juan Manuel Peña²

Received: 1 December 2020 / Revised: 23 February 2021 / Accepted: 15 March 2021
© SBMAC - Sociedade Brasileira de Matemática Aplicada e Computacional 2021

Abstract

It is proved that any B_{π}^R -matrix has positive determinant. For $\pi > 0$, norm bounds for the inverses of B_{π}^R -matrices and error bounds for linear complementarity problems associated with B_{π}^R -matrices are provided. In this last case, the bounds are simpler than previous bounds and also have the advantage that they can be used without previously knowing whether we have a B_{π}^R -matrix. Some numerical examples show that these new bounds can be considerably sharper than previous ones.

Keywords Error bounds · Linear complementarity problems · Norm bounds for the inverse · B_{π}^R -matrices

Mathematics Subject Classification 90C33 · 90C31 · 65G50 · 15A48

1 Introduction

This paper provides error bounds for linear complementarity problems (LCPs) associated with B_{π}^R -matrices as well as norms for the inverses of these matrices. The LCP (see Sect. 4) has many important applications, for instance, to problems in linear and quadratic programming, network equilibrium problems, or to the Nash equilibrium of a bimatrix game (see Berman and Plemmons 1994; Chen and Xiang 2006; Cottle et al. 1992; Schäffer 2004). A principal minor is the determinant of a submatrix involving the same rows and columns, and P -matrices

Communicated by Jinyun Yuan.

This work was partially supported through the Spanish research Grant PGC2018-096321-B-I00 (MCIU/AEI), by Gobierno de Aragón (E41-17R) and Feder 2014-2020 “Construyendo Europa desde Aragón”.

✉ Héctor Orera
hectororera@unizar.es

Juan Manuel Peña
jmpena@unizar.es

¹ Departamento de Matemática Aplicada, Universidad de Zaragoza, Zaragoza, Spain

² Departamento de Matemática Aplicada/IUMA, Universidad de Zaragoza, Zaragoza, Spain

are square matrices with all their principal minors positive. Let us recall a remarkable property of P -matrices: the solution to a LCP exists and is unique if and only if its associated matrix is a P -matrix (Cottle et al. 1992).

Error bounds for LCPs associated with several subclasses of P -matrices are presented in Chen and Xiang (2006), Chen et al. (2015), Dai et al. (2016, 2019a, 2019b), Gao and Li (2017), García-Esnaola and Peña (2009, 2012, 2014, 2019), Li et al. (2020) and Wang (2017). In particular, error bounds for LCPs associated with B_π^R -matrices with $\pi > 0$ were presented in Gao et al. (2019) and García-Esnaola and Peña (2017). The class of B_π^R -matrices was introduced by Neumann et al. (2013), generalizing the class of B -matrices (see Gao and Li 2017; García-Esnaola and Peña 2009; Mendes and Mendes-Gonçalves 2019; Peña 2001). If we do not know whether a given matrix is a B_π^R -matrix with a fixed $\pi > 0$, then we cannot apply the bounds of Gao et al. (2019) and García-Esnaola and Peña (2017). In this paper, we shall provide alternative bounds for any matrix with positive row sums that is a B_π^R -matrix with $\pi \geq 0$. Moreover, we shall characterize B_π^R -matrices with $\pi \geq 0$ and provide $\pi > 0$. In contrast to Gao et al. (2019) and García-Esnaola and Peña (2017), our new bound does not depend on an additional parameter ε , so that its application is simpler. In addition, we show in Sect. 4 with some test matrices used in Gao et al. (2019) and García-Esnaola and Peña (2017) that our new bound considerably improves those of Gao et al. (2019) and García-Esnaola and Peña (2017).

In Sect. 2, we first introduce B_π^R -matrices and clarify a result of Neumann et al. (2013), where it was claimed that any B_π^R -matrix is a P -matrix but the proof assumed that $\pi \geq 0$. We show in Example 1 that there exist B_π^R -matrices that are not P -matrices when π has a negative component. However, Theorem 2 proves that any B_π^R -matrix has positive determinant. We also present in Sect. 2 a characterization to determine whether a given matrix is a B_π^R -matrix with $\pi \geq 0$. This characterization also provides a positive vector π . Section 3 is devoted to bound the infinity norm of the inverse of B_π^R -matrices. Results of Sect. 3 are used in Sect. 4 to derive the new error bounds of LCPs associated with B_π^R -matrices with $\pi > 0$. Numerical examples are included at the end of Sect. 4.

Finally, let us recall some matrix definitions. We say that a matrix A is *nonnegative* (respectively, *positive*) if all its entries are nonnegative (respectively, positive) and we write $A \geq 0$ (respectively, $A > 0$). The same notation applies to vectors considering them as column matrices. A matrix $M = (m_{ij})_{1 \leq i, j \leq n}$ is a *strictly diagonally dominant matrix* if $|m_{ii}| > \sum_{j \neq i} |m_{ij}|$, for each $i = 1, \dots, n$. A *Z-matrix* is a square real matrix with nonpositive off-diagonal entries. A nonsingular *M-matrix* is a *Z-matrix* with nonnegative inverse. Nonsingular *M-matrices* form an important subclass of P -matrices and some fields where these matrices arise are dynamic systems, economics or the discretization of partial differential equations.

2 Some basic results on B_π^R -matrices

Let us start by recalling the definition of a B_π^R -matrix given in Neumann et al. (2013).

Definition 1 Let $\pi = (\pi_1, \dots, \pi_n)^T$ be a vector such that

$$0 < \sum_{j=1}^n \pi_j \leq 1. \quad (1)$$

Let $M = (m_{ij})_{1 \leq i, j \leq n}$ be a real matrix with positive row sums and let $R = (R_1, \dots, R_n)^T$ be the vector formed by the row sums of M . Then we say that M is a B_π^R -matrix if for all $i = 1, \dots, n$,

$$\pi_j R_i > m_{ij}, \quad \forall j \neq i. \quad (2)$$

When $\pi_j = 1/n$ for all j , the previous definition coincides with that of a B -matrix (see Peña 2001). The close relationship of P -matrices with the LCP was recalled in Introduction. In fact, in Theorem 3.4 of Neumann et al. (2013) it was proved that a B_π^R -matrix is also a P -matrix whenever the vector π is nonnegative. However, the condition on the sign of π is omitted as a hypothesis in the statement of that theorem. Precisely, as was commented in page 251 of Orera and Peña (2019), the nonnegativity of the vector π is sufficient to ensure that a B_π^R -matrix is also a P -matrix. So, we state the result that was proved in fact in Theorem 3.4 of Neumann et al. (2013).

Theorem 1 *If A is a B_π^R -matrix with $\pi \geq 0$, then A is a P -matrix.*

With the following example we show that the condition $\pi \geq 0$ can not be omitted to assure that a B_π^R -matrix is a P -matrix.

Example 1 Let us consider the vector $\pi = (1.1, -2.9, 2.1)^T$. Then the matrix

$$A := \begin{pmatrix} 2 & -3 & 2 \\ -1 & 1 & 1 \\ 0.1 & -1 & 1 \end{pmatrix}$$

is a B_π^R -matrix. However, A is not a P -matrix since the principal minor using the first and second rows and columns is -1 .

Let us also mention that, to derive bounds for LCPs associated with B_π^R -matrices, the condition $\pi > 0$ was used in Gao et al. (2019) and García-Esnaola and Peña (2017) as well as in the bounds that we shall present later. In contrast to the loss of the property of being a P -matrix seen in Example 1, we can see that $\det A > 0$ holds for any B_π^R -matrix A for any vector π .

Theorem 2 *Let $M = (m_{ij})_{1 \leq i, j \leq n}$ be a real matrix with positive row sums. If M is a B_π^R -matrix, then $\det M > 0$.*

Proof By (1) there exists $k \in \{1, \dots, n\}$ such that $\pi_k > 0$. Let us choose $\varepsilon > 0$ such that $\pi_k - \varepsilon > 0$ and $m_{ik} - (\pi_k - \varepsilon)R_i < 0$ for $i \neq k$. Then we can define a new parameter vector $\hat{\pi} = (\hat{\pi}_1, \dots, \hat{\pi}_n)^T$ with

$$\hat{\pi}_i = \begin{cases} \pi_i, & i \neq k, \\ \pi_k - \varepsilon, & i = k, \end{cases}$$

and use it to decompose M as

$$M = B^+ + C, \quad B^+ := (m_{ij} - \hat{\pi}_j R_i)_{1 \leq i, j \leq n}, \quad C := R\hat{\pi}^T. \quad (3)$$

Then B^+ is a Z -matrix with row sums $\bar{R} = (\bar{R}_1, \dots, \bar{R}_n)^T$. Observe that, by (1) and the definition of $\hat{\pi}$, $\sum_{j=1}^n \hat{\pi}_j < 1$. Hence, for $i = 1, \dots, n$, the row sum \bar{R}_i is given by

$$\bar{R}_i = \sum_{j=1}^n (m_{ij} - \hat{\pi}_j R_i) = R_i \left(1 - \sum_{j=1}^n \hat{\pi}_j \right) > 0. \quad (4)$$

Since B^+ is a Z -matrix with positive diagonal entries, the positivity of its row sums implies that it is also strictly diagonally dominant. Hence, B^+ is a nonsingular M -matrix and so $\det(B^+) > 0$. By the decomposition (3) and the relationship between R and \bar{R} given by (4), we have that

$$\begin{aligned} \det M &= \det(B^+ + C) = \det(B^+ + R\hat{\pi}^T) = \det(B^+)(1 + \hat{\pi}^T(B^+)^{-1}R) \\ &= \det(B^+) \left(1 + \hat{\pi}^T(B^+)^{-1} \left(1 - \sum_{j=1}^n \hat{\pi}_j \right)^{-1} \bar{R} \right). \end{aligned}$$

Given $e = (1, \dots, 1)^T$, observe that $B^+e = \bar{R}$, and so

$$\det M = \det(B^+) \left(1 + \hat{\pi}^T(B^+)^{-1} \left(1 - \sum_{j=1}^n \hat{\pi}_j \right)^{-1} B^+e \right).$$

Therefore, we deduce that

$$\begin{aligned} \det M &= \det(B^+) \left(1 + \hat{\pi}^T \left(1 - \sum_{j=1}^n \hat{\pi}_j \right)^{-1} e \right) = \det(B^+) \left(1 + \sum_{j=1}^n \hat{\pi}_j \left(1 - \sum_{j=1}^n \hat{\pi}_j \right)^{-1} \right) \\ &= \det(B^+) \left(1 - \sum_{j=1}^n \hat{\pi}_j \right)^{-1}, \end{aligned}$$

and, since $\det(B^+) > 0$, we conclude that $\det M > 0$. □

By Proposition 3.5 of Neumann et al. (2013), the class of matrices satisfying Definition 1 is closed under positive linear combinations. Then, by Theorem 2, it has positive determinant. Finally, Theorem 1 gives a sufficient condition to assure that the positive combination is a P -matrix. This information is gathered in the following corollary.

Corollary 1 *Let $A = (a_{ij})_{1 \leq i, j \leq n}$ and $B = (b_{ij})_{1 \leq i, j \leq n}$ be a B_π^R -matrix and a B_ψ^R -matrix, respectively. Let s and t be nonnegative numbers with $s + t > 0$. Then*

- (i) $\det(sA + tB) > 0$.
- (ii) *If $\pi, \psi > 0$, then $sA + tB$ is a P -matrix.*

We now present a characterization that allows us to determine whether a given matrix is a B_π^R -matrix with $\pi \geq 0$ and so, in the affirmative case, that it is in particular a P -matrix. Moreover, the characterization gives a suitable positive vector π satisfying (1) and so we can apply the bounds that will be presented later. This characterization will allow us to obtain bounds for B_π^R -matrices whenever the vector π is unknown. A characterization of a B_π^R -matrix for any π was obtained in Observation 3.2 of Neumann et al. (2013), but we are going to adapt it by imposing the additional condition $\pi \geq 0$.

Proposition 1 *Let A be a square matrix with positive row sums and let $R = (R_1, \dots, R_n)^T$ be the vector formed from the row sums of A . Then there exists a nonnegative vector π satisfying (1) such as A is a B_π^R -matrix if and only if*

$$\sum_{j=1}^n \max_{i \neq j} \left(\frac{a_{ij}}{R_i}, 0 \right) < 1. \tag{5}$$

Proof Let us first suppose that A is a B_{π}^R -matrix for a given nonnegative vector π satisfying (1). By (1) there exists $k \in \{1, \dots, n\}$ such that $\pi_k > 0$. Then we have that $\max_{i \neq k} \left(\frac{a_{ik}}{R_i}, 0\right) < \pi_k$ and, since $\max_{i \neq j} \left(\frac{a_{ij}}{R_i}, 0\right) \leq \pi_j$ for all $j \neq k$, we also have that

$$\sum_{j=1}^n \max_{i \neq j} \left(\frac{a_{ij}}{R_i}, 0\right) < \sum_{j=1}^n \pi_j \leq 1. \tag{6}$$

Conversely, let us now suppose that (5) holds. If we define

$$k := 1 - \sum_{j=1}^n \max_{i \neq j} \left(\frac{a_{ij}}{R_i}, 0\right), \tag{7}$$

then we have that the vector $\pi = (\pi_1, \dots, \pi_n)$ with

$$\pi_j := \max_{i \neq j} \left(\frac{a_{ij}}{R_i}, 0\right) + \frac{k}{n} \quad \text{for } j = 1, \dots, n \tag{8}$$

is positive and satisfies (1). Hence, A is a B_{π}^R -matrix. \square

Remark 1 Let us observe that the choice of π in (8) agrees with the natural parameter vector $\pi = (\frac{1}{n}, \dots, \frac{1}{n})^T$ of an $n \times n$ B -matrix in some extremal examples of B -matrices (see Neumann et al. 2013). A first example of these B -matrices is provided by any positive diagonal matrix. In this case, (7) gives $k = 1$ and so (8) gives $\pi_j = \frac{1}{n}$ for all $j = 1, \dots, n$. The other extremal example of a B -matrix is provided by a matrix of the form

$$A = \begin{pmatrix} 1 + \varepsilon & 1 & \dots & 1 \\ 1 & 1 + \varepsilon & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \dots & 1 & 1 + \varepsilon \end{pmatrix},$$

where $\varepsilon > 0$. In this case, $\frac{a_{ij}}{R_i} = \frac{1}{n+\varepsilon}$ for any $i \neq j$, and so (7) gives $k = 1 - \frac{n}{n+\varepsilon} = \frac{\varepsilon}{n+\varepsilon}$ and (8) gives $\pi_j = \frac{1}{n+\varepsilon} + \frac{\varepsilon}{n(n+\varepsilon)} = \frac{1}{n}$ for all $j = 1, \dots, n$.

Remark 2 Observe that in the proof of Proposition 1, we prove that, if the matrix A satisfies (5), then the vector π given by (8) is positive.

3 Norm bounds for the inverses of B_{π}^R -matrices

Given a B_{π}^R -matrix $M = (m_{ij})_{1 \leq i, j \leq n}$, in García-Esnaola and Peña (2017) a decomposition of M depending on a parameter ε was obtained and applied to derive error bounds of LCPs when the involved matrix is a B_{π}^R -matrix with $\pi_j > 0$ for all j . In the following result, we provide another decomposition of a B_{π}^R -matrix with $\pi_j > 0$ for all j , which will not depend on any parameter and which will be very useful in this paper.

Proposition 2 Let $M = (m_{ij})_{1 \leq i, j \leq n}$ be a B_{π}^R -matrix with $\pi_j > 0$ for all j and for each $i = 1, \dots, n$ let $\gamma_i := \max_{j \neq i} \{0, \frac{m_{ij}}{\pi_j}\}$. Then we can write $M = B^+ + C$, where $B^+ := (m_{ij} - \pi_j \gamma_i)_{1 \leq i, j \leq n}$ is a strictly diagonally dominant Z -matrix with positive diagonal entries and C is the rank one matrix given by $C := (\gamma_1, \dots, \gamma_n)^T (\pi_1, \dots, \pi_n)$.

Proof We only have to prove that the Z -matrix B^+ has positive row sums. As usual, let us denote by $R = (R_1, \dots, R_n)$ the vector of row sums of M , which are positive because M is a B_π^R -matrix. For each $i = 1, \dots, n$, from (1) we deduce that the sum of the i th row of B^+ is $R_i - \gamma_i (\sum_{j=1}^n \pi_j) \geq R_i - \gamma_i$. Then, by definition of γ_i , we conclude that it is bounded below by either R_i (and so, it is positive) or by $R_i - \frac{m_{ij}}{\pi_j}$ for some $j \in \{1, \dots, n\}$ (which is also positive by (2)). \square

The following result gives an upper bound for $\|M^{-1}\|_\infty$.

Theorem 3 Let $M = (m_{ij})_{1 \leq i, j \leq n}$ be a B_π^R -matrix with $\pi_j > 0$ for all j and let R_j, γ_j be given as in Definition 1 and Proposition 2, respectively. Then

$$\|M^{-1}\|_\infty \leq \frac{\max_{1 \leq i \leq n} \left\{ \frac{1}{\pi_i} - 1 \right\}}{\min_{1 \leq i \leq n} \left\{ R_i - \gamma_i \sum_{j=1}^n \pi_j \right\}}. \tag{9}$$

Proof By Proposition 2 and Theorem (2.3) of Chapter 6 of Berman and Plemmons (1994), B^+ is a nonsingular M -matrix. So, we can write $(B^+)^{-1} =: (\bar{b}_{ij})_{1 \leq i, j \leq n}$ with $\bar{b}_{ij} \geq 0$ for all i, j . Then we can express $M = B^+(I + (B^+)^{-1}C)$ and so

$$\|M^{-1}\|_\infty \leq \|(I + (B^+)^{-1}C)^{-1}\|_\infty \|(B^+)^{-1}\|_\infty. \tag{10}$$

Let us now provide an upper bound for $\|(B^+)^{-1}\|_\infty$. By Proposition 2, B^+ is a strictly diagonally dominant matrix with positive diagonal entries and so it has positive row sums:

$$R_i - \gamma_i \sum_{j=1}^n \pi_j > 0, \quad i = 1, \dots, n.$$

By Theorem 1 of Varah (1975), we deduce that

$$\|(B^+)^{-1}\|_\infty \leq \frac{1}{\min_{1 \leq i \leq n} \left\{ R_i - \gamma_i \sum_{j=1}^n \pi_j \right\}}. \tag{11}$$

Now we bound the other factor of (10). Observe that

$$I + (B^+)^{-1}C = \begin{pmatrix} 1 + a_1\pi_1 & a_1\pi_2 & \dots & a_1\pi_n \\ a_2\pi_1 & 1 + a_2\pi_2 & \dots & a_2\pi_n \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ a_n\pi_1 & a_n\pi_2 & \dots & 1 + a_n\pi_n \end{pmatrix}, \tag{12}$$

where $a_i := \sum_{j=1}^n \bar{b}_{ij}\gamma_j \geq 0$ for $i = 1, \dots, n$. Then (12) can be written as

$$I + (B^+)^{-1}C = I + AP, \tag{13}$$

where $A := (a_1, \dots, a_n)^T e^T (\geq 0)$, $P := \text{diag}(\pi_1, \pi_2, \dots, \pi_n)$ and $e := (1, \dots, 1)^T$. By our hypothesis on π , P is nonsingular and so $I + AP = P^{-1}(I + PA)P$. Denoting by $\bar{C} := PA$, we have

$$(I + AP)^{-1} = P^{-1}(I + \bar{C})^{-1}P. \tag{14}$$

Observe that $\bar{C} = \bar{a}e^T$, where $\bar{a}_i := \pi_i a_i \geq 0$, for each $i = 2, \dots, n$ and $\bar{a} := (\bar{a}_1, \dots, \bar{a}_n)^T$. So, since $e^T \bar{a} = \sum_{i=1}^n \bar{a}_i \geq 0$, we can derive from the Sherman–Morrison formula (see formula (2.1.5) of page 65 of Golub and Van Loan 2013)

$$(I + \bar{C})^{-1} = (I + \bar{a}e^T)^{-1} = I - \frac{\bar{a}e^T}{1 + e^T \bar{a}}. \tag{15}$$

Hence, by (14), we get that

$$(I + AP)^{-1} = \begin{pmatrix} 1 - \frac{\bar{a}_1}{1 + \sum_{i=1}^n \bar{a}_i} & \frac{\pi_2}{\pi_1} \left(\frac{-\bar{a}_1}{1 + \sum_{i=1}^n \bar{a}_i} \right) & \dots & \frac{\pi_n}{\pi_1} \left(\frac{-\bar{a}_1}{1 + \sum_{i=1}^n \bar{a}_i} \right) \\ \frac{\pi_1}{\pi_2} \left(\frac{-\bar{a}_2}{1 + \sum_{i=1}^n \bar{a}_i} \right) & 1 - \frac{\bar{a}_2}{1 + \sum_{i=1}^n \bar{a}_i} & \dots & \frac{\pi_n}{\pi_2} \left(\frac{-\bar{a}_2}{1 + \sum_{i=1}^n \bar{a}_i} \right) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\pi_1}{\pi_n} \left(\frac{-\bar{a}_n}{1 + \sum_{i=1}^n \bar{a}_i} \right) & \frac{\pi_2}{\pi_n} \left(\frac{-\bar{a}_n}{1 + \sum_{i=1}^n \bar{a}_i} \right) & \dots & 1 - \frac{\bar{a}_n}{1 + \sum_{i=1}^n \bar{a}_i} \end{pmatrix}. \tag{16}$$

Then since $\bar{a}_i \geq 0$ for all $i = 1, \dots, n$, we conclude that $\|(I + AP)^{-1}\|_{\infty}$ is given by

$$\|(I + AP)^{-1}\|_{\infty} = 1 - \frac{\bar{a}_i}{1 + \sum_{j=1}^n \bar{a}_j} + \sum_{j \neq i} \frac{\pi_j}{\pi_i} \frac{\bar{a}_i}{1 + \sum_{j=1}^n \bar{a}_j} \tag{17}$$

for some $i = 1, \dots, n$. Since $\sum_{j=1}^n \pi_j \leq 1$ and $\bar{a}_i \geq 0$ for all i , formula (17) can be bounded above by

$$\frac{\bar{a}_i}{1 + \sum_{j=1}^n \bar{a}_j} \left(\sum_{j \neq i} \frac{\pi_j}{\pi_i} - 1 \right) + 1 \leq \frac{1 - \pi_i}{\pi_i} - 1 + 1 = \frac{1 - \pi_i}{\pi_i}$$

and so,

$$\|(I + AP)^{-1}\|_{\infty} \leq \max_i \left\{ \frac{1}{\pi_i} - 1 \right\}. \tag{18}$$

Now the result follows from (10), (11), (13) and (18). □

Proposition 1, Remark 2 and Theorem 3 allow us to deduce the following corollary.

Corollary 2 *Let M be a square matrix with positive row sums $R = (R_1, \dots, R_n)^T$ satisfying (5), let $\pi = (\pi_1, \dots, \pi_n)$ be the positive vector given by (8) and let γ_j be given as in Proposition 2 for $j = 1, \dots, n$. Then M is a B_{π}^R -matrix and formula (9) holds.*

In the proof of Theorem 3, we have bounded the second factor of (10) using Varah’s bound for strictly diagonally dominant matrices of Theorem 1 of Varah (1975). If we use a sharper bound, then we obtain sharper bounds for the norm of the inverse of a B_{π}^R -matrix. To illustrate this fact, we are going to use the bound introduced in Kolotilina (2014) for Nekrasov matrices, which in particular improves Varah’s bound for SDD matrices (as proven in Theorem 2.4 of Kolotilina 2014):

$$\|A^{-1}\|_{\infty} \leq \max_{i \in N} \frac{z_i(A)}{|a_{ii}| - h_i(A)}, \tag{19}$$

where $z_i(A)$ and $h_i(A)$ are defined recursively for $i = 1, \dots, n$ by

$$z_1(A) := 1, \quad z_i(A) := \sum_{j=1}^{i-1} |a_{ij}| \frac{z_j(A)}{|a_{jj}|} + 1, \quad i = 2, \dots, n.$$

Table 1 Examples of B_π^R -matrices with their parameter vector π

Matrix	π	Source
$A_1(8)$	$(19/50, 19/50, 6/25)$	García-Esnaola and Peña (2017)
$M_1(21/25)$	$(7/24, 7/24, 1/4, 1/6)$	Gao et al. (2019)
$M_2(8/9)$	$(3/8, 3/8, 1/4)$	Gao et al. (2019)
$M_3(1/2)$	$(9/24, 7/24, 1/6, 1/6)$	Gao et al. (2019)

$$h_1(A) := \sum_{j \neq 1} |a_{1j}|, \quad h_i(A) := \sum_{j=1}^{i-1} |a_{ij}| \frac{h_j(A)}{|a_{jj}|} + \sum_{j=i+1}^n |a_{ij}|, \quad i = 2, \dots, n.$$

In particular, if we apply bound (19) to the second factor of (10) we deduce the following result:

Theorem 4 *Let $M = (m_{ij})_{1 \leq i, j \leq n}$ be a B_π^R -matrix with $\pi_j > 0$ for all j and let R_j, γ_j be given as in Definition 1 and Proposition 2, respectively. Then*

$$\|M^{-1}\|_\infty \leq \max_{1 \leq i \leq n} \left\{ \frac{1}{\pi_i} - 1 \right\} \max_{1 \leq i \leq n} \frac{z_i(B^+)}{m_{ii} - \gamma_i \pi_i - h_i(B^+)}, \tag{20}$$

where B^+ is given in Proposition 2, $h_i(B^+) = \sum_{j=1}^{i-1} \frac{\gamma_i \pi_j - m_{ij}}{m_{jj} - \gamma_j \pi_j} h_j(B^+) + \sum_{j=i+1}^n (\gamma_i \pi_j - m_{ij})$ and $z_i(B^+) = \sum_{j=1}^{i-1} \frac{\gamma_i \pi_j - m_{ij}}{m_{jj} - \gamma_j \pi_j} z_j(B^+) + 1$.

The next result follows from Proposition 1, Remark 2 and Theorem 4.

Corollary 3 *Let M be a square matrix with positive row sums $R = (R_1, \dots, R_n)^T$ satisfying (5), let $\pi = (\pi_1, \dots, \pi_n)$ be the positive vector given by (8) and let γ_j be given as in Proposition 2 for $j = 1, \dots, n$. Then M is a B_π^R -matrix and formula (20) holds.*

We now present some numerical examples to illustrate our new results. Our test matrices were introduced in previous articles that studied error bounds for LCPs of B_π^R -matrices. The matrix $A_1(m)$ corresponds to Example 1 from García-Esnaola and Peña (2017). $M_1(k)$, $M_2(h)$ and $M_3(m)$ are examples from Gao et al. (2019):

$$A_1(m) = \begin{pmatrix} 10m & -10m & 1 \\ -10m + 1 & 10m & 0 \\ 2 & 3 & 3 \end{pmatrix}, \quad M_1(k) = \begin{pmatrix} 4k & k & 0 & -k \\ k & 6k & 0 & 0 \\ 0 & k & 4k & -k \\ k & 0 & -k & 7k \end{pmatrix},$$

$$M_2(h) = \begin{pmatrix} 3h & h & -h \\ 3h & 10h & 3h \\ -h & h & 3h \end{pmatrix}, \quad M_3(m) = \begin{pmatrix} 3m & m & 0 & 0 \\ 0.5m & 4m & 0 & -0.5m \\ 0.5m & m & 3m & -0.5m \\ 0.5m & m & -0.5m & 3m \end{pmatrix}.$$

We have computed bounds for the infinity norm of the inverse using Theorems 3 and 4 and Corollaries 2 and 3. The previous theorems need a given vector π , so we are going to use the parameter vectors given in the original articles. In Table 1, we gather these parameters and we present our results in Table 2.

We can see that Theorem 4 only improves Theorem 3 for the matrix A_1 and that Corollary 2 (and Corollary 3) considerably improve Theorems 3 and 4.

Table 2 Bounds to $\|A^{-1}\|_\infty$

Matrix	$A_1(8)$	$M_1(21/25)$	$M_2(8/9)$	$M_3(1/2)$
$\ A^{-1}\ _\infty$	2.0000	0.40668	0.8839	1.0333
Theorem 3	30.083	10.4167	10.1250	17.500
Theorem 4	27.226	10.4167	10.1250	17.500
Corollary 2	7	4.4025	4.1720	7.0200
Corollary 3	7	4.4025	4.1720	7.0200

4 Error bounds for LCPs involving B_π^R -matrices

Let us recall that the linear complementarity problem (LCP) looks for a vector $x \in \mathbf{R}^n$ such that

$$x \geq 0, \quad Mx + q \geq 0, \quad x^T(Mx + q) = 0, \tag{21}$$

where M is the $n \times n$ associated real matrix and $q \in \mathbf{R}^n$. Some important applications of this problem have been mentioned in the Introduction.

By Theorem 2.3 of Chen and Xiang (2006), if M is a P -matrix, then the solution x^* of the LCP (21) satisfies

$$\|x - x^*\|_\infty \leq \max_{d \in [0,1]^n} \|M_D^{-1}\|_\infty \|r(x)\|_\infty, \tag{22}$$

where

$$M_D := I - D + DM, \tag{23}$$

I is the $n \times n$ identity matrix, D is the diagonal matrix $\text{diag}(d_i)$ with $0 \leq d_i \leq 1$, for all $i = 1, \dots, n$ and $r(x) := \min(x, Mx + q)$, where the min operator denotes the componentwise minimum of two vectors.

In García-Esnaola and Peña (2017), another decomposition of a B_π^R -matrix involving a parameter ε was obtained and applied to derive bounds for the error of the LCP when the associated matrix is a B_π^R -matrix with $\pi_j > 0$ for all j . It was also used in Gao et al. (2019). Let us now recall it to compare it with our new decomposition.

Given a B_π^R -matrix $M = (m_{ij})_{1 \leq i, j \leq n}$, by (1), there exists $j \in \{1, \dots, n\}$ such that $\pi_j > 0$. By (2), there exists an $\varepsilon > 0$ such that

$$\pi_j - \varepsilon > 0 \quad \text{and} \quad m_{ij} - (\pi_j - \varepsilon)R_i < 0, \quad \forall i \neq j. \tag{24}$$

Then we can write

$$M = B^+(\varepsilon) + C(\varepsilon), \tag{25}$$

where

$$B^+(\varepsilon) = \begin{pmatrix} m_{11} - \pi_1 R_1 & \dots & m_{1j} - (\pi_j - \varepsilon)R_1 & \dots & m_{1n} - \pi_n R_1 \\ \vdots & & \vdots & & \vdots \\ \vdots & & \vdots & & \vdots \\ m_{n1} - \pi_1 R_n & \dots & m_{nj} - (\pi_j - \varepsilon)R_n & \dots & m_{nn} - \pi_n R_n \end{pmatrix} \tag{26}$$

and

$$C(\varepsilon) = \begin{pmatrix} \pi_1 R_1 & \dots & \pi_{j-1} R_1 & (\pi_j - \varepsilon) R_1 & \pi_{j+1} R_1 & \dots & \pi_n R_1 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \pi_1 R_n & \dots & \pi_{j-1} R_n & (\pi_j - \varepsilon) R_n & \pi_{j+1} R_n & \dots & \pi_n R_n \end{pmatrix}. \tag{27}$$

To bound the error of the corresponding LCP, we have to provide an upper bound for $\|M_D^{-1}\|_\infty$, where M_D is given by (23) and M is a B_π^R -matrix for a vector $\pi = (\pi_1, \dots, \pi_n)$ with $\pi_i > 0$ for all $i = 1, \dots, n$. Let $B^+(\varepsilon)$ and $C(\varepsilon)$ be the matrices given by (26) and (27) and let

$$C_D := DC(\varepsilon), \quad B_D^+ := I - D + DB^+(\varepsilon). \tag{28}$$

By Proposition 2 of García-Esnaola and Peña (2017), $B_D^+(\varepsilon)$ is a strictly diagonally dominant Z -matrix with positive diagonal entries and so it has positive row sums. For each $i = 1, \dots, n$, let us denote by $\beta_i > 0$ the sum of the entries of the i th row of $B_D^+(\varepsilon)$ and let $\beta(\varepsilon) := \min_i \{\beta_i\}$. The following result shows the mentioned upper bound for $\max_{d \in [0,1]^n} \|M_D^{-1}\|_\infty$ given in Theorem 1 of García-Esnaola and Peña (2017). It uses the parameter ε .

Theorem 5 *Let M be a B_π^R -matrix for a vector $\pi = (\pi_1, \dots, \pi_n)$ with $\pi_i > 0$ for all $i = 1, \dots, n$ and let M_D, C_D , and B_D^+ be given by (23), (28) and $B^+(\varepsilon) =: (b_{ij})_{1 \leq i, j \leq n}$. Then*

$$\max_{d \in [0,1]^n} \|M_D^{-1}\|_\infty \leq \frac{\max_i \left\{ \frac{1}{\pi_i} - 1 \right\}}{\min\{\beta(\varepsilon), 1\}}, \tag{29}$$

where $\beta(\varepsilon) := \min_i \{\beta_i\}$ and $\beta_i := b_{ii} - \sum_{j \neq i} |b_{ij}|$, $i = 1, \dots, n$.

In this section, we present a new bound for the error of the LCP associated with a B_π^R -matrix for a vector $\pi = (\pi_1, \dots, \pi_n)$ with $\pi_i > 0$ for all $i = 1, \dots, n$ using the decomposition of Proposition 2. In contrast to the previous bound, it will not depend on a parameter.

Given M , a B_π^R -matrix for a vector $\pi = (\pi_1, \dots, \pi_n)$ with $\pi_i > 0$ for all $i = 1, \dots, n$, we can define again $M_D = (\bar{m}_{ij})_{1 \leq i, j \leq n}$ by (23) for any diagonal matrix $D = \text{diag}(d_i)$ with $0 \leq d_i \leq 1$ for all $i = 1, \dots, n$. If B^+ and C are the matrices given by the decomposition of M given in Proposition 2, then we can define the corresponding matrices B_D^+, C_D by

$$C_D := DC, \quad B_D^+ := I - D + DB^+, \quad B^+ = (b_{ij})_{1 \leq i, j \leq n}. \tag{30}$$

The following result gives an upper bound for $\|M_D^{-1}\|_\infty$.

Theorem 6 *Suppose that $M = (m_{ij})_{1 \leq i, j \leq n}$ is a B_π^R -matrix for a vector π with $\pi_i > 0$ for all $i = 1, \dots, n$ and let $M_D = (\bar{m}_{ij})_{1 \leq i, j \leq n}$, C_D and B_D^+ be the matrices given by (23) and (30). Then B_D^+ is a strictly diagonally dominant Z -matrix with positive diagonal entries and*

$$\max_{d \in [0,1]^n} \|M_D^{-1}\|_\infty \leq \frac{\max_{1 \leq i \leq n} \left\{ \frac{1}{\pi_i} - 1 \right\}}{\min_{1 \leq i \leq n} \left\{ 1, R_i - \gamma_i \sum_{j=1}^n \pi_j \right\}}, \tag{31}$$

where, for each $i = 1, \dots, n$, R_i and γ_i are given by Definition 1 and Proposition 2, respectively.

Proof It is easy to check that M_D is a $B_{\pi}^{\bar{R}}$ -matrix where $\bar{R} = (\bar{R}_1, \dots, \bar{R}_n)^T$ and that $\bar{R}_i = (1 - d_i) + d_i R_i$ for each $i = 1, \dots, n$. We can observe that the decomposition (10) of M_D is given by the matrices B_D^+ and C_D of (30). Then

$$\max_{d \in [0, 1]^n} \|M_D^{-1}\|_{\infty} \leq \max_{d \in [0, 1]^n} \|(I + (B_D^+)^{-1} C_D)^{-1}\|_{\infty} \max_{d \in [0, 1]^n} \|(B_D^+)^{-1}\|_{\infty}. \quad (32)$$

Following the argumentation given in the proof of Theorem 3, we can give the same bound for the first factor of (32). So, we have that

$$\max_{d \in [0, 1]^n} \|(I + (B_D^+)^{-1} C_D)^{-1}\|_{\infty} \leq \max_{1 \leq i \leq n} \left\{ \frac{1}{\pi_i} - 1 \right\}. \quad (33)$$

The matrix B_D^+ is a strictly diagonally dominant Z -matrix with positive diagonal entries, and so, taking into account (30), we can write

$$\alpha_i^D = (1 - d_i) + d_i b_{ii} - \sum_{j \neq i} d_i |b_{ij}| = 1 - d_i + d_i \sum_{j=1}^n (m_{ij} - \gamma_i \pi_j) > 0.$$

By Theorem 1 of Varah (1975), we deduce that

$$\|(B_D^+)^{-1}\|_{\infty} \leq \frac{1}{\min_{1 \leq i \leq n} \alpha_i^D} = \frac{1}{\min_{1 \leq i \leq n} \{1 - d_i + d_i \sum_{j=1}^n (m_{ij} - \gamma_i \pi_j)\}}. \quad (34)$$

Let us consider an index $k \in N$ such that $\alpha_k^D = \min_i \{\alpha_i^D\}$. Then

$$\alpha_k^D = 1 - d_k + d_k \sum_{j=1}^n (m_{kj} - \gamma_k \pi_j) = 1 - d_k + d_k \left(R_k - \sum_{j=1}^n \gamma_k \pi_j \right).$$

If $R_k - \sum_{j=1}^n \gamma_k \pi_j \geq 1$, then $\alpha_k^D \geq 1$ for any $d_k \in [0, 1]$, and so, $\|(B_D^+)^{-1}\|_{\infty} \leq 1$. Otherwise, we have that $\alpha_k^D \leq R_k - \sum_{j=1}^n \gamma_k \pi_j$ for any $d_k \in [0, 1]$. Taking into account these cases, we can bound (34) as follows:

$$\|(B_D^+)^{-1}\|_{\infty} \leq \frac{1}{\min_{1 \leq i \leq n} \left\{ 1, R_i - \gamma_i \sum_{j=1}^n \pi_j \right\}}. \quad (35)$$

So we conclude that (31) holds since it is the product of the bound (35) for $\|(B_D^+)^{-1}\|_{\infty}$ and the bound (33) for $\|(I + (B_D^+)^{-1} C_D)^{-1}\|_{\infty}$. \square

We can deduce the next result from Proposition 1, Remark 2 and Theorem 6.

Corollary 4 *Let M be a square matrix with positive row sums $R = (R_1, \dots, R_n)^T$ satisfying (5), let $\pi = (\pi_1, \dots, \pi_n)$ be the positive vector given by (8) and let γ_j be given as in Proposition 2 for $j = 1, \dots, n$. Then M is a B_{π}^R -matrix and formula (31) holds.*

Finally, we are going to present some numerical examples to compare our new results with previous ones. The test matrices are those used in the previous section. In this case, we have computed bounds for the error of the LCP using Theorem 6 (that used the given vector π in Table 1) and Corollary 4. We show the results obtained following this approach in the third and fourth rows of Table 3. We compare the results with those obtained using the bounds introduced in García-Esnaola and Peña (2017) and Gao et al. (2019), which are included in the first two rows of Table 3. We borrowed the data from the original articles

Table 3 Bounds for the LCP

Matrix	$A_1(8)$	$M_1(21/25)$	$M_2(8/9)$	$M_3(1/2)$
LCP García-Esnaola and Peña (2017)	26.389	10	6	10
LCP Gao et al. (2019)	20.192	9.9125	6.6667	9
Theorem 6	30.083	10.4167	10.1250	17.500
Corollary 4	7	5.5882	4.1720	7.0200

whenever possible, and we computed the corresponding bound when it was not available. These bounds also use the parameter vector π given by Table 1.

Table 3 shows that the bounds obtained with Theorem 6 using a given vector π are not necessarily sharper. However, we can see that the new bounds given by Corollary 4 are sharper in all cases. Moreover, another advantage of this approach is that it can be applied to any matrix with positive row sums to first identify if it is a B_π^R -matrix. If so, it computes a compatible vector π and then we can apply our new bounds without further modifications.

References

- Berman A, Plemmons RJ (1994) Nonnegative matrices in the mathematical sciences. Classics in applied mathematics. SIAM, Philadelphia
- Chen X, Xiang S (2006) Computation of error bounds for P-matrix linear complementarity problems. Math Program 106:513–525
- Chen T, Li W, Wu X, Vong S (2015) Error bounds for linear complementarity problems of MB -matrices. Numer Algorithms 70:341–356
- Cottle RW, Pang J-S, Stone RE (1992) The linear complementarity problems. Academic Press, Boston
- Dai P-F, Li J-C, Li Y-T, Zhang C-Y (2016) Error bounds to linear complementarity problem of QN-matrices. Calcolo 53:647–657
- Dai P-F, Li J, Bai J, Dong L (2019a) Notes on new error bounds for linear complementarity problems of Nekrasov matrices, B -Nekrasov matrices and QN -matrices. Numer Math Theory Methods Appl 12:1191–1212
- Dai P-F, Li J, Bai J, Dong L (2019b) New error bounds for linear complementarity problems of S -Nekrasov matrices and B - S -Nekrasov matrices. Comput Appl Math 38:61. <https://doi.org/10.1007/s40314-019-0818-4>
- Gao L, Li C (2017) An improved error bound for linear complementarity problems for B -matrices. J Inequal Appl 2017:144. <https://doi.org/10.1186/s13660-017-1414-z>
- Gao L, Li C, Li Y (2019) Parameterized error bounds for linear complementarity problems of B_π^R -matrices and their optimal values. Calcolo 56:31. <https://doi.org/10.1007/s10092-019-0328-1>
- García-Esnaola M, Peña JM (2009) Error bounds for linear complementarity problems for B -matrices. Appl Math Lett 22:1071–1075
- García-Esnaola M, Peña JM (2012) Error bounds for linear complementarity problems of B^S -matrices. Appl Math Lett 25:1379–1383
- García-Esnaola M, Peña JM (2014) Error bounds for linear complementarity problems of Nekrasov matrices. Numer Algorithms 67:655–667
- García-Esnaola M, Peña JM (2017) B_π^R -matrices and error bounds for linear complementarity problems. Calcolo 54:813–822
- García-Esnaola M, Peña JM (2019) On the asymptotic optimality of error bounds for some linear complementarity problems. Numer Algorithms 80:521–532
- Golub GH, Van Loan CF (2013) Matrix computations, 4th edn. The Johns Hopkins University Press, Baltimore
- Kolotilina LY (2014) On bounding inverses to Nekrasov matrices in the infinity norm. J Math Sci 199:432–437
- Li C, Yang S, Huang H, Li Y, Wei Y (2020) Note on error bounds for linear complementarity problems of Nekrasov matrices. Numer Algorithms 83:355–372
- Mendes C, Mendes-Gonçalves S (2019) On a class of nonsingular matrices containing B -matrices. Linear Algebra Appl 578:356–369

- Neumann M, Peña JM, Pryporova O (2013) Some classes of nonsingular matrices and applications. *Linear Algebra Appl* 438:1936–1945
- Orera H, Peña JM (2019) B_{π}^R -tensors. *Linear Algebra Appl* 581:247–259
- Peña JM (2001) A class of P -matrices with applications to the localization of the eigenvalues of a real matrix. *SIAM J Matrix Anal Appl* 22:1027–1037
- Schäffer U (2004) A linear complementarity problem with a P -matrix. *SIAM Rev* 46:189–201
- Varah JM (1975) A lower bound for the smallest singular value of a matrix. *Linear Algebra Appl* 11:3–5
- Wang F (2017) Error bounds for linear complementarity problems of weakly chained diagonally dominant B -matrices. *J Inequal Appl* 2017:33. <https://doi.org/10.1186/s13660-017-1303-5>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Part III

THEMATIC UNIT AND SUMMARY OF THE ARTICLES

Chapter 4

Totally positive matrices

This chapter shows the thematic unit of the articles on totally positive matrices [16–20] and the article about tridiagonal Toeplitz P -matrices [21]:

- [17] **Article 1:** J. Delgado, H. Orera and J. M. Peña. Accurate computations with Laguerre matrices. *Numer. Linear Algebra Appl.* 26 (2019), e2217, 10 pp.
- [16] **Article 2:** J. Delgado, H. Orera and J. M. Peña. Accurate algorithms for Bessel matrices. *J. Sci. Comput.* 80 (2019), 1264-1278.
- [18] **Article 6:** J. Delgado, H. Orera and J. M. Peña. Accurate bidiagonal decomposition and computations with generalized Pascal matrices. *J. Comput. Appl. Math.* 391 (2021), Paper No. 113443, 10 pp.
- [20] **Article 7 :** J. Delgado, H. Orera and J. M. Peña. High relative accuracy with matrices of q -integers. *Numer. Linear Algebra Appl.* 28 (2021), Paper No. e2383, 20 pp.
- [19] **Article 8:** J. Delgado, H. Orera and J. M. Peña. Optimal properties of tensor product of B-bases. *Appl. Math. Lett.* 121 (2021), Paper No. 107473, 5 pp.
- [21] **Article 9:** J. Delgado, H. Orera and J. M. Peña. Characterizations and accurate computations for tridiagonal Toeplitz matrices, *Linear and Multilinear Algebra* (2021), Published online, DOI: 10.1080/03081087.2021.1884180.

Totally positive (TP) matrices are matrices whose minors are all nonnegative. Even though this definition might sound too restrictive, they appear in plenty of applications and its strong structure translates into many useful properties [3, 36, 89]. For example, TP matrices appear in Approximation Theory, Combinatorics, Graph Theory and Computer Aided Geometric Design (CAGD) . One of the nice properties that justifies their role in CAGD and approximation theory is their shape preserving properties. In fact, TP matrices present variation diminishing properties, i.e., linear transformations given by TP matrices do not increase the number of sign changes of their input vector (see Chapter 3 of [89] or Chapter 4 of [36] for formal definitions on variation diminishing properties).

This chapter is organized as follows. Section 4.1 recalls the basic results and tools used to assure HRA with TP matrices. In particular, it presents the bidiagonal decomposition and the functions that can be used to achieve accurate computation if this representation is known accurately. Then it is divided in subsections that introduce our new results for different classes of matrices. First, we introduce collocation matrices of generalized Laguerre polynomials [17], of Bessel polynomials and of reverse Bessel polynomials [16]. We have characterized when these matrices are TP and shown that their bidiagonal decomposition can be computed to HRA, and, hence, that it is possible to compute their eigenvalues, singular values, inverses as well as the solution to some linear systems of equations with HRA. Then we have considered some extensions of the Pascal matrix appearing in Combinatorics [18]. We have deduced the bidiagonal decomposition of these classes of matrices and, depending on the sign of the multipliers of this representation, we have characterized whether they are TP or not. And, in the case that they are TP, we have also studied when the bidiagonal decomposition can be computed accurately and used as a parametrization to assure HRA. Our last examples of TP matrices come from q -calculus [54]. We included a section devoted to the introduction of some q -analogues such as the q -Pascal matrix, matrices of q -Stirling numbers and an extension of generalized Laguerre polynomials [20]. We have shown that these matrices are TP and how to compute their bidiagonal decomposition to HRA. In Section 4.2, we introduce optimal properties on the minimal singular value, eigenvalue and condition number of tensor products of B-bases [19] compared to the tensor product of other NTP bases of their generated space of functions (see Section 3.3). Finally, Section 4.3 presents our results about tridiagonal Toeplitz matrices. We have characterized the cases when these matrices are TP, M -matrices or general P -matrices. We also show how to perform accurate computations in some cases. In particular, that it is possible to compute the determinants and inverse accurately of skew-symmetric sign tridiagonal P -matrices as well as the bidiagonal decomposition of nonsingular tridiagonal Toeplitz TP matrices and nonsingular tridiagonal Toeplitz M -matrices. In this particular case, both cases are closely related. The bidiagonal decomposition allows us to achieve accurate computations with both classes of matrices. Moreover, the inverses of nonsingular tridiagonal Toeplitz M -matrices are TP matrices, and we have also obtained the bidiagonal decomposition of these TP inverses.

4.1 Accurate computations with TP matrices

In Section 3.3 we have introduced the class of TP matrices and we provided a short survey on some of the properties that are used to achieve accurate computations. The work on accurate computations presented in [16–18, 20] takes as a basis the representation of a nonsingular TP matrix given by its bidiagonal decomposition. Let us recall that any nonsingular TP matrix can be expressed in terms of the unique bidiagonal decomposition given by the following theorem.

Theorem 4.1. (cf. Theorem 4.2 of [47]). *Let A be a nonsingular $n \times n$ TP matrix. Then A admits a decomposition of the form*

$$A = F_{n-1} \cdots F_1 D G_1 \cdots G_{n-1}, \quad (4.1)$$

where F_i and G_i , $i \in \{1, \dots, n-1\}$, are the lower and upper triangular nonnegative bidiagonal matrices given by

$$F_i = \begin{pmatrix} 1 & & & & & \\ 0 & 1 & & & & \\ & \ddots & \ddots & & & \\ & & 0 & & & \\ & & & m_{i+1,1} & & \\ & & & & 1 & \\ & & & & & \ddots & \ddots \\ & & & & & & m_{n,n-i} & 1 \end{pmatrix}, \quad G_i^T = \begin{pmatrix} 1 & & & & & \\ 0 & 1 & & & & \\ & \ddots & \ddots & & & \\ & & 0 & & & \\ & & & \tilde{m}_{i+1,1} & & \\ & & & & 1 & \\ & & & & & \ddots & \ddots \\ & & & & & & \tilde{m}_{n,n-i} & 1 \end{pmatrix}, \quad (4.2)$$

and D a diagonal matrix $\text{diag}(p_{11}, \dots, p_{nn})$ with positive diagonal entries. If, in addition, the entries m_{ij} , \tilde{m}_{ij} satisfy

$$m_{ij} = 0 \Rightarrow m_{hj} = 0 \quad \forall h > i$$

and

$$\tilde{m}_{ij} = 0 \Rightarrow m_{ik} = 0 \quad \forall k > j,$$

then the decomposition (4.1) is unique.

The bidiagonal decomposition given by (4.1) and (4.2) is defined by n^2 parameters and can be represented by the following abbreviated notation introduced in [60]:

$$(\mathcal{BD}(A))_{ij} = \begin{cases} m_{ij}, & \text{if } i > j, \\ \tilde{m}_{ji}, & \text{if } i < j, \\ p_{ii}, & \text{if } i = j. \end{cases} \quad (4.3)$$

As we mentioned in Section 3.3, the parameters m_{ij} , \tilde{m}_{ji} and p_{ii} from the bidiagonal decomposition (4.3) are the multipliers and pivots associated to an elimination procedure called Neville elimination (3.11). For $1 \leq j < i \leq n$, m_{ij} and p_{ii} are the multipliers and the diagonal pivots when applying Neville elimination to A and \tilde{m}_{ij} are the multipliers when applying Neville elimination to A^T .

For nonsingular totally positive matrices, the bidiagonal decomposition can be used as a parametrization to achieve accurate computations. In [59, 60], Plamen Koev devised algorithms to solve many algebraic problems with nonsingular TP matrices to high relative accuracy using the bidiagonal decomposition as input. He implemented these algorithms and they are available in the library TNTool to be used in Matlab and Octave. The library can be downloaded from Koev's personal webpage [58], and it also includes subsequent contributions of more authors. Some of the functions from the library that have been key to achieving high relative accuracy in our work are the following:

- `TNEigenvalues`: Computes the eigenvalues of A to HRA from $\mathcal{BD}(A)$.
- `TNSingularValues` Computes the singular values of A to HRA from $\mathcal{BD}(A)$.
- `TNInverseExpand` Computes the explicit inverse A^{-1} to HRA from $\mathcal{BD}(A)$. This function was contributed by Ana Marco and José Javier Martínez [77].
- `TNSolve` Computes the solution to the linear system of equations $Ax = b$ and assures the HRA whenever b has an alternating sign pattern. It takes as input $\mathcal{BD}(A)$ and b .

- `TNProduct` Computes $\mathcal{BD}(AB)$, the bidiagonal decomposition of the product of two nonsingular TP matrices A and B , from $\mathcal{BD}(A)$ and $\mathcal{BD}(B)$ to HRA.
- `TNVandBD` Computes the bidiagonal decomposition of the Vandermonde matrix on the points $t_1 < \dots < t_n$ to HRA. It requires the nodes as input.

Thanks to these algorithms, we can achieve accurate computations with nonsingular TP matrices if we know their bidiagonal decomposition accurately. Therefore, apparently it seems that we have all the ingredients necessary to achieve HRA for a nonsingular TP matrix. Let us remember that obtaining the bidiagonal decomposition through NE implies subtractions, and hence, we cannot assure HRA for it. Moreover, many TP matrices are ill-conditioned in the traditional sense. Therefore, we need to find a different method to obtain the bidiagonal decomposition accurately in order to take advantage of the high relative accuracy of the functions provided by the library `TNTool`.

One method for deriving the bidiagonal decomposition accurately could be using (3.14) if we know the explicit expression of the minors of the nonsingular TP matrix with HRA. This idea has been sometimes used for obtaining the bidiagonal decomposition of TP matrices.

4.1.1 Accurate computations with Laguerre matrices

Our first new class of nonsingular TP matrices comes from collocation matrices (3.20) of generalized Laguerre polynomials. Let us recall that, for $\alpha > -1$, the *generalized Laguerre polynomials* are given by

$$L_n^{(\alpha)}(t) = \sum_{k=0}^n (-1)^k \binom{n+\alpha}{n-k} \frac{t^k}{k!}, \quad n = 0, 1, 2, \dots \quad (4.4)$$

They are orthogonal polynomials on $[0, \infty)$ with respect to the weight function $x^\alpha e^{-x}$. Let us observe that $\alpha = 0$ corresponds to the classical Laguerre polynomials. These polynomials appear in many applications, like in the use of Gaussian quadrature rules to numerically compute integrals. Moreover, these polynomials and their extension have important applications in Quantum Mechanics (see [62]).

Given a real number x and a positive integer k , let us denote the corresponding *falling factorial* by

$$x^{(k)} := x(x-1)(x-2) \cdots (x-k+1). \quad (4.5)$$

Let us also denote $x^{(0)} := 1$. Let $M := \left(L_{j-1}^{(\alpha)}(t_{i-1}) \right)_{1 \leq i, j \leq n+1}$ be the collocation matrix of the generalized Laguerre polynomials at $(0 >) t_0 > t_1 > \dots > t_n$, let P_U be the $(n+1) \times (n+1)$ upper triangular Pascal matrix with $\binom{j-1}{i-1}$ as its (i, j) -entry for $j \geq i$ and let S_α and J be the $(n+1) \times (n+1)$ diagonal matrices:

$$S_\alpha := \text{diag} \left((\alpha + i)^i \right)_{0 \leq i \leq n}, \quad J := \text{diag} \left((-1)^i \right)_{0 \leq i \leq n}. \quad (4.6)$$

The following result assures that, given the parameters $(0 >) t_0 > t_1 > \dots > t_n$, many algebraic computations with these collocation matrices M can be performed with HRA. It also shows that these matrices are STP and it gives a particular factorization for these matrices.

Theorem 4.2. (Theorem 2 of [17]) Let $M := \left(L_{j-1}^{(\alpha)}(t_{i-1}) \right)_{1 \leq i, j \leq n+1}$ for $(0 >) t_0 > t_1 > \dots > t_n$ with $\alpha > -1$, let P_U be the $(n+1) \times (n+1)$ upper triangular Pascal matrix, let S_α and J be the $(n+1) \times (n+1)$ diagonal matrices given by (4.6) and let $V := (t_{i-1}^{j-1})_{1 \leq i, j \leq n+1}$. Then $M = VJS_\alpha^{-1}P_US_0^{-1}S_\alpha$ is an STP matrix and, given the parametrization t_i ($0 \leq i \leq n$), the following computations can be performed with HRA: all the eigenvalues, all the singular values and the inverse of M , as well as the solution of the linear systems $Mx = b$, where $b = (b_0, \dots, b_n)^T$ has alternating signs.

In particular, Theorem 4.2 includes the collocation matrices of classical Laguerre polynomials when $\alpha = 0$. One of the main ideas used in the proof of the previous Theorem will be used in other of our results on accurate computations with collocation matrices. For a collocation matrix of the generalized Laguerre polynomials M , we have seen that $M = BC$, with $B = VJ$ being a Vandermonde matrix and $C = S_\alpha^{-1}P_US_0^{-1}S_\alpha$ (see proof of Theorem 2 in [17]). In fact, for computing the bidiagonal decomposition of M to HRA we first compute accurately both $\mathcal{BD}(B)$ and $\mathcal{BD}(C)$, and then we compute the bidiagonal decomposition of their product using the function `TNProduct` from the library `TNTool` (corresponding to Algorithm 5.1 in [60]). Since B is a Vandermonde matrix, $\mathcal{BD}(B)$ is known and it can be computed accurately (in fact, this can be done with the function `TNVandBD` available in `TNTool`). We can derive $\mathcal{BD}(C)$ using the fact that the bidiagonal decomposition of the upper triangular Pascal matrix is known. From the decomposition $C = S_\alpha^{-1}P_US_0^{-1}S_\alpha$, we first factorize P_U using its bidiagonal decomposition (as the product of upper bidiagonal matrices with nonzero off-diagonal entries equal to one) and we “move” the diagonal matrices so we end up rewriting C as a product of upper bidiagonal matrices and a diagonal matrix satisfying the hypotheses of Theorem 4.1.

Then, from $\mathcal{BD}(B)$ and $\mathcal{BD}(C)$ we can compute $\mathcal{BD}(M)$ accurately with `TNProduct`. In fact, this reasoning could be applied to other polynomial basis. If we want to compute the bidiagonal decomposition of its collocation matrices to high relative accuracy (so we can obtain accurate results using the functions from `TNTool`), we need to find the accurate bidiagonal decomposition of the matrix of change of basis between the polynomial basis object of study and the monomials.

For the matrix of change of basis C appearing in Theorem 4.2, we have that its bidiagonal decomposition is given by

$$(\mathcal{BD}(C))_{ij} = \begin{cases} 0, & \text{if } i < j, \\ \frac{i+\alpha}{i}, & \text{if } i > j, \\ (i-1)!, & \text{if } i = j. \end{cases} \quad (4.7)$$

4.1.2 Accurate computations with Bessel matrices

In [16] we studied the collocation matrices of Bessel polynomials: we proved that their matrix of change of basis between the monomial basis and the Bessel polynomials is TP, we obtained its bidiagonal decomposition and we used it to solve many algebraic problems with HRA

using the library TNTTool. The Bessel polynomials are defined by

$$B_n(x) = \sum_{k=0}^n \frac{(n+k)!}{2^k(n-k)!k!} x^k, \quad n = 0, 1, 2, \dots, \quad (4.8)$$

Bessel polynomials occur in many different areas such as number theory, partial differential equations and statistics (see [48]). These polynomials are also very important for some problems of Static Potentials, Signal Processing and Electronics.

Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be the matrix of change of basis between the Bessel polynomials and the monomial basis,

$$(B_0(x), B_1(x), \dots, B_{n-1}(x))^T = A(1, x, \dots, x^{n-1})^T, \quad (4.9)$$

that is, the lower triangular matrix A is defined by

$$a_{ij} := \begin{cases} \frac{(i+j-2)!}{2^{j-1}(i-j)!(j-1)!} = \frac{(2j-2)!}{2^{j-1}(j-1)!} \binom{i+j-2}{i-j}, & \text{if } i \geq j, \\ 0, & \text{if } i < j. \end{cases} \quad (4.10)$$

We proved the total positivity of the matrix of change of basis A and we obtained $\mathcal{BD}(A)$.

Theorem 4.3. (Theorem 3 of [16]) *Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be the lower triangular matrix in (4.9) defined by (4.10). Then we have that*

(i) *the pivots of the NE of A are given by*

$$p_{ij} = \frac{1}{2^{j-1}} \frac{(i-1)!}{(i-j)!} \prod_{r=1}^{j-1} \frac{(2i-r-1)}{(i-j+r)}, \quad 1 \leq j \leq i \leq n, \quad (4.11)$$

and the multipliers by

$$m_{ij} = \frac{(2i-2)(2i-3)}{(2i-j-1)(2i-j-2)}, \quad 1 \leq j < i \leq n, \quad (4.12)$$

(ii) *A is a nonsingular TP matrix*

(iii) *and the bidiagonal factorization of A is given by*

$$\mathcal{BD}(A)_{ij} = \begin{cases} \frac{(2i-2)(2i-3)}{(2i-j-1)(2i-j-2)}, & \text{if } i > j, \\ 1, & \text{if } i = j = 1, \\ (2i-3)!!, & \text{if } i = j > 1, \\ 0, & \text{if } i < j, \end{cases} \quad (4.13)$$

and can be computed to HRA. The notation “!!” corresponds to the semifactorial given by $n!! = \prod_{k=0}^{\lfloor n/2 \rfloor - 1} (n - 2k)$.

As a consequence of Theorem 4.3, we have that the system of functions formed by the Bessel polynomials of degree less than n on $(0, \infty)$ is an STP system.

Given a sequence of parameters $0 < t_0 < t_1 < \dots < t_{n-1}$, we call the collocation matrix of the Bessel polynomials (B_0, \dots, B_{n-1}) at that sequence a *Bessel matrix*:

$$M = M \begin{pmatrix} B_0, \dots, B_{n-1} \\ t_0, \dots, t_{n-1} \end{pmatrix} = (B_{j-1}(t_{i-1}))_{1 \leq i, j \leq n}, \quad (4.14)$$

Theorem 4.4. (Theorem 4 of [16]) Given a sequence of parameters $0 < t_0 < t_1 < \dots < t_{n-1}$, the corresponding Bessel matrix M is an STP matrix and given the parametrization t_i ($0 \leq i \leq n-1$), we can compute its bidiagonal decomposition to HRA.

By formula (4.9) we have that $M = VA^T$, where M is the Bessel matrix at t_0, \dots, t_{n-1} , A is the lower triangular matrix defined by (4.10) and V is the Vandermonde matrix corresponding to the collocation matrix of the monomial basis of degree $n-1$ at t_0, \dots, t_{n-1} . $\mathcal{BD}(V)$ is known and can be computed to HRA using the function TNVandBD from TNTTool. Then, we can compute $\mathcal{BD}(M)$ from $\mathcal{BD}(V)$ and, by (3.19), from $\mathcal{BD}(A)^T$ using TNProduct.

Reversing the order of the coefficients of $B_n(x)$ in (4.8) we can define the *reverse Bessel polynomials*:

$$B_n^r(x) = \sum_{k=0}^n \frac{(n+k)!}{2^k(n-k)!k!} x^{n-k}, \quad n = 0, 1, 2, \dots, \quad (4.15)$$

The reverse Bessel polynomials occur in applications such as Electrical Engineering. In particular, they play a key role in network analysis of electrical circuits (see page 145 of [48] and references therein). Their coefficients are known as signless Bessel numbers of the first kind in Combinatorics. They are closely related to the Stirling numbers [49, 98].

Let $C = (c_{ij})_{1 \leq i, j \leq n}$ be the matrix of change of basis between the reverse Bessel polynomials and the monomial basis,

$$(B_0^r(x), B_1^r(x), \dots, B_{n-1}^r(x))^T = C(1, x, \dots, x^{n-1})^T, \quad (4.16)$$

i.e., the lower triangular matrix C defined by

$$c_{ij} = \begin{cases} \frac{(2i-j-1)!}{2^{i-j}(j-1)!(i-j)!}, & i \geq j, \\ 0, & i < j. \end{cases} \quad (4.17)$$

Theorem 4.5 proves the total positivity of C , and provides $\mathcal{BD}(C)$. In addition, its proof gives the explicit form of all the entries of the matrices $C^{(k)}$ computed through the NE of C .

Theorem 4.5. (Theorem 5 of [16]) Let $C = (c_{ij})_{1 \leq i, j \leq n}$ be the lower triangular matrix in (4.16) defined by (4.17). Then, we have that

(i) the pivots of the NE of C are given by

$$\begin{aligned} p_{ij} &= \frac{(2i-2j)!}{2^{i-j}(i-j)!} & 1 \leq j \leq i \leq n & \quad \text{if } j \text{ is odd,} \\ p_{ij} &= 0 & 1 \leq j < i \leq n, & \quad p_{jj} = 1 \quad 1 \leq j \leq n \quad \text{if } j \text{ is even,} \end{aligned} \quad (4.18)$$

and the multipliers by

$$\begin{aligned} m_{ij} &= 2i-1-2j & 1 \leq j < i \leq n & \quad \text{if } j \text{ is odd,} \\ m_{ij} &= 0, & 1 \leq j < i \leq n & \quad \text{if } j \text{ is even,} \end{aligned} \quad (4.19)$$

(ii) C is a nonsingular TP matrix

(iii) and the bidiagonal factorization of C is given by

$$\mathcal{BD}(C)_{ij} = \begin{cases} 2i - 2j - 1, & \text{if } i > j \text{ with } j \text{ odd,} \\ 1, & \text{if } i = j, \\ 0, & \text{otherwise,} \end{cases} \quad (4.20)$$

and can be computed to HRA.

Given a sequence of parameters $0 < t_0 < t_1 < \dots < t_{n-1}$, we call the collocation matrix of the reverse Bessel polynomials $(B_0^r, \dots, B_{n-1}^r)$ at that sequence a *reverse Bessel matrix*:

$$M_r = M \begin{pmatrix} B_0^r, \dots, B_{n-1}^r \\ t_0, \dots, t_{n-1} \end{pmatrix} = (B_{j-1}^r(t_{i-1}))_{1 \leq i, j \leq n}. \quad (4.21)$$

The following result proves that the reverse Bessel matrices are STP and that some usual algebraic problems with these matrices can be solved to HRA.

Theorem 4.6. (Theorem 6 of [16]) *Given a sequence of parameters $0 < t_0 < t_1 < \dots < t_{n-1}$, the corresponding reverse Bessel matrix M_r (4.21) is an STP matrix and given the parametrization t_i ($0 \leq i \leq n-1$), its bidiagonal decomposition can be computed to HRA.*

We can build an algorithm to compute the bidiagonal decomposition of the reverse Bessel matrices to HRA following the same strategy used with the Bessel matrices.

4.1.3 Accurate bidiagonal decomposition and computations with generalized Pascal matrices

In Section 3.5 we introduced Pascal matrices as an example of TP matrices that are ill-conditioned but have a really simple representation in terms of the bidiagonal decomposition. Let us recall that the lower triangular Pascal matrix $P_L = (p_{ij})_{1 \leq i, j \leq n}$ has entries $p_{ij} := \binom{i-1}{j-1}$ for $1 \leq j \leq i \leq n+1$ and $p_{ij} := 0$ whenever $j > i$ and the symmetric Pascal matrix $R = (r_{ij})$ has entries $r_{ij} := \binom{i+j-2}{j-1}$. These matrices satisfy that $R = P_L P_L^T$ and their bidiagonal decomposition is given by:

$$\mathcal{BD}(P_L) = \begin{cases} 1, & \text{if } i \geq j, \\ 0, & \text{otherwise,} \end{cases} \quad (4.22)$$

and $\mathcal{BD}(R) = (1)_{1 \leq i, j \leq n}$ (see [2, 60]). In [18], we considered some classical extensions of Pascal matrices. These classes of generalized Pascal matrices appear in applications such as Filter Design, Probability, Combinatorics, Signal Processing or Electrical Engineering (see [66] and references therein). We have obtained their bidiagonal decomposition and we have studied the cases when these extensions of Pascal matrices are TP and their bidiagonal decomposition can be computed to HRA. Let us start by giving the definitions of the matrices that we have considered:

Definition 4.7. (see [57, 100]) For a real number x , the generalized Pascal matrix of the first kind, $P_n[x]$, is defined as the $(n+1) \times (n+1)$ lower triangular matrix with 1's on the main diagonal and

$$(P_n[x])_{ij} := x^{i-j} \binom{i-1}{j-1}, \quad 1 \leq j \leq i \leq n+1$$

and the symmetric generalized Pascal $(n+1) \times (n+1)$ matrix $R_n[x]$ is given by

$$(R_n[x])_{ij} := x^{i+j-2} \binom{i+j-2}{j-1}, \quad 1 \leq i, j \leq n+1.$$

For $x, y \in \mathbb{R}$, the $(n+1) \times (n+1)$ matrix $R_n[x, y]$ is given by

$$(R_n[x, y])_{ij} := x^{j-1} y^{i-1} \binom{i+j-2}{j-1}, \quad 1 \leq i, j \leq n+1.$$

Let us notice how these matrices and the classical Pascal matrices are related: $R_n[x] = R_n[x, x]$, $P_n[1]$ is the lower triangular Pascal matrix and $R_n[1]$ is the symmetric Pascal matrix. The following definition gives other families of matrices that we have also studied.

Definition 4.8. (see [101]) For $x, y \in \mathbb{R}$, the extended generalized Pascal matrix $\Phi_n[x, y]$ is defined as

$$(\Phi_n[x, y])_{ij} = x^{i-j} y^{i+j-2} \binom{i-1}{j-1}, \quad 1 \leq j \leq i \leq n+1$$

and the extended generalized symmetric Pascal matrix $\Psi_n[x, y]$ is given by

$$(\Psi_n[x, y])_{ij} = x^{i-j} y^{i+j-2} \binom{i+j-2}{j-1}, \quad 1 \leq i, j \leq n+1.$$

We have found the bidiagonal decomposition of the matrices given by definitions 4.7 and 4.8 as particular cases of two wider classes of matrices, one related to the triangular matrices and the other related to the symmetric ones. Given two real numbers x , λ and a nonnegative integer n , we define the notation $x^{n|\lambda}$ as:

$$x^{n|\lambda} := \begin{cases} x(x+\lambda) \cdots (x+(n-1)\lambda), & \text{if } n > 0, \\ 1, & \text{if } n = 0. \end{cases} \quad (4.23)$$

In [5], the generalized lower triangular Pascal matrix $P_{n,\lambda}[x]$ is defined by

$$(P_{n,\lambda}[x])_{i,j} := x^{(i-j)|\lambda} \binom{i-1}{j-1}, \quad 1 \leq j \leq i \leq n+1, \quad (4.24)$$

where n is a natural number and λ and x are both real numbers. The case $\lambda = 0$ leads to the generalized Pascal matrix of the first kind $P_{n,0}[x] = P_n[x]$. The following result provides the bidiagonal decomposition of the generalized Pascal matrix $P_{n,\lambda}[x]$.

Theorem 4.9. (Theorem 5 of [18]) Given $x, \lambda \in \mathbb{R}$ and $n \in \mathbb{N}$, let $P_{n,\lambda}[x]$ be the $(n+1) \times (n+1)$ lower triangular matrix given by (4.24).

i) If $x \neq k\lambda$ for $k = -n+1, \dots, 0, \dots, n-1$, we have that

$$(\mathcal{B}\mathcal{D}(P_{n,\lambda}[x]))_{ij} = \begin{cases} 1, & i = j, \\ x + (i-2j)\lambda, & i > j, \\ 0, & i < j. \end{cases} \quad (4.25)$$

ii) If $x = k\lambda$ for some $k \in \{0, \dots, n-1\}$, we have that

$$(\mathcal{B}\mathcal{D}(P_{n,\lambda}[x]))_{ij} = \begin{cases} 1, & i = j, \\ x + (i-2j)\lambda, & i > j, j \leq k, \\ 0, & \text{otherwise.} \end{cases} \quad (4.26)$$

iii) If $x = -k\lambda$ for some $k \in \{0, \dots, n-1\}$, we have that

$$(\mathcal{B}\mathcal{D}(P_{n,\lambda}[x]))_{ij} = \begin{cases} 1, & i = j, \\ x + (i-2j)\lambda, & 0 < i-j \leq k, \\ 0, & \text{otherwise.} \end{cases} \quad (4.27)$$

Let us notice that we have computed the bidiagonal decomposition of $P_{n,\lambda}[x]$ in general, so for some values of its parameters the matrix is not totally positive. The following result characterizes the values of the parameters x and λ for which $P_{n,\lambda}[x]$ is a TP matrix.

Proposition 4.10. (Corollary 7 of [18]) *Let $P_{n,\lambda}[x]$ be the lower triangular matrix given by (4.24) with $x, \lambda \in \mathbb{R}$ and with $n \in \mathbb{N}$. Then $P_{n,\lambda}[x]$ is a TP matrix if and only if one of the following conditions holds:*

i) $x \geq (n-1)|\lambda|$.

ii) $x = k|\lambda|$ for $k = 0, \dots, n-1$.

In [5], a generalization of $P_{n,\lambda}[x]$ is given in terms of a second real number y and an arbitrary sequence $\mathbf{a} = \{a_n\}_{n \geq 0}$:

$$(P_{n,\lambda}[x, y, \mathbf{a}])_{i,j} := a_{j-1} x^{(i-j)|\lambda} y^{(j-1)|\lambda} \binom{i-1}{j-1}. \quad (4.28)$$

Let us notice that we can write this matrix as a product of $P_{n,\lambda}[x]$ and a diagonal matrix:

$$P_{n,\lambda}[x, y, \mathbf{a}] = P_{n,\lambda}[x] \operatorname{diag}(a_0, a_1 y^{1|\lambda}, \dots, a_n y^{n|\lambda}). \quad (4.29)$$

By (4.29) and Theorem 4.9, we can deduce the bidiagonal decomposition of the matrix $\mathcal{B}\mathcal{D}(P_{n,\lambda}[x, y, \mathbf{a}])$. For example, if $x \neq k\lambda$ for $k = -n+1, \dots, 0, \dots, n-1$, its bidiagonal decomposition is given by

$$(\mathcal{B}\mathcal{D}(P_{n,\lambda}[x, y, \mathbf{a}]))_{ij} = \begin{cases} a_{j-1} y^{(j-1)|\lambda}, & i = j, \\ x + (i-2j)\lambda, & i > j, \\ 0, & i < j. \end{cases} \quad (4.30)$$

We also studied the families given in definitions 4.7 and 4.8 as particular cases of the *lattice path matrices*. We obtained their bidiagonal decomposition and characterized whether they are total positive in terms of the parameters defining them. Let us now introduce the $(n + 1) \times (n + 1)$ lattice path matrix $Lp_n(\alpha, \beta, \gamma) = (k_{ij})_{1 \leq i, j \leq n+1}$, whose entries are given by the recurrence relation

$$\alpha k_{i,j-1} + \beta k_{i-1,j} + \gamma k_{i-1,j-1} = k_{ij}, \quad 2 \leq i, j \leq n + 1, \tag{4.31}$$

with $k_{1j} = \alpha^{j-1}$ for $j \in \{1, \dots, n + 1\}$ and $k_{i1} = \beta^{i-1}$ for $i \in \{1, \dots, n + 1\}$. These matrices were studied in [57]. In Theorem 2.3 of [57] it is shown that $Lp_n(\alpha, \beta, \gamma)$ admits the following factorization

$$Lp_n(\alpha, \beta, \gamma) = P_n[\alpha] D_{\alpha\beta+\gamma}^n (P_n[\beta])^T, \tag{4.32}$$

where $D_{\alpha\beta+\gamma}^n = \text{diag}(1, \alpha\beta + \gamma, \dots, (\alpha\beta + \gamma)^n)$ and $P_n[\delta] = P_{n,0}[\delta]$. Observe that the matrix $Lp_n(\alpha, \beta, \gamma)$ is nonsingular if and only if $\alpha\beta + \gamma \neq 0$. Based on the LDU decomposition given by (4.32), we obtained the bidiagonal decomposition of $Lp_n(\alpha, \beta, \gamma)$.

Theorem 4.11. (Theorem 8 of [18]) *Let $Lp_n(\alpha, \beta, \gamma) = (k_{ij})_{1 \leq i, j \leq n+1}$ be the matrix whose entries are defined by (4.31) with $\alpha\beta + \gamma \neq 0$. Then its bidiagonal decomposition is given by*

$$(\mathcal{BD}(Lp_n(\alpha, \beta, \gamma)))_{ij} = \begin{cases} (\alpha\beta + \gamma)^{i-1}, & \text{if } i = j, \\ \alpha, & \text{if } i > j, \\ \beta, & \text{if } i < j. \end{cases} \tag{4.33}$$

As it was the case with matrices $P_{n,\lambda}[x]$, the lattice path matrices are not always totally positive. The following proposition introduces a case where these matrices are TP and its bidiagonal decomposition can be computed to high relative accuracy.

Proposition 4.12. (Corollary 9 of [18]) *Let $Lp_n(\alpha, \beta, \gamma) = (k_{ij})_{1 \leq i, j \leq n+1}$ be the matrix whose entries are defined by (4.31). If $\alpha, \beta > 0$ and $\alpha\beta + \gamma > 0$, then $Lp_n(\alpha, \beta, \gamma) = (k_{ij})_{1 \leq i, j \leq n+1}$ is an STP matrix. Moreover, if $\gamma \geq 0$, then its bidiagonal decomposition (4.33) can be computed to HRA and it can be used to obtain the eigenvalues, singular values and the inverse of $Lp_n(\alpha, \beta, \gamma)$ with HRA as well as the solution of the linear systems $Lp_n(\alpha, \beta, \gamma)x = b$, where $b = (b_1, \dots, b_{n+1})$ has alternating signs.*

As we have mentioned earlier, the matrices $Lp_n(\alpha, \beta, \gamma)$ have the generalizations of Pascal matrices given in definitions 4.7 and 4.8 as particular cases. Hence, Theorem 4.11 can be used to obtain the bidiagonal decomposition of those classes of matrices. In fact, Theorem 3.1 of [57] gives the following relationship between the classes that can be used to obtain their bidiagonal decompositions from Theorem 4.11.

$$Lp_n(\alpha, \beta, \gamma) = \begin{cases} P_n[x, y], & \text{if } \alpha = 0, \beta = y, \gamma = x, \\ R_n[x, y], & \text{if } \alpha = x, \beta = y, \gamma = 0, \\ \Phi_n[x, y], & \text{if } \alpha = 0, \beta = xy, \gamma = y^2, \\ \Psi_n[x, y], & \text{if } \alpha = y/x, \beta = xy, \gamma = 0. \end{cases} \tag{4.34}$$

There is another interesting extension of Pascal matrices in terms of the q -integers. In the following section we have considered this extension as well as other q -analogues of well-known TP matrices.

4.1.4 High relative accuracy with matrices of q -integers

Some of the recent examples of TP matrices whose bidiagonal decomposition has been obtained to HRA come from quantum calculus [54]. These matrices are usually q -analogues of other well-known families of matrices, such as Pascal matrices or Jacobi-Stirling matrices. There are also examples from other areas, like from extensions of orthogonal polynomials. Given a positive real number q and a natural number r we define the q -integer $[r]$ as

$$[r] := \begin{cases} 1 + q + \cdots + q^{r-1} = \frac{1-q^r}{1-q}, & \text{if } q \neq 1, \\ r, & \text{if } q = 1. \end{cases}$$

Let us define the following q -analogues in terms of the q -integers. The q -factorial $[r]!$ (see [54]) is given by

$$[r]! := \begin{cases} [r][r-1] \cdots [1], & \text{if } q \neq 1, \\ r!, & \text{if } q = 1, \end{cases}$$

and the q -binomial coefficient $\begin{bmatrix} i \\ j \end{bmatrix}$ is defined as

$$\begin{bmatrix} i \\ j \end{bmatrix} := \frac{[i]!}{[j]![i-j]!} \quad (4.35)$$

if $i \geq j \geq 0$ and as 0 otherwise. Let us recall the recurrence relation that defines the classical binomial coefficients $\binom{n}{k}$,

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}. \quad (4.36)$$

The q -binomial coefficients satisfy the following recurrence relations, which are q -analogues of (4.36):

$$\begin{bmatrix} i \\ j \end{bmatrix} = \begin{bmatrix} i-1 \\ j-1 \end{bmatrix} + q^j \begin{bmatrix} i-1 \\ j \end{bmatrix}, \quad (4.37)$$

$$\begin{bmatrix} i \\ j \end{bmatrix} = q^{i-j} \begin{bmatrix} i-1 \\ j-1 \end{bmatrix} + \begin{bmatrix} i-1 \\ j \end{bmatrix}. \quad (4.38)$$

They also satisfy a q -analogue of the Vandermonde identity:

$$\begin{bmatrix} m+n \\ k \end{bmatrix} = \sum_{j=0}^k q^{(k-j)(m-j)} \begin{bmatrix} m \\ j \end{bmatrix} \begin{bmatrix} n \\ k-j \end{bmatrix}. \quad (4.39)$$

Let us also define the lower triangular matrix of q -binomial coefficients, $P_{L,q}$, whose nonzero entries are given by

$$(P_{L,q})_{i,j} = \begin{bmatrix} i-1 \\ j-1 \end{bmatrix}, \quad 1 \leq j \leq i \leq n+1, \quad (4.40)$$

and its upper triangular counterpart $P_{U,q} := P_{L,q}^T$. The matrix $P_{L,q}$ is a TP matrix (see page 198 of [7]) and we obtained the following bidiagonal decomposition

Theorem 4.13. (Theorem 3 of [20]) Let $P_{L,q}$ be the $(n + 1) \times (n + 1)$ matrix given by (4.40). Then $P_{L,q}$ is TP and its bidiagonal decomposition is given by

$$(\mathcal{BD}(P_{L,q}))_{i,j} = \begin{cases} 1, & i = j, \\ q^{j-1}, & i > j, \\ 0, & \text{otherwise,} \end{cases} \quad (4.41)$$

which can be computed to HRA.

We see that $q = 1$ gives the bidiagonal decomposition of the lower triangular Pascal matrix P_L . Let us now define the symmetric matrix of q -binomial coefficients P_q :

$$(P_q)_{i,j} = \begin{bmatrix} i+j-2 \\ i-1 \end{bmatrix}, \quad 1 \leq i, j \leq n+1. \quad (4.42)$$

The matrix P_q is the q -analogue of the Pascal matrix R defined in the previous section. We can derive the bidiagonal decomposition of this matrix from $\mathcal{BD}(P_{L,q})$.

Proposition 4.14. (Proposition 1 of [20]) Let P_q be the matrix of q -binomial coefficients given by (4.42). Then P_q is STP and its bidiagonal decomposition is given by

$$(\mathcal{BD}(P_q))_{i,j} = \begin{cases} q^{(j-1)^2}, & i = j, \\ q^{j-1}, & i > j, \\ q^{i-1}, & \text{otherwise,} \end{cases} \quad (4.43)$$

which can be computed to HRA.

We can also check that the case $q = 1$ gives the bidiagonal decomposition of the symmetrical classical Pascal matrix.

Our next family of examples comes from q -analogues of the Stirling numbers of the first and the second kind [4]. The bidiagonal decomposition of the matrices formed by the Stirling numbers (called Stirling matrices) was studied in [25]. The q -Stirling numbers of the second kind, $B_q = (b_{ij})_{1 \leq i, j \leq n+1}$, are given by the recurrence relation (see [35])

$$b_{ij} = b_{i-1, j-1} + [j-1]b_{i-1, j}, \quad (4.44)$$

with $b_{00} = 1, b_{i0} = 0$ for $i > 0$ and $b_{0j} = 0$ for $j > 0$. The q -Stirling numbers of the first kind, $S_q = (s_{ij})_{1 \leq i, j \leq n+1}$, follow the relationship (see [35])

$$s_{ij} = s_{i-1, j-1} - [i-1]s_{i-1, j}, \quad (4.45)$$

with $s_{00} = 1, s_{i0} = 0$ for $i > 0$ and $s_{0j} = 0$ for $j > 0$. Let us define the unsigned q -Stirling numbers of the first kind, $C_q = (c_{ij})_{1 \leq i, j \leq n+1}$, by the following relationship

$$c_{ij} = c_{i-1, j-1} + [i-1]c_{i-1, j}, \quad (4.46)$$

with $c_{00} = 1, c_{i0} = 0$ for $i > 0$ and $c_{0j} = 0$ for $j > 0$. The entries of S_q are equal in absolute value to those of C_q . The difference lies on their sign pattern: S_q has a checkerboard pattern of alternating signs while $C_q \geq 0$. The following proposition gives the bidiagonal decomposition of C_q .

Proposition 4.15. Let $C_q = (c_{ij})_{1 \leq i, j \leq n+1}$ be the matrix whose (i, j) entry is the unsigned q -Stirling number of the first kind c_{ij} given by (4.46). Then C_q is TP and

$$\mathcal{BD}(C_q) = \begin{cases} 1, & i = j, \\ [i - j], & i > j, \\ 0, & \text{otherwise.} \end{cases}$$

By using (4.45) instead of (4.46), the same proof of Proposition 4.15 leads to the bidiagonal decomposition of S_q

$$\mathcal{BD}(S_q) = \begin{cases} 1, & i = j, \\ -[i - j], & i > j, \\ 0, & \text{otherwise.} \end{cases} \quad (4.47)$$

In spite that S_q is not a TP matrix, it is closely related to this class of matrices since it is the inverse of the matrix B_q (by Theorem 3.16 of [35]). Based on that fact, from (4.47) we deduced $\mathcal{BD}(B_q)$.

Proposition 4.16. (Corollary 3 of [20]) Let $B_q = (b_{ij})_{1 \leq i, j \leq n+1}$ be the matrix whose (i, j) entry is the q -Stirling number of the second kind b_{ij} given by (4.44). Then B_q is TP and

$$\mathcal{BD}(B_q) = \begin{cases} 1, & i = j, \\ [j], & i > j, \\ 0, & \text{otherwise.} \end{cases}$$

Finally, we have considered a q -analogue of the Laguerre polynomials. Let us define the q -Laguerre polynomials $L_{n,q}^{(\alpha)}$ (see p. 552 of [52]):

$$L_{n,q}^{(\alpha)}(x) := \frac{(q^{\alpha+1}; q)_n}{(q; q)_n} \sum_{k=0}^n \begin{bmatrix} n \\ k \end{bmatrix} q^{\alpha k + k^2} \frac{(-x)^k}{(q^{\alpha+1}; q)_k}. \quad (4.48)$$

The following result shows the strict total positivity of collocation matrices M of q -Laguerre polynomials and guarantees HRA for many algebraic computations with whenever $\alpha > -1$ is a rational number.

Theorem 4.17. (Theorem 6 and Corollary 4 of [20]) Let $M := (L_{j-1,q}^{(\alpha)}(t_{i-1}))_{1 \leq i, j \leq n+1}$ be the collocation matrix of the q -Laguerre polynomials at the nodes $(0 >) t_0 > t_1 > \dots > t_n$ with $\alpha > -1$ and $0 < q < 1$. Then

i) M is an STP matrix.

ii) If $\alpha \in \mathbb{Q}$, given the nodes t_i ($0 \leq i \leq n$) we can compute $\mathcal{BD}(M)$ with HRA and hence, the following computations can be performed with HRA: all the eigenvalues and singular values, the inverse of M , and the solution of the linear systems $Mx = b$ where $b = (b_0, \dots, b_n)$ has alternating signs.

4.2 Optimal properties of tensor products of B -bases

In Section 3.3 we introduced totally positive bases and the concept of B -basis. Normalized B -bases have the optimal shape preserving properties with respect to all the other NTP bases of their spanned function space. In [28], it was shown that the minimal eigenvalue and singular value of a collocation matrix of the normalized B -basis of a space of functions is bounded below by the minimal eigenvalue and singular value, respectively, of the corresponding collocation matrix of any other normalized totally positive basis of the same space. It is also proved that the collocation matrix of the normalized B -basis of a space of functions has optimal conditioning in the ∞ -norm with respect to all the normalized TP bases of the space. In [19], we have considered the collocation matrices of the tensor product of normalized B -bases. Given two systems of functions $u^1 = (u_0^1, \dots, u_m^1)$ and $u^2 = (u_0^2, \dots, u_n^2)$ defined in $[a, b]$ and $[c, d]$, respectively, the system $u^1 \otimes u^2 := (u_i^1(x) \cdot u_j^2(y))_{i=0, \dots, m}^{j=0, \dots, n}$ is called a tensor product system and generates a tensor product surface. Let us consider two increasing sequences of nodes $\mathbf{t} = (t_i)_{i=0}^m$ in $[a, b]$ and $\mathbf{r} = (r_i)_{i=0}^n$ in $[c, d]$ and the collocation matrices A_1 and A_2 of the bases u^1 and u^2 at \mathbf{t} and \mathbf{r} , respectively. Then the collocation matrix of the system $u^1 \otimes u^2$ at the points (t_i, r_j) with $i = 0, \dots, m$ and $j = 0, \dots, n$ can be formed directly from A_1 and A_2 using the *Kronecker product*. The Kronecker product of two matrices $A = (a_{ij}) \in \mathbb{R}^{m_1 \times n_1}$ and $B = (b_{ij}) \in \mathbb{R}^{m_2 \times n_2}$ is defined to be the $m_1 m_2 \times n_1 n_2$ block matrix

$$A \otimes B := \begin{pmatrix} a_{11}B & \cdots & a_{1n_1}B \\ \vdots & \ddots & \vdots \\ a_{m_1 1}B & \cdots & a_{m_1 n_1}B \end{pmatrix}. \quad (4.49)$$

The Kronecker product presents a lot of useful properties. It serves as a great tool for studying high dimensional problems taking advantage of the theory and techniques known for the lower dimensional case. Many of the fundamental structures and properties desired for matrices are inherited by the Kronecker product if both of the smaller factors present them. For instance, this is true for the Kronecker product of two nonsingular matrices, of two symmetric or triangular matrices, of two positive definite matrices, or of two orthogonal matrices (see [96]). Besides, the eigenvalues and singular values of the Kronecker product are given by products of the eigenvalues and singular values of the smaller matrices defining it. Also, whenever A and B are nonsingular, we have that $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$. Unfortunately, total positivity or the structure of a P -matrix is in general not inherited by the Kronecker product of two TP matrices or of two P -matrices, respectively. In our work, exploiting the nice properties of the Kronecker product has allowed us to obtain the following result for the optimality of the tensor product of normalized B -bases.

Theorem 4.18. (Theorem 1 of [19]) *Let $u^1 = (u_0^1, \dots, u_m^1)$ be an NTP basis on $[a, b]$ of a space of functions \mathcal{U}_1 , $u^2 = (u_0^2, \dots, u_n^2)$ be an NTP basis on $[c, d]$ of a space of functions \mathcal{U}_2 and let $v^1 = (v_0^1, \dots, v_m^1)$ and $v^2 = (v_0^2, \dots, v_n^2)$ be the normalized B -bases of \mathcal{U}_1 and \mathcal{U}_2 , respectively. Given the increasing sequences of nodes $\mathbf{t} = (t_i)_{i=0}^m$ on $[a, b]$ and $\mathbf{r} = (r_i)_{i=0}^n$ on $[c, d]$, the nonsingular collocation matrices A_1 and M_1 of the bases u^1 and v^1 , respectively, at \mathbf{t} , and A_2 and M_2 of the bases u^2 and v^2 , respectively, at \mathbf{r} , the following properties hold*

- i) The matrix $|(A_1 \otimes A_2)^{-1}|$ dominates $(M_1 \otimes M_2)^{-1}$.
- ii) The minimal eigenvalue (resp., singular value) of $A_1 \otimes A_2$ is bounded above by the minimal eigenvalue (resp., singular value) of $M_1 \otimes M_2$.
- iii) $\kappa_\infty(M_1 \otimes M_2) \leq \kappa_\infty(A_1 \otimes A_2)$.

In [19] we have also included a section with numerical experiments that illustrate the results from Theorem 4.18. We have compared the minimal singular values, minimal eigenvalues and the condition number using the ∞ -norm of the Kronecker product of two collocation matrices of the Bernstein polynomials on $[0, 1]$ with the Kronecker product of two collocation matrices of the DP basis [23] and two collocation matrices of the Said-Ball basis [91]. For the comparison, all the collocation matrices have been built using the same sequences of increasing positive nodes. We have also compared the case of collocation matrices of rational Bernstein basis with the associated rational basis of the other studied NTP bases. In this case, the weights defining the rational functions were chosen such as the space generated by all the rational bases was always the same, so we were under the hypotheses of Theorem 4.18 and the results were comparable. In both cases, we checked that the minimal eigenvalues and singular values for the Kronecker product of the normalized B -bases are larger than the minimal eigenvalue and the minimal singular value, respectively, of the other matrices. We also showed that the ∞ -norm condition number of the Kronecker product of the collocation matrices of B -bases is smaller than the ∞ -norm condition number of the Kronecker product of the collocation matrices of the other normalized TP bases.

4.3 Tridiagonal Toeplitz P -matrices

In spite of the fact that, for a general Toeplitz matrix, we cannot assure the HRA for a simple algebraic computation such as the determinant (see Section 3.5), in [21] we show that this fact changes if the Toeplitz matrix is also tridiagonal. Let us recall that an $n \times n$ Toeplitz matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ is a real matrix such that all its diagonals are constant. These matrices can be defined through a sequence of $2n - 1$ real numbers $\{\alpha_k\}_{-n+1}^{n-1}$ with

$$a_{ij} := \alpha_{i-j}, \quad 1 \leq i, j \leq n. \quad (4.50)$$

If an $n \times n$ Toeplitz matrix is also tridiagonal, it can be uniquely represented with 3 parameters:

$$T_n(a, b, c) := \begin{pmatrix} a & c & & & \\ b & a & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & c \\ & & & b & a \end{pmatrix}. \quad (4.51)$$

Given a positive matrix $A = (a_{ij})_{1 \leq i, j \leq n}$, the following condition is sufficient for its total positivity (see [56] or section 2.6 of [89]):

$$a_{ij}a_{i+1, j+1} \geq 4 \cos^2 \left(\frac{\pi}{n+1} \right) a_{i, j+1}a_{i+1, j}, \quad (4.52)$$

with $i, j = 1, \dots, n-1$. If all these inequalities are strict, then A is STP. In the following proposition we have characterized whenever the tridiagonal Toeplitz matrix $T_n(a, b, c)$ is TP or an M -matrix. The condition that characterizes these classes is closely related to the condition given by (4.52).

Proposition 4.19. (Proposition 3.1 and Corollary 3.2 of [21]) Let $A = T_n(a, b, c)$ be the tridiagonal Toeplitz matrix given by (4.51). Then A is TP if and only if

$$a, b, c \geq 0, \quad a \geq 2\sqrt{bc} \cos \left(\frac{\pi}{n+1} \right), \quad (4.53)$$

and A is an M -matrix if and only if

$$a \geq 2\sqrt{bc} \cos \left(\frac{\pi}{n+1} \right) \text{ and } b, c \leq 0. \quad (4.54)$$

It is known (see page 59 of [92]) that the eigenvalues of the $n \times n$ tridiagonal Toeplitz matrix $T_n(a, b, c)$ are given by

$$\lambda_k = a + 2\sqrt{bc} \cos \left(\frac{k\pi}{n+1} \right), \quad k = 1, \dots, n. \quad (4.55)$$

As it can be seen in the proof of Proposition 3.1 of [21], the conditions (4.53) and (4.54) correspond to the condition that all the eigenvalues of $T_n(a, b, c)$ are positive. Finally, we have also studied the cases where $T_n(a, b, c)$ is a P -matrix. By the definition of P -matrix we always have that $a > 0$. The conditions on b and c are summarized in the following theorem.

Theorem 4.20. (Theorem 3.4 of [21]) Let $A = T_n(a, b, c)$ be the tridiagonal Toeplitz matrix given by (4.51). Then A is a P -matrix if and only if one of the following two conditions holds:

- (i) $bc \leq 0$ and $a > 0$.
- (ii) $bc \geq 0$ and $a > 2\sqrt{bc} \cos \left(\frac{\pi}{n+1} \right)$.

Until now, we have achieved HRA with TP matrices. For the case of tridiagonal P -matrices, it is possible to achieve accurate computations in more cases. For the case of a sign skew-symmetric tridiagonal matrix, it is possible to compute its bidiagonal decomposition, all its minors and its inverse with HRA.

Theorem 4.21. (Proposition 4.1 and Theorem 4.2 of [21]) Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a tridiagonal matrix such that $a_{ii} > 0$ for $i = 1, \dots, n$ and $a_{i+1,i}a_{i,i+1} \leq 0$ for $i = 1, \dots, n-1$. Then

$$\mathcal{BD}(A) = \begin{pmatrix} \delta_1 & \frac{a_{12}}{\delta_1} & & & \\ \frac{a_{21}}{\delta_1} & \delta_2 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \frac{a_{n-1,n}}{\delta_{n-1}} \\ & & & \frac{a_{n,n-1}}{\delta_{n-1}} & \delta_n \end{pmatrix}, \quad (4.56)$$

where δ_i are the diagonal pivots associated to the NE of A . The diagonal pivots satisfy the following recurrence relation:

$$\delta_1 = a_{11}, \quad \delta_i = a_{ii} - \frac{a_{i,i-1}a_{i-1,i}}{\delta_{i-1}} \quad i = 2, \dots, n. \quad (4.57)$$

If we know the entries of A with HRA then we can compute $\mathcal{BD}(A)$ (4.56) to HRA, and hence, all the minors and the inverse of A can be computed to HRA.

Let us notice that Theorem 4.21 is the only result for tridiagonal matrices that are not necessarily Toeplitz matrices. Finally, we have also considered the case where $T_n(a, b, c)$ is a sign symmetric tridiagonal Toeplitz P -matrix. By Theorem 4.19, the P -matrices corresponding to this case are either nonsingular M -matrices or nonsingular TP matrices. In this case, we require the additional parameter $a^2 - 4bc$ to be positive and known to HRA.

Let us recall that the inverse of a nonsingular tridiagonal M -matrix is TP (see [84]). We are going to obtain the bidiagonal decomposition of an M -matrix $A = T_n(a, -b, -c)$. From the $\mathcal{BD}(A)$ obtained in Theorem 4.22, in Theorem 4.23 we shall deduce $\mathcal{BD}(A^{-1})$.

Theorem 4.22. Let $A = T_n(a, -b, -c)$ be a nonsingular M -matrix given by (4.51). Then

$$\mathcal{BD}(A) = \begin{pmatrix} \delta_1 & -\frac{c}{\delta_1} & & & \\ -\frac{b}{\delta_1} & \delta_2 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & -\frac{c}{\delta_{n-1}} \\ & & & -\frac{b}{\delta_{n-1}} & \delta_n \end{pmatrix}, \quad (4.58)$$

where δ_i are the diagonal pivots associated to the NE of A and are given by:

$$\delta_1 = a, \quad \delta_i = a - \frac{bc}{\delta_{i-1}} \quad \text{with } i = 2, \dots, n. \quad (4.59)$$

Moreover, if we know a, b, c with HRA and $a^2 - 4bc$ is a positive number known with HRA, then we can compute $\mathcal{BD}(A)$ (4.58) to HRA.

The following result provides the bidiagonal decomposition of the inverse of a nonsingular tridiagonal Toeplitz M -matrix.

Theorem 4.23. *Let $A = T_n(a, -b, -c)$ be a nonsingular M -matrix. Then A^{-1} is a TP matrix and*

$$\mathcal{BD}(A^{-1}) = \begin{pmatrix} 1/\delta_n & c/\delta_{n-1} & c/\delta_{n-2} & \cdots & c/\delta_1 \\ b/\delta_{n-1} & 1/\delta_{n-1} & 0 & \cdots & 0 \\ b/\delta_{n-2} & 0 & 1/\delta_{n-2} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ b/\delta_1 & 0 & \cdots & 0 & 1/\delta_1 \end{pmatrix}, \quad (4.60)$$

where δ_i are the diagonal pivots associated to the NE of A for $i = 1, \dots, n$.

Finally, let us notice that we can get the bidiagonal decomposition of a TP matrix $T_n(a, b, c)$ from Theorem 4.22. In that case, we could get the following result

Theorem 4.24. *Let $A = T_n(a, b, c)$ be a nonsingular TP matrix given by (4.51). Then*

$$\mathcal{BD}(A) = \begin{pmatrix} \delta_1 & \frac{c}{\delta_1} & & & \\ \frac{b}{\delta_1} & \delta_2 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \frac{c}{\delta_{n-1}} \\ & & & \frac{b}{\delta_{n-1}} & \delta_n \end{pmatrix}, \quad (4.61)$$

where δ_i are the diagonal pivots associated to the NE of A and are given by:

$$\delta_1 = a, \quad \delta_i = a - \frac{bc}{\delta_{i-1}} \quad \text{with } i = 2, \dots, n. \quad (4.62)$$

Moreover, if we know a, b, c with HRA and $a^2 - 4bc$ is a positive number known with HRA, then we can compute $\mathcal{BD}(A)$ (4.58) to HRA.

As we can see, for tridiagonal Toeplitz matrices the difference between TP matrices and M -matrices lies on the sign of the off-diagonal entries, or analogously, on the sign of the multipliers of their associated bidiagonal decomposition.

Chapter 5

M-matrices and related problems

Let us recall that the core of this thesis is formed by a collection of publications that can be divided into two main groups: articles on *M*-matrices and problems related to them and articles on TP matrices. In this chapter we provide the thematic unit of the articles [79–83], belonging to the first class:

- [79] **Article 3:** H. Orera and J. M. Peña. Accurate inverses of Nekrasov *Z*-matrices. *Linear Algebra Appl.* 574 (2019), 46-59.
- [81] **Article 4:** H. Orera and J. M. Peña. Infinity norm bounds for the inverse of Nekrasov matrices using scaling matrices. *Appl. Math. Comput.* 358 (2019), 119-127.
- [80] **Article 5:** H. Orera and J. M. Peña. B_{π}^R -tensors. *Linear Algebra Appl.* 581 (2019), 247-259.
- [82] **Article 10:** H. Orera and J. M. Peña. Accurate determinants of some classes of matrices. *Linear Algebra Appl.* 630 (2021), 1-14.
- [83] **Article 11:** H. Orera and J. M. Peña. Error bounds for linear complementarity problems of B_{π}^R -matrices. *Comput. Appl. Math.* 40 (2021), Paper No. 94, 13 pp.

We introduced *M*-matrices in Section 3.2. Let us recall that *M*-matrices have a particular sign structure: nonnegative diagonal entries and nonpositive off-diagonal entries. Moreover, nonsingular *M*-matrices have an entrywise nonnegative inverse and can be characterized by a wide range of properties (see Theorem 3.7). They play an important role in many applications, which credits the interest that raised in their study for theoretical reasons and practical use. For example, they are studied for the establishment of convergence criteria for iterative methods used to solve large sparse systems of linear equations, in the solution of the linear complementarity problem or in economics when considering a Leontief's input-output analysis. One of the great properties discovered for this class is the one object of study in this dissertation: their structure can be exploited to achieve HRA while solving some of the more commons problems in linear algebra. In [1] it was shown that, even though an *M*-matrix might be ill-conditioned in the traditional sense, if we know its row sums and off-diagonal

entries with enough accuracy we can assure that the determinant, the inverse and the smallest eigenvalue can be computed accurately. For that, we will use an algorithm that takes as input this different representation of the matrix. In Section 3.5 we introduced this representation and we called it the DD-parameters (3.32).

One of our main objectives was extending the classes of M -matrices for which computations with HRA can be achieved. In [79, 82] we found new classes and we discussed adequate parametrizations for achieving high relative accuracy. In the next two sections we will introduce these classes of matrices, the parametrizations and the problems that can be solved to HRA. Section 5.3 introduces the error bounds for the LCP of Nekrasov matrices as well as the infinity norm bounds for their inverses obtained in [81]. Finally, Section 5.4 is devoted to B_π^R -matrices and their extension to the higher order case. It presents the error bounds for the LCP of B_π^R -matrices from [83] as well as the class of B_π^R -tensors and the theoretical properties of this class obtained in [80].

5.1 High relative accuracy for Nekrasov Z -matrices with positive diagonal entries

The first class of matrices that we have studied is called *Nekrasov* matrices. Let $N := \{1, \dots, n\}$. Given a complex matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ with $a_{ii} \neq 0$ for all $i \in N$, let us define

$$h_1(A) := \sum_{j \neq 1} |a_{1j}|, \quad h_i(A) := \sum_{j=1}^{i-1} |a_{ij}| \frac{h_j(A)}{|a_{jj}|} + \sum_{j=i+1}^n |a_{ij}|, \quad i = 2, \dots, n. \quad (5.1)$$

The matrix A is called a *Nekrasov matrix* if $|a_{ii}| > h_i(A)$ for all $i \in N$ (see [13–15, 94]). Nekrasov matrices are nonsingular H -matrices. A Nekrasov Z -matrix with positive diagonal entries is a nonsingular M -matrix.

The parametrization that we consider for an $n \times n$ Nekrasov Z -matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ with positive diagonal entries is given by the following n^2 parameters, which we introduced with the name *N -parameters* in [79]:

$$\begin{cases} a_{ij}, & i \neq j, \\ \Delta_j(A) := a_{jj} - h_j(A), & j \in N. \end{cases} \quad (5.2)$$

As we have seen in Theorem 3.9 of Section 3.2, any H -matrix A is characterized by the existence of a positive diagonal matrix D such as AD is SDD. And, if the matrix A has the sign structure of a Z -matrix with positive diagonal entries, then it is an M -matrix.

The idea of finding a good scaling diagonal matrix for an M -matrix gave us a hint about finding good representations for more classes of matrices. In fact, if we can exploit this property with the right scaling matrix, we can apply the accurate algorithms known for nonsingular DD M -matrices to more classes of nonsingular M -matrices. For achieving this goal, we do not require that the product is SDD, but only DD and nonsingular (which made the development of the HRA methods easier in our experience). For a Nekrasov matrix A , the simple diagonal matrix

$$S = \begin{pmatrix} \frac{h_1(A)}{a_{11}} & & & \\ & \frac{h_2(A)}{a_{22}} & & \\ & & \ddots & \\ & & & \frac{h_n(A)}{a_{nn}} \end{pmatrix} \quad (5.3)$$

holds that AS is a DD matrix (see Lemma 2.2 of [79]). The following theorem (Theorem 2.3 of [79]) shows that we can compute the DD-parameters of the DD M -matrix AS if we know the N -parameters (5.2) of a Nekrasov Z -matrix A with positive diagonal entries. Hence, it serves as a base for developing accurate algorithms for this new class of matrices based on the techniques known for DD M -matrices.

Theorem 5.1. (Theorem 2.3 of [79]) *Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a Nekrasov Z -matrix with positive diagonal entries and let S be the matrix given by (5.3). Given the n^2 N -parameters (5.2), we can compute the row sums and the off-diagonal entries of AS (its DD-parameters (3.32)) by a SF algorithm (and so, with HRA), with at most $3n(n-1)/2$ additions, $2n(n-1)$ multiplications and $2n-1$ quotients.*

So, computing the DD-parameters of AS takes $\mathcal{O}(n^2)$ elementary operations. Using these parameters as input, we can adapt Gaussian elimination (or Gauss-Jordan elimination) to compute the inverse of A , its determinant and the solution to linear systems of equations $Ax = b$ with $b \geq 0$ to HRA. In [79], we introduced the algorithm for computing the inverse and the solution of linear systems of equations to HRA. In [82], we showed that we can also compute the determinant to HRA, and based on that method we computed the determinants of B -Nekrasov matrices (see Section 3.2) to HRA. Algorithm 1 of [82] showed how an adapted version of Gauss-Jordan elimination can be implemented to work with the DD-parameters of a nonsingular DD M -matrix. The following result combines Theorem 3.2 of [79] and Theorem 3.2 of [82].

Theorem 5.2. *Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a Nekrasov Z -matrix with positive diagonal entries. If we know its n^2 N -parameters (5.2), then we can compute its determinant or its inverse to HRA using a SF algorithm of $\mathcal{O}(n^3)$ elementary operations.*

As a consequence, we can also compute to HRA the solution to linear systems of equations $Ax = b$ whenever $b \geq 0$, since A being a nonsingular M -matrix implies that $A^{-1} \geq 0$. In [79], we implemented and tested the accuracy of the HRA methods for computing the inverse and the solution to these particular linear systems of equations. We compared the results to the ones obtained with the Matlab functions `inv` for computing the inverse and `\` for solving linear systems of equations that use the original matrix (instead of the N -parameters) as input.

5.2 Accurate computation of the determinant of B -matrices

In [82] we considered the problem of computing the determinant of two classes of matrices: B -matrices and B -Nekrasov matrices to HRA. The techniques and discussion developed in

that manuscript are a direct continuation of the work started in [79].

Let us first present the classes of matrices studied. *B*-matrices were introduced in [85], where they were used to develop criteria for the localization of eigenvalues. One of the advantages of this class is that it gives an easy to check condition to identify some *P*-matrices. We have introduced *P*-matrices in Section 3.1. One of the underlying problems for this class is that they are difficult to identify in practice. Algorithms for checking that a general matrix is a *P*-matrix usually have a huge computational cost, so looking for good criteria for identifying them is an interesting research topic. Moreover, this problem also extends to the higher dimension case: the identification of classes of tensors (hypermatrices) having similar properties has an even higher computational cost. So, simple criteria for identifying subclasses of *P*-matrices can be of interest in this area, as they could also be extended to identify some classes of tensors as *P*-tensors taking a reasonable computational effort. Let us start by recalling the definition of a *B*-matrix [85].

Definition 5.1. A square real matrix $A := (a_{ij})_{1 \leq i, j \leq n}$ with positive row sums is a *B*-matrix if all its off-diagonal elements are bounded above by the corresponding row means, i.e., for all $i = 1, \dots, n$,

$$\sum_{j=1}^n a_{ij} > 0, \quad \frac{1}{n} \left(\sum_{k=1}^n a_{ik} \right) > a_{ij} \quad \forall j \neq i. \quad (5.4)$$

B-matrices admit the following decomposition, which we will use to achieve accurate computations with them. Let us first introduce the following notation. Given a real matrix $B = (b_{ij})_{1 \leq i, j \leq n}$, we define for each $i = 1, \dots, n$, $r_i^+ := \max_{j \neq i} \{0, b_{ij}\}$. Then B can be decomposed in the form

$$B = B^+ + C, \quad (5.5)$$

$$B^+ = \begin{pmatrix} b_{11} - r_1^+ & \dots & b_{1n} - r_1^+ \\ \vdots & & \vdots \\ b_{n1} - r_n^+ & \dots & b_{nn} - r_n^+ \end{pmatrix}, \quad C = \begin{pmatrix} r_1^+ & \dots & r_1^+ \\ \vdots & & \vdots \\ r_n^+ & \dots & r_n^+ \end{pmatrix}. \quad (5.6)$$

Observe that, if B is a *B*-matrix, then B^+ is an SDD *Z*-matrix (see Proposition 3.11). Therefore, for each $i = 1, \dots, n$,

$$d_{ii} = \sum_{j=1}^n (b_{ij} - r_i^+) > 0. \quad (5.7)$$

Given a DD *Z*-matrix $A = (a_{ij})_{1 \leq i, j \leq n}$, the n^2 parameters used to assure that many algebraic computations can be performed with HRA are the off-diagonal entries of A and the n (nonnegative) row sums of A (i.e., the DD-parameters (3.32)). The n^2 parameters of a *B*-matrix $B = (b_{ij})_{1 \leq i, j \leq n}$ that will be used to compute its determinant with HRA will be again its off-diagonal entries (as with diagonally dominant *Z*-matrices and with Nekrasov *Z*-matrices) and the n positive parameters given by (5.7). We called these n^2 parameters of a *B*-matrix its *B*-parameters in [82]:

$$\begin{cases} b_{ij}, & i \neq j, \\ d_{ii}, & i \in N. \end{cases} \quad (5.8)$$

If we know the B -parameters of a B -matrix with HRA, then we can compute its determinant to HRA.

Theorem 5.2. (Theorem 2.2. of [82]) Let $B = (b_{ij})_{1 \leq i, j \leq n}$ be a B -matrix. Given its B -parameters (see (5.8)) we can compute $\det B$ with HRA.

Moreover, the method developed in [82] for computing the determinant of a B -matrix to HRA has a computational cost of $\mathcal{O}(n^3)$ elementary operations. It is described in Algorithm 2 of [82].

We could have defined B -matrices from the decomposition given by (5.5) and (5.6), saying that a B -matrix is any matrix that can be written in the form (5.5) with C being a nonnegative matrix and B^+ being an SDD Z -matrix with positive diagonal entries. This definition shows the close relationship between B -matrices and SDD M -matrices. In fact, from this point of view new classes of matrices have been defined imposing different conditions than diagonal dominance on B^+ . If that condition implies that the Z -matrix B^+ is also an M -matrix, some of the nice properties of B -matrices can also be derived for the new class of matrices. For example, this is the case for the class of B -Nekrasov matrices [43].

We say that B is a B -Nekrasov matrix if given the decomposition defined by (5.5) and (5.6), the matrix B^+ in (5.6) is a Nekrasov Z -matrix with positive diagonal entries. This class of matrices includes both B -matrices and Nekrasov matrices.

We have seen that it is possible to compute the determinants of those two classes of matrices to HRA whenever the adequate parametrization is known with HRA. For B -Nekrasov matrices, we found the following n^2 parameters that we called BN -parameters in [82]:

$$\begin{cases} b_{ij}, & i \neq j, \\ \Delta_j(B^+), & j \in N, \end{cases} \quad (5.9)$$

where $\Delta_j(\cdot)$ is given by (5.1) and (5.2). Based on this parametrization we obtained the following result for B -Nekrasov matrices.

Theorem 5.3. (Theorem 4.1 of [82]) Let $B = (b_{ij})_{1 \leq i, j \leq n}$ be a B -Nekrasov matrix. Given its BN -parameters (see (5.9)) we can compute $\det B$ with HRA.

As it was the case with B -matrices and Nekrasov matrices, our method for computing the determinant of a B -Nekrasov matrix to HRA has a computational cost of $\mathcal{O}(n^3)$ elementary operations. The method for computing the determinant of a B -Nekrasov matrix to HRA is described in pseudocode in Algorithm 5 of [82].

5.3 Bounds based on diagonal scaling for Nekrasov matrices

Finding an appropriate scaling matrix S for a Nekrasov M -matrix has allowed us to achieve the accurate computation of the inverse, the determinant and the solution of some linear systems of equations. We considered the application of this scaling matrix to more problems. With a modification, these scaling matrices can be used to derive upper bounds for the norm of the inverse of a Nekrasov matrix. In this case, the bound is achieved for any sign structure, not only that of an M -matrix, and so we have considered H -matrices.

5.3.1 Infinity norm bounds for the inverse of Nekrasov matrices

SDD matrices are a clear example of H -matrices (see Theorem 3.9). In [97], the following simple bound for the norm of the inverse of an SDD matrix was introduced.

Theorem 5.4. *Let A be an SDD matrix and let $\alpha := \min_k (|a_{kk}| - \sum_{j \neq k} |a_{kj}|)$. Then $\|A^{-1}\|_\infty < 1/\alpha$.*

An H -matrix A is characterized by the existence of a diagonal scaling matrix S such that AS is SDD. If we know such a matrix S for a subclass of H -matrices, then we can take advantage of Theorem 5.4 to develop bounds for the norm of the inverse of matrices belonging in that class. A simple way to achieve this would be:

$$\|A^{-1}\|_\infty = \|S(S^{-1}A^{-1})\|_\infty = \|S(AS)^{-1}\|_\infty \leq \|S\|_\infty \|(AS)^{-1}\|_\infty, \quad (5.10)$$

which corresponds to equation (3) in the proof of Theorem 3.2 of [81]. Precisely, in [81] we studied the use of this technique for finding suitable bounds for Nekrasov matrices. Our starting point was the diagonal matrix S (5.3) that we used in [79] to develop accurate methods for Nekrasov M -matrices. However, that matrix only satisfied that its product with a Nekrasov matrix is DD, so it needed a modification.

With that idea, we considered the diagonal matrices introduced in Theorem 2.1 and Theorem 2.2 of [81]. The following result introduces the two diagonal scaling matrices considered in the theorems.

Theorem 5.5. *(Theorem 2.1 and Theorem 2.2 of [81]) Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a Nekrasov matrix. Then the diagonal matrix*

$$S_1 = \begin{pmatrix} \frac{h_1(A) + \varepsilon_1}{|a_{11}|} & & \\ & \ddots & \\ & & \frac{h_n(A) + \varepsilon_n}{|a_{nn}|} \end{pmatrix},$$

with

$$\begin{cases} \varepsilon_1 > 0, \\ 0 < \varepsilon_i \leq |a_{ii}| - h_i(A), \quad \varepsilon_i > \sum_{j=1}^{i-1} \frac{|a_{ij}|\varepsilon_j}{|a_{jj}|} \text{ for } i = 2, \dots, n, \end{cases} \quad (5.11)$$

is a positive diagonal matrix such that AS_1 is SDD.

Let $k \in N$ be the first index such that there does not exist $j > k$ with $a_{kj} \neq 0$. Then the diagonal matrix

$$S_2 = \begin{pmatrix} \frac{h_1(A) + \varepsilon_1}{|a_{11}|} & & \\ & \ddots & \\ & & \frac{h_n(A) + \varepsilon_n}{|a_{nn}|} \end{pmatrix},$$

with

$$\begin{cases} \varepsilon_i = 0, & \text{for } i = 1, \dots, k-1, \\ 0 < \varepsilon_i < |a_{ii}| - h_i(A), \quad \varepsilon_i > \sum_{j=k}^{i-1} \frac{|a_{ij}|\varepsilon_j}{|a_{jj}|}, & \text{for } i = k, \dots, n. \end{cases} \quad (5.12)$$

is a positive diagonal matrix such that AS_2 is SDD.

The particular case $k = n$ for the diagonal matrix S_2 corresponds to a diagonal matrix that was already introduced in [41] to derive error bounds for LCPs of Nekrasov matrices. We have used the matrices introduced in Theorem 5.5 to derive the following bounds for the infinity norm of the inverse of a Nekrasov matrix.

Theorem 5.6. (Theorem 3.2 of [81]) Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a Nekrasov matrix. Then

$$\|A^{-1}\|_{\infty} \leq \frac{\max_{i \in N} \left(\frac{h_i(A) + \varepsilon_i}{|a_{ii}|} \right)}{\min_{i \in N} (\varepsilon_i - w_i + p_i)}, \quad (5.13)$$

where $(\varepsilon_1, \dots, \varepsilon_n)$ are given either by (5.11) or (5.12) from Theorem 5.5, $w_i := \sum_{j=1}^{i-1} |a_{ij}| \frac{\varepsilon_j}{|a_{jj}|}$, and $p_i := \sum_{j=i+1}^n |a_{ij}| \frac{|a_{jj}| - h_j(A) - \varepsilon_j}{|a_{jj}|}$ for all $i \in N$.

Bounds from Theorem 5.6 are based on the bound given in Theorem 5.4 for SDD matrices combined with the use of an adequate diagonal matrix. If we use the same diagonal matrices but we take a better bound for SDD matrices, we can derive tighter bounds for Nekrasov matrices. In [79] we have illustrated this fact using as bound for SDD matrices the bound introduced in [61] for Nekrasov matrices, where it is shown that it gives better results when applied to an SDD matrix than the bound presented in Theorem 5.4.

5.3.2 Error bounds for the LCP of Nekrasov matrices

In Chapter 3 we have seen the strong relationship between P -matrices and the LCP. By Theorem 2.3 of [10], if M is a P -matrix, then the solution x^* of the LCP (3.3) satisfies

$$\|x - x^*\|_{\infty} \leq \max_{d \in [0,1]^n} \|M_D^{-1}\|_{\infty} \|r(x)\|_{\infty}, \quad (5.14)$$

where

$$M_D := I - D + DM, \quad (5.15)$$

I is the $n \times n$ identity matrix, D is the diagonal matrix $\text{diag}(d_i)$ with $0 \leq d_i \leq 1$, for all $i = 1, \dots, n$ and $r(x) := \min(x, Mx + q)$, where the min operator denotes the componentwise minimum of two vectors. Any H -matrix with positive diagonal entries is a P -matrix, so we will be able to apply this result to any Nekrasov matrix with positive diagonal entries. Moreover, if we know the right diagonal scaling matrix to transform an H -matrix into an SDD matrix, we can apply the following bound introduced in [40].

Theorem 5.7. (Theorem 2.1 of [40]) Suppose that $A = (a_{ij})_{1 \leq i, j \leq n}$ is an H -matrix with all its diagonal entries positive. Let $S = \text{diag}(s_i)_{i=1}^n$, $s_i > 0$ for all $i \in N$, be a diagonal matrix such that AS is SDD. For any $i = 1, \dots, n$, let $\beta_i := a_{ii}s_i - \sum_{j \neq i} |a_{ij}| s_j$. Then

$$\max_{d \in [0,1]^n} \|(I - D + DA)^{-1}\|_{\infty} \leq \max \left\{ \frac{\max_i \{s_i\}}{\min_i \{\beta_i\}}, \frac{\max_i \{s_i\}}{\min_i \{s_i\}} \right\}. \quad (5.16)$$

Using Theorem 5.7 and the scaling matrix S_2 introduced in Theorem 5.5, we obtained the following bound for Nekrasov matrices.

Theorem 5.8. (Theorem 5.2 of [81]) Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a Nekrasov matrix with all its diagonal entries positive. Let $S_2 = \text{diag}(s_i)_{i=1}^n$ and ε_i ($i \in N$) be the diagonal matrix and positive real numbers, respectively, defined in Theorem 5.5. Then

$$\max_{d \in [0,1]^n} \|(I - D + DA)^{-1}\|_{\infty} \leq \max \left\{ \frac{1}{\min_i \{\varepsilon_i - w_i + p_i\}}, \frac{1}{\min_i \{s_i\}} \right\}, \quad (5.17)$$

where, for each $i \in N$, p_i and w_i are defined in Theorem 5.6.

We also included numerical experiments that illustrate our new bounds and we compared them with some of the error bounds already known in the literature for Nekrasov matrices. Let us notice that the bound for the infinity norm of the inverse of an SDD matrix given by Theorem 5.4 can be arbitrarily large depending just on the behaviour of any of the matrix rows because of its dependency on the quantities $|a_{kk}| - \sum_{j \neq k} |a_{kj}|$ for all $k = 1, \dots, n$. For the bounds on Nekrasov matrices, this behaviour can be observed with respect to the differences $|a_{kk}| - h_k(A)$ for some sets of indices $k \in N$. For the bound given by Theorem 5.8, we can expect that the closeness to zero of the condition $|a_{kk}| - h_k(A)$ will not affect the quality of the bound for any indices $k \in N \setminus \{n\}$. And so, we can expect to obtain useful bounds for Nekrasov matrices that could be troublesome for the other known bounds.

5.4 B_{π}^R -matrices and B_{π}^R -tensors

The class of B_{π}^R -matrices was introduced in [78], giving a new class of matrices with positive determinant that contains the class of B -matrices.

Definition 5.9. Let $\pi = (\pi_1, \dots, \pi_n)^T$ be a vector such that

$$0 < \sum_{j=1}^n \pi_j \leq 1. \quad (5.18)$$

Let $M = (m_{ij})_{1 \leq i, j \leq n}$ be a real matrix with positive row sums and let $R = (R_1, \dots, R_n)^T$ be the vector formed by the row sums of M . Then we say that M is a B_{π}^R -matrix if for all $i = 1, \dots, n$,

$$\pi_j R_i > m_{ij}, \quad \forall j \neq i. \quad (5.19)$$

Moreover, this new class is a subset of P -matrices whenever $\pi \geq 0$ (Theorem 3.4 of [78]), so it gives a new easily checkable condition to identify some P -matrices. The definition of P -matrices is an extension of the definition of strictly positive definiteness: a symmetric matrix is positive definite if and only if it is a P -matrix. Precisely, the interest of the P -problem motivated the extension of the class of B -matrices to the multidimensional case (see [90]). After that, some related classes of matrices have also been defined for the multidimensional case, such as double B -tensors [63] and MB -tensors [64]. We considered an extension of B_π^R -matrices to the higher dimensional case and studied its properties and characterized the cases where these tensors are P -tensors.

5.4.1 B_π^R -tensors

Let us start by introducing some basic concepts about tensors that we can think of as an extension of the matrix definitions. A real m -th order n -dimensional tensor $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ is a multi-array of real entries $a_{i_1 \dots i_m} \in \mathbb{R}$, where $i_k \in N := \{1, \dots, n\}$ for $k = 1, \dots, m$. We say that \mathcal{A} is a *symmetric* tensor if its entries are invariant under any permutation of its indices. A tensor \mathcal{A} is called *diagonally dominant* if

$$|a_{i \dots i}| \geq \sum_{i_2, \dots, i_m \neq (i, \dots, i)}^n |a_{ii_2 \dots i_m}|, \quad i \in N. \quad (5.20)$$

If (5.20) holds strictly, then \mathcal{A} is called *strictly diagonally dominant*. We say that $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ is a B -tensor (B_0 -tensor) if

$$R_i(\mathcal{A}) > 0 \quad (\geq 0), \quad i \in N, \quad (5.21)$$

and

$$\frac{R_i(\mathcal{A})}{n^{m-1}} > a_{ij_2 \dots j_m} \quad (\geq a_{ij_2 \dots j_m}), \quad \forall (j_2, \dots, j_m) \neq (i, \dots, i). \quad (5.22)$$

We say that a tensor is *nonnegative* if all its entries are nonnegative, and that it is a Z -tensor if all its off-diagonal entries are nonpositive. Let us also define the *identity tensor* \mathcal{I} , whose entries are ones on the main diagonal (i.e., entries such that $i_1 = \dots = i_m$) and zeros elsewhere. A tensor $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ is called an (a *strong*) M -tensor if there exists a nonnegative tensor $\mathcal{B} = (b_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ and a positive scalar $s \geq \rho(\mathcal{B})$ ($> \rho(\mathcal{B})$) such that $\mathcal{A} = s\mathcal{I} - \mathcal{B}$, where $\rho(\mathcal{B})$ is the *spectral radius* of \mathcal{B} (see page 15 of [90]). (Strictly) diagonally dominant Z -tensors are also (strong) M -tensors (as it happened in the 2-dimensional case with SDD Z -matrices and nonsingular M -matrices).

A tensor \mathcal{A} is called positive semidefinite (definite) if for each (nonzero) $x \in \mathbb{R}^n$

$$\mathcal{A}x^m \geq 0 \quad (> 0),$$

where $\mathcal{A}x^m = \sum_{i_1, \dots, i_m=1}^n a_{i_1 i_2 \dots i_m} x_{i_1} \cdots x_{i_m}$. There are not any nontrivial positive semidefinite tensors when m is odd. Let us recall that, given an m -th order tensor $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ and $x \in \mathbb{R}^n$, then $\mathcal{A}x^{m-1} \in \mathbb{R}^n$ is given by

$$(\mathcal{A}x^{m-1})_i := \sum_{i_2, \dots, i_m=1}^n a_{ii_2 \dots i_m} x_{i_2} \cdots x_{i_m}, \quad \text{for each } i = 1, \dots, n.$$

Definition 5.10. (see [34] or page 192 of [90]) A tensor $\mathcal{A} \in \mathbb{R}^{[m,n]}$ is called a P -tensor if for each nonzero $x \in \mathbb{R}^n$ there exists an index $i \in N$ such that

$$x_i^{m-1}(\mathcal{A}x^{m-1})_i > 0. \quad (5.23)$$

A tensor $\mathcal{A} \in \mathbb{R}^{[m,n]}$ is called a P_0 -tensor if for each nonzero $x \in \mathbb{R}^n$ there exists some index $i \in N$ such that

$$x_i \neq 0 \text{ and } x_i^{m-1}(\mathcal{A}x^{m-1})_i \geq 0. \quad (5.24)$$

In [93] it was shown that in the even order case a symmetric tensor is positive definite (semidefinite) if and only if it is a P -tensor (P_0 -tensor). Finally, let us introduce the definition of B_π^R -tensor from [80].

Definition 5.11. Let $\pi = (\pi_1, \dots, \pi_n)$ be a nonnegative vector satisfying (5.18), let $i_1, \dots, i_m \in N$ and let $\pi_{i_1 i_2 \dots i_k} := \pi_{i_1} \pi_{i_2} \dots \pi_{i_k}$ with $k \leq m$. Given a tensor $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ and the vector $R = (R_i)_{i \in N}$ formed by its row sums, we say that \mathcal{A} is a B_π^R -tensor ($(B_\pi^R)_0$ -tensor) if R is positive (nonnegative) and, for all $k \in N$,

$$\pi_{i_2 \dots i_m} R_k > a_{k i_2 \dots i_m} (\geq a_{k i_2 \dots i_m}), \quad \text{with } \delta_{k i_2 \dots i_m} = 0. \quad (5.25)$$

When $\pi_j = \frac{1}{n}$ for $j \in N$ this definition of a B_π^R -tensor ($(B_\pi^R)_0$ -tensor) coincides with that of a B -tensor (B_0 -tensor).

B_π^R -tensors give a new subclass of P -tensors whenever they have odd order and of positive definite tensors whenever they are symmetric with even order. The key to prove these results lies on the use of the right decomposition of the tensor. Our first result gives a decomposition that relates B_π^R -tensors to SDD M -tensors.

Theorem 5.12. (Theorem 3.1 of [80]) Let $\mathcal{A} \in \mathbb{R}^{[m,n]}$ be a B_π^R -tensor. Then we can write \mathcal{A} as

$$\mathcal{A} = \mathcal{B} + \mathcal{C},$$

where \mathcal{B} is a strictly diagonally dominant M -tensor and \mathcal{C} is a nonnegative rank-one tensor.

Let $\Pi = \pi^m \in \mathbb{R}^{[m,n]}$ and let $J \subseteq N$. Then we denote by Π^J a tensor $\mathcal{A} = (a_{i_1 \dots i_m}) \in \mathbb{R}^{[m,n]}$ such that $a_{i_1 \dots i_m} = (\pi^m)_{i_1 \dots i_m} = \pi_{i_1} \dots \pi_{i_m}$ whenever $i_j \in J$ for all $j = 1, \dots, m$ and such that all its remaining entries are zero. The tensors Π^J play a key role in the new decomposition for symmetric B_π^R -tensors introduced in [80]. This decomposition was used to prove that a symmetric B_π^R -tensor of even order is positive definite.

Theorem 5.13. (Theorem 3.3 of [80]) Let $\mathcal{A} \in \mathbb{R}^{[m,n]}$ be a symmetric B_π^R -tensor. Then either \mathcal{A} is a strictly diagonally dominant symmetric Z -tensor or it can be written as

$$\mathcal{A} = \mathcal{M} + \sum_{i=1}^s h_i \Pi^{J_i}, \quad (5.26)$$

where \mathcal{M} is a strictly diagonally dominant Z -tensor, s is a positive integer, $h_k > 0$, $J_k \subseteq N$ for $k = 1, \dots, s$ and $J_s \subsetneq J_{s-1} \subsetneq \dots \subsetneq J_1$.

Based on the decompositions introduced in Theorem 5.12 and Theorem 5.13, we have the following theorem for B_π^R -tensors that presents the main results from [80].

Theorem 5.14. (Theorem 4.1 and Theorem 4.3 of [80]) *Let $\mathcal{A} \in \mathbb{R}^{[m,n]}$ be a B_π^R -tensor. Then we have that:*

- i) \mathcal{A} is a P -tensor whenever it has odd order.
- ii) if \mathcal{A} is a symmetric tensor of even order, then it is positive definite. In that case, \mathcal{A} is also a P -tensor.

Hence, our extension of B_π^R -matrices to the higher order case provides a condition based on the tensor entries to identify P -tensors and positive definite tensors in some new cases.

5.4.2 Error bounds for LCPs of B_π^R -matrices

As we recalled earlier, the class of B_π^R -matrices was introduced in [78] and proven to be a class with positive determinant and of P -matrices whenever $\pi \geq 0$. This condition is not required in the definition of a B_π^R -matrix (see Definition 5.9), but it was used when proving these properties. In fact, it is a necessary condition for the class to be P -matrices, but not for the weaker condition of being a class of matrices with positive determinant. We revised these properties in [83].

Theorem 5.15. (Theorem 1 of [83]) *If A is a B_π^R -matrix with $\pi \geq 0$, then A is a P -matrix.*

The condition $\pi \geq 0$ seems to be a general requirement to work with this class of matrices. In fact, since B_π^R -matrices are P -matrices whenever $\pi \geq 0$, they have been studied in the context of the LCP. Both [42] and [38] presented bounds for LCPs associated to B_π^R -matrices whenever $\pi > 0$. But, any B_π^R -matrix for any vector π satisfying (5.18) has positive determinant.

Theorem 5.16. (Theorem 2 of [83]) *Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a real matrix with positive row sums. If A is a B_π^R -matrix, then $\det A > 0$.*

In [83] we focused on the study of B_π^R -matrices with $\pi \geq 0$. For that, we started by presenting a characterization of B_π^R -matrices that gives a method for recognizing the class under the additional condition $\pi \geq 0$. Our characterization also gives a suitable vector π satisfying (5.18) that can be used to compute bounds for the norm of the inverses of these matrices as well as bounds for the error of the associated LCPs.

Proposition 5.17. (Proposition 1 of [83]) *Let A be a square matrix with positive row sums and let $R = (R_1, \dots, R_n)^T$ be the vector formed by the row sums of A . Then there exists a nonnegative vector π satisfying (5.18) such as A is a B_π^R -matrix if and only if*

$$\sum_{j=1}^n \max_{i \neq j} \left(\frac{a_{ij}}{R_i}, 0 \right) < 1. \quad (5.27)$$

Remark 5.18. The proof of Proposition 1 in [83] also gives a suitable vector π for a B_π^R -matrix $A = (a_{ij})_{1 \leq i, j \leq n}$. This vector is defined as $\pi := (\pi_1, \dots, \pi_n)$ with

$$\pi_j := \max_{i \neq j} \left(\frac{a_{ij}}{R_i}, 0 \right) + \frac{k}{n} \quad \text{for } j = 1, \dots, n, \quad (5.28)$$

where k is defined as

$$k := 1 - \sum_{j=1}^n \max_{i \neq j} \left(\frac{a_{ij}}{R_i}, 0 \right). \quad (5.29)$$

As we did with Nekrasov matrices in [81], we have considered two problems with B_π^R -matrices: finding bounds for the norm of their inverses and finding error bounds for the LCP. The first step for approaching both problems has been finding an appropriate decomposition for these matrices. The main difference with the decomposition introduced in [42] is that their decomposition, and also their error bounds for the LCP, depended on an additional parameter ε .

Proposition 5.19. (Proposition 2 of [83]) Let $A = (a_{ij})_{1 \leq i, j \leq n}$ be a B_π^R -matrix with $\pi_j > 0$ for all j and for each $i = 1, \dots, n$ let $\gamma_i := \max_{j \neq i} \{0, \frac{m_{ij}}{\pi_j}\}$. Then we can write $M = B^+ + C$, where $B^+ := (m_{ij} - \pi_j \gamma_i)_{1 \leq i, j \leq n}$ is a strictly diagonally dominant Z -matrix with positive diagonal entries and C is the rank one matrix given by $C := (\gamma_1, \dots, \gamma_n)^T (\pi_1, \dots, \pi_n)$.

Based on the decomposition given by Proposition 5.19, we obtained the following bound for the infinity norm of the inverse.

Theorem 5.20. (Theorem 3 of [83]) Let $M = (m_{ij})_{1 \leq i, j \leq n}$ be a B_π^R -matrix with $\pi_j > 0$ for all j and let R_j, γ_j be given as in Definition 5.9 and Proposition 5.19, respectively. Then

$$\|M^{-1}\|_\infty \leq \frac{\max_{1 \leq i \leq n} \left\{ \frac{1}{\pi_i} - 1 \right\}}{\min_{1 \leq i \leq n} \left\{ R_i - \gamma_i \sum_{j=1}^n \pi_j \right\}}. \quad (5.30)$$

Let us notice that we can apply Theorem 5.20 to any B_π^R -matrix satisfying (5.27) using the vector π given by (5.28). The bound presented in Theorem 5.20 is based on the bound of Theorem 5.4 for SDD matrices, as it was the case with the bound given by Theorem 5.6 for Nekrasov matrices. As we did in [81], we also considered using a different bound for SDD matrices that it is always as good as the one introduced in Theorem 5.4.

Theorem 5.21. (Theorem 4 of [83]) Let $M = (m_{ij})_{1 \leq i, j \leq n}$ be a B_π^R -matrix with $\pi_j > 0$ for all j and let R_j, γ_j be given as in Definition 5.9 and Proposition 5.19, respectively. Then

$$\|M^{-1}\|_\infty \leq \max_{1 \leq i \leq n} \left\{ \frac{1}{\pi_i} - 1 \right\} \max_{1 \leq i \leq n} \frac{z_i(B^+)}{m_{ii} - \gamma_i \pi_i - h_i(B^+)}, \quad (5.31)$$

where B^+ is given in Proposition 5.19, $h_i(B^+) = \sum_{j=1}^{i-1} \frac{\gamma_j \pi_j - m_{ij}}{m_{jj} - \gamma_j \pi_j} h_j(B^+) + \sum_{j=i+1}^n (\gamma_j \pi_j - m_{ij})$ and $z_i(B^+) = \sum_{j=1}^{i-1} \frac{\gamma_j \pi_j - m_{ij}}{m_{jj} - \gamma_j \pi_j} z_j(B^+) + 1$.

Let us recall that the error of the solution of a LCP (3.3) satisfies bound (5.14) (given by Theorem 2.3 of [10]) if the matrix defining the LCP is a P -matrix. Hence, the problem of looking for an error bound transforms into the problem of finding an upper bound for the infinity norm of the matrix M_D^{-1} given by (5.15).

Given M , a B_π^R -matrix for a vector $\pi = (\pi_1, \dots, \pi_n)$ with $\pi_i > 0$ for all $i = 1, \dots, n$, we can define $M_D = (\bar{m}_{ij})_{1 \leq i, j \leq n}$ by (5.15) for any diagonal matrix $D = \text{diag}(d_i)$ with $0 \leq d_i \leq 1$ for all $i = 1, \dots, n$. If B^+ and C are the matrices given by the decomposition of M given in Proposition 5.19, then we can define the corresponding matrices B_D^+, C_D by

$$C_D := DC, \quad B_D^+ := I - D + DB^+. \quad (5.32)$$

Hence, in [83] we presented the following upper bound for $\|M_D^{-1}\|_\infty$.

Theorem 5.22. (Theorem 6 of [83]) *Suppose that $M = (m_{ij})_{1 \leq i, j \leq n}$ is a B_π^R -matrix for a vector π with $\pi_i > 0$ for all $i = 1, \dots, n$ and let $M_D = (\bar{m}_{ij})_{1 \leq i, j \leq n}$, C_D and B_D^+ be the matrices given by (5.15) and (5.32). Then B_D^+ is a strictly diagonally dominant Z -matrix with positive diagonal entries and*

$$\max_{d \in [0, 1]^n} \|M_D^{-1}\|_\infty \leq \frac{\max_{1 \leq i \leq n} \left\{ \frac{1}{\pi_i} - 1 \right\}}{\min_{1 \leq i \leq n} \left\{ 1, R_i - \gamma_i \sum_{j=1}^n \pi_j \right\}}, \quad (5.33)$$

where, for each $i = 1, \dots, n$, R_i and γ_i are given by Definition 5.9 and Proposition 5.19, respectively.

Part IV

CONCLUSIONS AND FUTURE WORK

Chapter 6

Conclusions and future work

The main topic of the work presented in this dissertation has been the development of efficient and accurate methods to work with structured classes of matrices. The considered problems are very common in linear algebra, like solving linear systems of equations, computing eigenvalues, singular values, inverses or determinants. We considered special classes of matrices and showed how their structure can be exploited to develop methods that can perform much better than the general ones for those matrices. This comparison does not mean that the usual methods are not valid, but for structured matrices better methods can be developed. The usual methods in linear algebra are fundamental and widely used in many applications in industry, science, engineering... In fact, the ubiquitous presence of numerical linear algebra translates into many different challenges for researchers. For example, overcoming numerical errors, dealing with data dimensionality and looking for efficiency. In every case, it seems that one way to face the incoming challenges in this area should be exploiting all the information available while developing new methods so that they can be optimized to their particular framework. In this dissertation, we illustrate this fact computing to high relative accuracy with subclasses of P -matrices. For us, the key lies in the proper use of the structure of these matrices. For nonsingular totally positive matrices and nonsingular M -matrices, the use of the right parametrization means that we can develop methods that avoid subtractions. And as we have showed with the numerical experiments included in the articles, the error of the solutions computed this way stays very close to the order of the unit roundoff of the floating point arithmetic. In double precision, that is of the order of 10^{-16} . We have also considered more problems, either with these classes of matrices or with some closely related ones. Some of these problems are finding infinity norm bounds for the inverses, error bounds for the linear complementarity problem, the study of simple conditions that assure the positivity of the determinant or the optimal properties of the collocation matrices of tensor products of normalized B -bases. We also studied the generalization of some classes of matrices to the higher order case with the objective of finding new simple criteria to identify P -tensors and positive definite tensors. We now give a summary of the conclusions of the work presented in this dissertation.

It is known that nonsingular totally positive matrices can be characterized by the ex-

istence of a bidiagonal decomposition. The bidiagonal decomposition provides a natural parametrization for this class of matrices that has some interesting applications. For instance, if this factorization is known accurately, it serves as a parametrization to compute the eigenvalues, singular values, the inverse or the solution to some linear systems of equations to high relative accuracy. Hence, finding a method to compute accurately the bidiagonal decomposition of a nonsingular totally positive matrix gives a method to solve these problems to high relative accuracy. Moreover, finding that a class of matrices admits a bidiagonal decomposition under the right hypotheses gives a way to prove that the class is formed by nonsingular totally positive matrices. Our main contributions in this area have been the following:

- We have studied the collocation matrices of generalized Laguerre polynomials at decreasingly ordered negative nodes. We have proved that these matrices are totally positive, we have obtained their bidiagonal decomposition and we have showed that it can be computed to high relative accuracy. See Article 1, [17].
- We have studied the collocation matrices of Bessel polynomials and of reverse Bessel polynomials at increasingly ordered positive nodes. In both cases, we showed that these matrices are totally positive and that their bidiagonal decomposition can be computed to high relative accuracy. See Article 2, [16].
- We have also found the bidiagonal decomposition of multiple generalizations of the Pascal matrix arising in Combinatorics. We identified when these extensions are totally positive and when their bidiagonal decomposition can be computed to high relative accuracy. See Article 6, [18].
- We also considered some matrices based on the q -integers. Some of these matrices are q -analogues of some well-known examples of totally positive matrices. For instance, we obtained the bidiagonal decomposition of the q -analogue of the triangular Pascal matrices and of the symmetrical Pascal matrix and we showed that it can be computed accurately. We also derived the bidiagonal decomposition of matrices formed by the q -analogues of the Stirling numbers and we showed that they are totally positive. Finally, we considered an extension of the generalized Laguerre polynomials based on the q -integers. We showed that these matrices are totally positive under the same hypotheses introduced for the generalized Laguerre polynomials and we studied the cases where their bidiagonal decomposition can be computed to high relative accuracy. See Article 7, [20].

In each of these articles there are numerical experiments that illustrate the great accuracy achievable using the bidiagonal decomposition. However, for many families of totally positive matrices, finding a method to compute the bidiagonal decomposition to high relative accuracy is still an open question. In the presented work we have showcased some techniques that could give a way for finding this representation with the required accuracy for more matrices, and hence, give a method to solve many problems with them with high relative accuracy.

Some important examples of totally positive matrices come from Computer Aided Geometric Design, where normalized B -bases are fundamental because of their optimal shape preserving properties. It was recently shown that the collocation matrices of these bases also satisfy that their minimal eigenvalue and singular value is larger than the minimal eigenvalue or singular value, respectively, of any collocation matrix of any other normalized totally positive basis of the same space of functions at the same nodes. Moreover, the ∞ -norm condition number of these collocation matrices is a lower bound for the ∞ -norm condition number of the collocation matrices of any other normalized totally positive basis at the same nodes.

- We showed the optimal conditioning of the collocation matrices of the tensor product of normalized B -bases with respect to the collocation matrices of tensor products of all normalized totally positive bases of its spanned function space. Moreover, we proved that the minimal singular value and eigenvalue of these matrices are larger than the minimal singular value and eigenvalue, respectively, of any other collocation matrix of any tensor product of normalized totally positive bases of that space. See Article 8, [19].

Nonsingular M -matrices are another important subclass of P -matrices. It is known that the row sums and off-diagonal entries of a nonsingular diagonally dominant M -matrix serves as a parametrization that can be used to compute their inverse, determinant or singular values to high relative accuracy. We searched for more classes of nonsingular M -matrices that admitted a representation in terms of a parametrization that could be used to achieve accurate computations. Our main contributions in this area are the following:

- We found a parametrization for $n \times n$ Nekrasov Z -matrices with positive diagonal entries that can be used to compute the determinant and the inverse of these matrices to high relative accuracy with a computational cost of $\mathcal{O}(n^3)$ elementary operations. See Article 3, [79].
- We also found a parametrization for $n \times n$ B -matrices that can be used to compute their determinant to high relative accuracy and we showed that this can be achieved with a computational cost of $\mathcal{O}(n^3)$ elementary operations. Based on this method and the techniques that we studied for Nekrasov Z -matrices, we derived a parametrization for B -Nekrasov matrices and a method to compute the determinant of this class to high relative accuracy. In this case, the computational cost is also of the order of $\mathcal{O}(n^3)$ elementary operations. See Article 10, [82].

We can see that our parametrization for Nekrasov matrices resembles the one known for diagonally dominant M -matrices. In fact, these classes are connected by the existence of a diagonal scaling matrix that can be used to transform a Nekrasov matrix into a diagonally dominant (or a strictly diagonally dominant) matrix. We took advantage of this relationship to develop accurate methods for Nekrasov matrices based on the techniques known for diagonally dominant M -matrices. The existence of such a scaling matrix is a general property that characterizes H -matrices. In general, this property would let us derive a method to perform accurate computations with more subclasses of M -matrices if the right scaling/parametrization

is available. For example, the class of QN -matrices (Quasi-Nekrasov matrices) gives an extension of Nekrasov matrices and an adequate scaling matrix has been found under certain additional hypotheses, so this would give a good starting point for the search of more conditions that assure computations with high relative accuracy. The family of B -matrices and their extensions are also closely related to M -matrices. Based on this relationship, we managed to find a good parametrization that allowed us to compute the determinant of this family and of B -Nekrasov matrices to high relative accuracy. For these classes, our algorithms are based on a decomposition of the form $B = B^+ + C$, where C is a rank one matrix. From this decomposition, we based our method on the well-known lemma for the computation of the determinant of a matrix with a rank-one perturbation. However, we can not assure high relative accuracy for the computation of the inverse following the same approach. The natural way of computing the inverse from that decomposition would be using the Sherman-Morrison formula for the inverse. However, it implies subtractions, and hence, the high relative accuracy is not assured anymore. Therefore, the question of whether it is possible to compute accurate inverses for these classes from our suggested parametrizations (or from different ones) remains open.

One interesting question arising in the study of the linear complementarity problem would be developing good error bounds whenever the matrix defining it is a P -matrix, since in that case the existence and uniqueness of its solution is assured. For the particular case of subclasses of nonsingular M -matrices, the knowledge of adequate scaling matrices can be used to derive new error bounds for this problem.

- We introduced two different scaling matrices whose product with a Nekrasov matrix is a strictly diagonally dominant matrix. Based on these scaling matrices, we developed new bounds for the infinity norm of the inverses of Nekrasov matrices as well as new error bounds for linear complementarity problems whose associated matrices are Nekrasov. See Article 4, [81].
- We also considered another subclass of P -matrices that extends the class of B -matrices called B_π^R -matrices. For this class, we showed that a simple decomposition can be used to derive infinity norm bounds for their inverses as well as error bounds for the linear complementarity problem. We also revised some theoretical results on the class, we proved that they have positive determinant and that they are P -matrices whenever the vector π that defines the class is nonnegative. See Article 11, [83].

In many applications, the underlying data structure might encourage the use of tensors (hypermatrices) to capture special features. Some of the classical problems encountered working with matrices can be amplified when we consider a higher order case. For example, recognizing a general P -matrix is already a tasking problem and the cost increases when we consider the recognition of P -tensors. Hence, it is of interest to find easy criteria based on the tensor entries that can be used to recognize some subclasses of P -tensors and of positive definite tensors in polynomial time. In fact, this problem was one of the motivations for the extension of the class of B -matrices to B -tensors.

- We defined the class of B_π^R -tensors as a extension of B_π^R -matrices to the higher order case. Our main results for the class are that B_π^R -tensors of odd order are P -tensors and

that symmetric B_{π}^R -tensors of even order are P -tensors or, equivalently, positive definite. See Article 5, [80].

A related open problem is finding other classes of matrices with positive determinant that can be extended to the higher order case, providing new subclasses of P -tensors and positive definite tensors. However, this is not always possible. Some matrix definitions do not have a tensor counterpart that inherits the desired properties. There are important classes of matrices, not related to the class of B -matrices, that could have applications in this area. For instance, we are interested in the study of the extension of simple conditions related to the family of totally positive matrices. For example, the family of TP_2 matrices are characterized by the sign of their 2×2 minors, which are all nonnegative. We would like to consider the extension of conditions of this kind based on the use of hyperdeterminants and see if they could be used to find useful eigenvalue localization criteria for tensors.

Another example of well-known structured matrices is given by Toeplitz matrices, which are matrices with constant diagonals. The case of tridiagonal Toeplitz matrices is quite illustrative in our work because any totally positive tridiagonal Toeplitz matrix can be transformed into a nonsingular M -matrix just by changing the sign of its off-diagonal entries. For the family of tridiagonal Toeplitz matrices, we have studied and classified when they are examples of P -matrices. We also considered the bidiagonal decomposition of both M -matrices and totally positive matrices of this class, showed the condition that assures that it can be computed to high relative accuracy and used this decomposition to solve many problems to high relative accuracy. As another remarkable property, the inverse of a nonsingular tridiagonal Toeplitz M -matrix is a totally positive matrix, so we also derived the bidiagonal decomposition of those inverses. Finally, we showed that we can compute the bidiagonal decomposition of any sign skew-symmetric tridiagonal matrix with positive diagonal entries to high relative accuracy. These matrices are always P -matrices, and we showed that the bidiagonal decomposition can be used to compute their inverse and all their minors to HRA. See Article 9, [21].

The Kronecker product has proven to be an incredibly useful tool that can be used to build fast and practical numerical methods. We showcased its good properties when deriving optimal properties for the tensor product of normalized B -bases. The Kronecker product can be used for computing approximations for matrices and tensors whose size can be out of any manageable range, which is usually one of the main challenges found in applications that deal with a lot of data. The right use of approximation techniques should allow us to keep the fundamental structure and information from the data while providing a noticeable reduction on the order of the problem. For that purpose, the Kronecker product can be used to find approximations with good properties. One of the future tasks involves developing a method that allows us to compute approximations of the Fisher matrix associated to the probability distribution of a deep neural network. The size of this matrix is the square of the number of parameters of the net, which can be of the order of millions in recent networks. Hence, we would like to find techniques that let us approximate this matrix by the Kronecker product of smaller ones, in a way that the resulting matrix is positive semidefinite and can be used as regularization while training a network. The interest on this problem comes from my research

stay at FORWISS (University of Passau).

The Bernstein polynomials in one and multiple variables have been fundamental in the development of Computer Aided Geometric Design. Different extensions of these bases have been proposed by many authors, looking for good properties useful in the context of design and approximation. For example, the q -Bernstein polynomials have been introduced as an alternative based on the q -integers for the approximation of functions of one variable. Following our work studying totally positive matrices based on the q -integers, we started the study of an extension of q -Bernstein polynomials to a triangular domain. Our aim is introducing a corner cutting algorithm that can be used in the design of surfaces based on these polynomials with a shape parameter.

Chapter 7

Conclusiones y trabajo futuro

El tema principal del trabajo de investigación presentado en esta memoria ha sido el desarrollo de métodos numéricos eficientes y precisos para trabajar con clases de matrices estructuradas. Los problemas considerados son muy comunes en álgebra lineal, como resolver sistemas de ecuaciones lineales, calcular valores propios, valores singulares, inversas o determinantes. Hemos estudiado clases de matrices especiales y hemos mostrado cómo se puede aprovechar su estructura al desarrollar métodos numéricos para obtener resultados mucho mejores que los que se obtendrían utilizando los métodos generales. Esta comparación no quiere decir que los métodos comunes no sirvan, pero cuando se trabaja con este tipo de matrices su estructura especial permite desarrollar métodos con mejores propiedades. Las técnicas habituales en álgebra lineal numérica son fundamentales en muchas aplicaciones en industria, ciencia, ingeniería... De hecho, la presencia generalizada del álgebra lineal numérica se traduce en nuevos desafíos para los investigadores, como son controlar el efecto de los errores numéricos, lidiar con el problema de la dimensionalidad de los datos o la búsqueda de la eficiencia en los métodos desarrollados. En cualquier caso, una forma adecuada de afrontar estos desafíos debería ser aprovechar toda la información de la que se dispone para desarrollar métodos numéricos optimizados y adaptados a su contexto particular. En esta tesis, hemos desarrollado métodos que buscan aprovechar la estructura conocida de matrices especiales para lograr cálculos con alta precisión relativa. En todo caso, las matrices consideradas son subclases de P -matrices. Para nosotros, la clave ha sido utilizar una representación o parametrización distinta de estas matrices. Para las matrices totalmente positivas no singulares así como para las M -matrices diagonal dominantes, la utilización de una parametrización adecuada implica que podemos desarrollar métodos numéricos que evitan restas. Y, como se muestra en los experimentos numéricos presentados en los artículos, el error de las soluciones calculadas siguiendo esta estrategia permanece muy próximo al orden de la unidad de redondeo de la aritmética de punto flotante utilizada. En doble precisión, eso se traduce en un valor próximo a 10^{-16} . Además de obtener resultados con alta precisión relativa, también hemos estudiado otros problemas con estas matrices o con otras clases muy relacionadas. Algunos de los problemas considerados son la obtención de cotas para la norma infinito de la inversa, el desarrollo de cotas para el error del problema de complementariedad lineal, la búsqueda de condiciones sencillas que aseguren la positividad del determinante o el estudio de propiedades óptimas de las matrices

de colocación del producto tensorial de B -bases normalizadas. También hemos estudiado la extensión de clases de matrices al caso de mayor dimensión con el objetivo de desarrollar nuevas condiciones para identificar clases de P -tensores y de tensores definidos positivos. A continuación damos un resumen de las principales conclusiones obtenidas en el trabajo de investigación presentado en esta memoria.

Las matrices totalmente positivas no singulares se caracterizan por la existencia de una factorización bidiagonal. Esta factorización bidiagonal sirve como parametrización natural de la clase y puede ser de mucha utilidad para trabajar con ella. En particular, si esta representación es conocida de forma precisa, sirve como parametrización para el cálculo de valores propios, valores singulares, inversas así como para el cálculo de la solución de ciertos sistemas de ecuaciones lineales con alta precisión relativa. Por tanto, encontrar un método que permita obtener la factorización bidiagonal de una matriz totalmente positiva no singular con alta precisión relativa nos da una forma de resolver los problemas mencionados previamente también con alta precisión relativa. Además, la factorización bidiagonal también puede servir como herramienta para identificar nuevas clases de matrices totalmente positivas, puesto que la existencia de una descomposición de este estilo bajo las hipótesis adecuadas caracteriza a la clase. Nuestra aportación en este área ha sido la siguiente:

- Hemos estudiado las matrices de colocación de los polinomios de Laguerre generalizados en nodos negativos ordenados de forma decreciente. Hemos demostrado que estas matrices son totalmente positivas, hemos obtenido su factorización bidiagonal y hemos mostrado como se puede obtener con alta precisión relativa. Ver Article 1, [17].
- Se han estudiado las matrices de colocación de los polinomios de Bessel y de los polinomios de Bessel reversos en nodos positivos ordenados de forma creciente. En ambos casos, hemos visto que estas matrices son totalmente positivas y que su factorización bidiagonal puede calcularse con alta precisión relativa. Ver Article 2, [16].
- Hemos obtenido la factorización bidiagonal de diversas generalizaciones de la matriz de Pascal utilizadas en Combinatoria. Hemos identificado los casos en los que estas generalizaciones son matrices totalmente positivas y los casos en los que su factorización bidiagonal puede ser calculada con alta precisión relativa. Ver Article 6, [18].
- También hemos estudiado varias clases de matrices de q -enteros. Algunas de estas matrices son q -análogos de matrices totalmente positivas muy conocidas. En particular, hemos obtenido la factorización bidiagonal del q -análogo de la matriz triangular de Pascal así como de la matriz de Pascal simétrica y hemos mostrado que se puede calcular de forma precisa. También hemos mostrado que se puede calcular la factorización bidiagonal de matrices formadas por q -análogos de los números de Stirling con alta precisión relativa y que estas matrices son totalmente positivas. Finalmente, hemos estudiado una extensión de los polinomios de Laguerre generalizados basada en los q -enteros. Hemos demostrado que estas matrices son totalmente positivas bajo las mismas hipótesis utilizadas al estudiar los polinomios generalizados de Laguerre y hemos identificado los casos en los que su factorización bidiagonal puede ser calculada con alta precisión relativa. Ver Article 7, [20].

En cada uno de estos artículos se han incluido experimentos numéricos que ilustran la gran precisión que se puede lograr partiendo de la factorización bidiagonal. Sin embargo, para muchas clases de matrices totalmente positivas todavía no se ha encontrado un método para conseguir esta parametrización con alta precisión relativa. En esta memoria hemos mostrado técnicas que pueden servir para obtener esta representación de forma precisa al estudiar nuevas clases de matrices totalmente positivas, y, por tanto, de lograr un método para realizar cálculos con alta precisión relativa al trabajar con dichas clases.

Algunos de los ejemplos más importantes de matrices totalmente positivas vienen del campo del Diseño Geométrico Asistido por Ordenador, donde las B -bases normalizadas juegan un papel fundamental debido a sus óptimas propiedades de preservación de forma. Recientemente, se ha demostrado que las matrices de colocación de estas bases también cumplen que su valor propio y su valor singular más pequeños son una cota superior de los valores propios y valores singulares más pequeños, respectivamente, de las matrices de colocación (utilizando los mismos nodos) de cualquier otra base totalmente positiva normalizada del espacio de funciones que genera dicha B -base. Además, el número de condición en la norma infinito de estas matrices de colocación es una cota inferior del número de condición en norma infinito de todas las matrices de colocación (en los mismos nodos) de bases totalmente positivas normalizadas de ese espacio de funciones.

- Hemos mostrado el condicionamiento óptimo de las matrices de colocación del producto tensorial de B -bases normalizadas con respecto a las matrices de colocación de productos tensoriales de cualquier base totalmente positiva normalizada de su espacio de funciones generado. Además, hemos demostrado que el valor singular y el valor propio más pequeños de estas matrices son mayores que los valores singulares y valores propios más pequeños, respectivamente, de cualquier matriz de colocación de un producto tensorial de bases totalmente positivas normalizadas de dicho espacio de funciones. Ver Article 8, [19].

Las M -matrices no singulares forman otra subclase muy importante de las P -matrices. Es conocido que las sumas de filas y las entradas extradiagonales de una M -matriz no singular diagonal dominante sirve como parametrización para calcular su inversa, determinante y valores singulares con alta precisión relativa. Hemos buscado nuevas clases de M -matrices que permitan una representación en términos de una parametrización que se pueda utilizar para lograr cálculos con alta precisión relativa. Nuestros resultados en este área han sido los siguientes:

- Hemos obtenido una parametrización para Z -matrices de Nekrasov $n \times n$ con entradas diagonales positivas que se puede utilizar para calcular el determinante y la inversa de estas matrices con alta precisión relativa con un coste computacional de $\mathcal{O}(n^3)$ operaciones elementales. Ver Article 3, [79].
- También hemos obtenido una parametrización para las B -matrices $n \times n$ que se puede utilizar para calcular su determinante con alta precisión relativa y hemos demostrado que se puede lograr con un método con un coste computacional de $\mathcal{O}(n^3)$ operaciones

elementales. Basándonos en este método y en las técnicas que hemos estudiado para las Z -matrices de Nekrasov, hemos obtenido una parametrización adecuada para las matrices B -Nekrasov y un método que, partiendo de dicha parametrización, permite calcular su determinante con alta precisión relativa con un coste computacional también del orden de $\mathcal{O}(n^3)$ operaciones elementales. Ver Article 10, [82].

Podemos ver que nuestra parametrización se asemeja a la parametrización utilizada con las M -matrices diagonal dominantes. De hecho, estas clases están conectadas por la existencia de una matriz diagonal de escalado que puede utilizarse para transformar una matriz de Nekrasov en una matriz diagonal dominante (o estrictamente diagonal dominante). Nos hemos aprovechado de esta relación para lograr métodos que aseguren la alta precisión relativa al trabajar con matrices de Nekrasov basándonos en las técnicas conocidas para M -matrices diagonal dominantes. La existencia de una matriz de escalado de esta forma es una propiedad que caracteriza las H -matrices. En general, esta propiedad nos debería permitir desarrollar métodos precisos para más subclases de M -matrices no singulares si conocemos un escalado o parametrización adecuados. Por ejemplo, la clase formada por las QN -matrices (matrices cuasi Nekrasov) proporciona una generalización de las matrices Nekrasov para la que se ha encontrado una matriz de escalado adecuada bajo ciertas hipótesis adicionales, por lo que serviría como un buen punto de partida en la búsqueda de más condiciones que permitan lograr métodos con alta precisión relativa. La familia de B -matrices y sus generalizaciones están muy relacionadas con las M -matrices. Apoyándonos en esa relación, obtuvimos una parametrización para las B -matrices y otra para las matrices B -Nekrasov que nos permite calcular sus determinantes con alta precisión relativa. Para estas clases, nuestros algoritmos parten de una descomposición de la forma $B = B^+ + C$, donde C es una matriz de rango uno. Partiendo de dicha descomposición, desarrollamos métodos numéricos muy precisos basándonos en la conocida fórmula para el cálculo del determinante de una matriz a la que se ha aplicado una perturbación de rango uno. Sin embargo, no podemos asegurar la alta precisión relativa para el cálculo de la inversa siguiendo una estrategia similar. La forma natural de calcular la inversa partiendo de dicha descomposición sería utilizar la fórmula de Sherman-Morrison para la inversa. Sin embargo, la utilización de esta fórmula acarrea restas, por lo que ya no se podría asegurar que el cálculo se llevara a cabo con alta precisión relativa. Por tanto, la cuestión de si es posible calcular la inversa de las matrices pertenecientes a estas clases con alta precisión relativa permanece abierto.

En el estudio del problema de complementariedad lineal, un problema que atrae mucha atención es el desarrollo de buenas cotas para el error cuando la matriz asociada es una P -matriz, puesto que en este caso la existencia y unicidad de la solución está asegurada. Para el caso particular de subclases de M -matrices no singulares, el conocimiento de una matriz de escalado adecuada puede ser utilizada para obtener nuevas cotas de error para este problema.

- Hemos introducido dos matrices de escalado diferentes cuyo producto con una matriz de Nekrasov es una matriz estrictamente diagonal dominante. Basándonos en estas matrices de escalado, hemos desarrollado nuevas cotas para la norma infinito de la inversa de una matriz Nekrasov así como nuevas cotas del error del problema de complementariedad lineal cuando su matriz asociada es una matriz Nekrasov. Ver Article 4, [81].

- También hemos trabajado con otra subclase de las P -matrices, que extiende la clase de las B -matrices, llamada B_{π}^R -matrices. Para esta clase, hemos conseguido una sencilla descomposición que puede ser usada para obtener cotas de la norma infinito de sus inversas así como cotas para el error del problema de complementariedad lineal. Además revisamos varios resultados teóricos sobre la clase. Demostramos que tienen determinante positivo y que son P -matrices siempre que el vector π que define la clase sea no negativo. Ver Article 11, [83].

En muchas aplicaciones, la utilización de tensores (hipermatrices) permite capturar propiedades estructurales importantes de los datos con los que se trabaja. Pero, al considerar este caso de mayor orden, algunos de los problemas comunes encontrados al trabajar con matrices se ven amplificados. Por ejemplo, el problema de reconocer un P -tensor general es aún más complejo que el problema de reconocer una P -matriz. Por tanto, es interesante desarrollar criterios sencillos basados en las entradas del tensor que puedan ser usados para reconocer más clases de P -tensores y de tensores definidos positivos con un coste computacional de orden polinómico. De hecho, este problema fue una de las principales razones que motivó la extensión de la definición de la clase de las B -matrices a B -tensores.

- Hemos definido B_{π}^R -tensor, dando una extensión de la clase formada por las B_{π}^R -matrices al caso tensorial. Nuestros resultados principales para esta nueva clase de tensores son que los B_{π}^R -tensores de orden impar son P -tensores y que los B_{π}^R -tensores simétricos de orden par son P -tensores y, de forma equivalente, definidos positivos. Ver Article 5, [80].

Un problema abierto relacionado sería buscar más clases de matrices con determinante positivo que puedan ser extendidas al caso tensorial, de forma que proporcionen nuevas clases de P -tensores y/o de tensores definidos positivos. Sin embargo, esto no es siempre posible. Algunas definiciones matriciales no tienen un análogo para tensores que herede las propiedades deseadas. Por otro lado, hay clases muy importantes de matrices, no relacionadas con las B -matrices, que podrían tener aplicaciones potenciales en este campo. En particular, estamos interesados en el estudio de la extensión de condiciones sencillas relacionadas con las matrices totalmente positivas. Por ejemplo, las matrices TP_2 se caracterizan por el signo de sus menores 2×2 , que son todos no negativos. Nos gustaría considerar la extensión de condiciones de este estilo basadas en el uso de hiperdeterminantes para comprobar si se podrían aplicar en el desarrollo de criterios útiles de localización de valores propios para tensores.

Otro ejemplo de matriz estructurada muy conocido es el de las matrices de Toeplitz, caracterizadas porque en sus diagonales se repite siempre el mismo elemento. El caso de las matrices tridiagonales de Toeplitz es muy ilustrativo en este trabajo porque cualquier matriz tridiagonal de Toeplitz que sea totalmente positiva puede ser transformada en una M -matriz cambiando simplemente el signo de sus entradas extradiagonales. Hemos estudiado y clasificado los casos en los que las matrices de Toeplitz tridiagonales son ejemplos de P -matrices. También hemos estudiado la factorización bidiagonal tanto de las M -matrices como de las matrices totalmente positivas de esta clase, hemos hallado la condición que nos permite asegurar que esta parametrización se puede conseguir con alta precisión relativa y la hemos utilizado para resolver diversos problemas con alta precisión relativa. Otra propiedad notoria de

esta clase es que la inversa de una M -matriz de Toeplitz tridiagonal no singular es una matriz totalmente positiva, por lo que también hemos obtenido la factorización bidiagonal de esas inversas. Finalmente, hemos demostrado que se puede obtener la factorización bidiagonal de cualquier matriz tridiagonal con estructura de signos antisimétrica y entradas diagonales positivas con alta precisión relativa. Estas matrices son siempre P -matrices, y hemos mostrado cómo se puede utilizar su factorización bidiagonal para calcular sus inversas y todos sus menores con alta precisión relativa. Ver Article 9, [21].

El producto de Kronecker proporciona una herramienta muy útil para el desarrollo de métodos numéricos rápidos y prácticos. En esta memoria hemos mostrado alguna de las muy buenas propiedades del producto de Kronecker utilizándolo para demostrar propiedades óptimas del producto tensorial de B -bases normalizadas. El producto de Kronecker se puede utilizar para calcular aproximaciones manejables de matrices y tensores cuyo tamaño original excede el límite manejable en nuestro contexto, que suele ser uno de las dificultades más comunes encontradas en aplicaciones que manejan grandes cantidades de datos. El uso adecuado de técnicas de aproximación nos debería permitir mantener la estructura e información fundamentales de nuestros datos proporcionándonos a la vez una reducción notable en el orden del problema. Con este fin, el producto de Kronecker es una herramienta con muchos usos potenciales. Una de las tareas futuras consistirá en el desarrollo de un método que nos permita calcular aproximaciones de la matriz de información de Fisher asociada a la distribución de probabilidad de una red neuronal. El tamaño de esta matriz es el cuadrado del número de parámetros empleados en la red, lo que puede ser del orden de millones en aplicaciones actuales. Por tanto, nos gustaría desarrollar técnicas que nos permitan aproximar esta matriz por el producto de Kronecker de otras más pequeñas, de modo que la aproximación resultante sea semidefinida positiva y pueda ser utilizada como regularización al reentrenar una red neuronal. El interés en el estudio de este problema surgió durante mi estancia en el instituto FORWISS, en la Universidad de Passau.

Los polinomios de Bernstein en una y varias variables han sido fundamentales en el desarrollo del Diseño Geométrico asistido por Ordenador. Muchos autores han estudiado distintas extensiones de estos polinomios buscando nuevas bases con buenas propiedades en el contexto del diseño y la aproximación. Por ejemplo, los polinomios de q -Bernstein han proporcionado una alternativa basada en los q -enteros para la aproximación de funciones de una variable. Continuando nuestro trabajo sobre matrices totalmente positivas de q -enteros, hemos comenzado el estudio de una extensión de los polinomios de q -Bernstein al caso de dominio triangular. Nuestro objetivo es desarrollar un algoritmo de corte de esquinas que permita diseñar superficies basadas en estos polinomios con un parámetro de forma.

Appendix

Journal impact factor of the presented articles

- [17] J. Delgado, H. Orera and J. M. Peña. Accurate computations with Laguerre matrices. Numer. Linear Algebra Appl. 26 (2019), e2217, 10 pp.

The JCR journal impact factor of Numerical Linear Algebra with Applications in 2019 is 1,298 (Q1 in the category “Mathematics”).

- [16] J. Delgado, H. Orera and J. M. Peña. Accurate algorithms for Bessel matrices. J. Sci. Comput. 80 (2019), 1264-1278.

The JCR journal impact factor of Journal of Scientific Computing in 2019 is 2,228 (Q1 in the category “Mathematics, Applied”).

- [79] H. Orera and J. M. Peña. Accurate inverses of Nekrasov Z-matrices. Linear Algebra Appl. 574 (2019), 46-59.

The JCR journal impact factor of Linear Algebra and its Applications in 2019 is 0,988 (Q2 in the category “Mathematics”).

- [81] H. Orera and J. M. Peña. Infinity norm bounds for the inverse of Nekrasov matrices using scaling matrices. Appl. Math. Comput. 358 (2019), 119-127.

The JCR journal impact factor of Applied Mathematics and Computation in 2019 is 3,472 (Q1 in the category “Mathematics, Applied”).

- [80] H. Orera and J. M. Peña. B_{π}^R -tensors. Linear Algebra Appl. 581 (2019), 247-259.

The JCR journal impact factor of Linear Algebra and its Applications in 2019 is 0,988 (Q2 in the category “Mathematics”).

- [18] J. Delgado, H. Orera and J. M. Peña. Accurate bidiagonal decomposition and computations with generalized Pascal matrices. J. Comput. Appl. Math. 391 (2021), Paper No. 113443, 10 pp.

The JCR journal impact factor of Journal of Computational and Applied Mathematics in 2020 is 2,621 (Q1 in the category “Mathematics, Applied”).

- [20] J. Delgado, H. Orera and J. M. Peña. High relative accuracy with matrices of q-integers. Numer. Linear Algebra Appl. 28 (2021), Paper No. e2383, 20 pp.

The JCR journal impact factor of Numerical Linear Algebra with Applications in 2020 is 2,109 (Q1 in the category “Mathematics”).

- [19] J. Delgado, H. Orera and J. M. Peña. Optimal properties of tensor product of B-bases. *Appl. Math. Lett.* 121 (2021), Paper No. 107473, 5 pp.

The JCR journal impact factor of Applied Mathematics Letters in 2020 is 4,055 (Q1 in the category “Mathematics, Applied”).

- [21] J. Delgado, H. Orera and J. M. Peña. Characterizations and accurate computations for tridiagonal Toeplitz matrices, *Linear and Multilinear Algebra* (2021), Published online, DOI: 10.1080/03081087.2021.1884180.

The JCR journal impact factor of Linear and Multilinear Algebra in 2020 is 1,736 (Q1 in the category “Mathematics”).

- [82] H. Orera and J. M. Peña. Accurate determinants of some classes of matrices. *Linear Algebra Appl.* 630 (2021), 1-14.

The JCR journal impact factor of Linear Algebra and its Applications in 2020 is 1,401 (Q2 in the category “Mathematics, Applied”).

- [83] H. Orera and J. M. Peña. Error bounds for linear complementarity problems of B_{π}^R -matrices. *Comput. Appl. Math.* 40 (2021), Paper No. 94, 13 pp.

The JCR journal impact factor of Computational and Applied Mathematics in 2020 is 2,239 (Q1 in the category “Mathematics, Applied”).

Co-Authorship justification

All the articles presented reflect the research work carried out by the author of this dissertation during his time as a predoctoral researcher. They have been prepared with the collaboration of his research supervisors, Prof. Juan Manuel Peña Ferrández and Prof. Jorge Delgado Gracia. This author’s contribution to the publications is embodied in the following tasks:

- Conceptualization.
- Methodology.
- Investigation.
- Writing mathematical proofs.
- Design and implementation of the numerical algorithms.
- Numerical experiments.
- Analysis and discussion of the results.
- Manuscript writing.

References

- [1] A. S. Alfa, J. Xue, and Q. Ye. Entrywise perturbation theory for diagonally dominant M-matrices with applications. *Numer. Math.*, 90(3):401–414, 2002.
- [2] P. Alonso, J. Delgado, R. Gallego, and J. M. Peña. Conditioning and accurate computations with Pascal matrices. *J. Comput. Appl. Math.*, 252:21–26, 2013.
- [3] T. Ando. Totally positive matrices. *Linear Algebra Appl.*, 90:165–219, 1987.
- [4] G. E. Andrews, E. S. Egge, W. Gawronski, and L. L. Littlejohn. The Jacobi-Stirling numbers. *Journal of Combinatorial Theory, Series A*, 120(1):288–303, 2013.
- [5] M. Bayat and H. Teimoori. The linear algebra of the generalized Pascal functional matrix. *Linear Algebra Appl.*, 295(1-3):81–89, 1999.
- [6] A. Berman and R. J. Plemmons. *Nonnegative matrices in the mathematical sciences*. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1979.
- [7] F. Brenti. Combinatorics and total positivity. *J. Combin. Theory Ser. A*, 71(2):175–218, 1995.
- [8] R. Bru, I. Giménez, and A. Hadjidimos. Is $A \in \mathbb{C}^{n,n}$ a general H -matrix? *Linear Algebra Appl.*, 436(2):364–380, 2012.
- [9] 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(6):633–654, 1994.
- [10] X. Chen and S. Xiang. Computation of error bounds for P-matrix linear complementarity problems. *Math. Program.*, 106(3, Ser. A):513–525, 2006.
- [11] R. W. Cottle, J.-S. Pang, and R. E. Stone. *The linear complementarity problem*. Computer Science and Scientific Computing. Academic Press, Inc., Boston, MA, 1992.
- [12] G. E. Coxson. The P-matrix problem is co-NP-complete. *Math. Programming*, 64(2, Ser. A):173–178, 1994.
- [13] L. Cvetković, P.-F. Dai, K. Doroslovački, and Y.-T. Li. Infinity norm bounds for the inverse of Nekrasov matrices. *Appl. Math. Comput.*, 219(10):5020–5024, 2013.

- [14] L. Cvetković, V. Kostić, and M. Nedović. Generalizations of Nekrasov matrices and applications. *Open Math.*, 13(5):96–105, 2015.
- [15] L. Cvetković, T. Szulc, and M. Nedović. Scaling technique for partition-Nekrasov matrices. *Appl. Math. Comput.*, 271:201–208, 2015.
- [16] J. Delgado, H. Orera, and J. M. Peña. Accurate algorithms for Bessel matrices. *J. Sci. Comput.*, 80(2):1264–1278, 2019.
- [17] J. Delgado, H. Orera, and J. M. Peña. Accurate computations with Laguerre matrices. *Numer. Linear Algebra Appl.*, 26(1):e2217, 10, 2019.
- [18] J. Delgado, H. Orera, and J. M. Peña. Accurate bidiagonal decomposition and computations with generalized Pascal matrices. *J. Comput. Appl. Math.*, 391:Paper No. 113443, 10, 2021.
- [19] J. Delgado, H. Orera, and J. M. Peña. Optimal properties of tensor product of B-bases. *Appl. Math. Lett.*, 121:Paper No. 107473, 5, 2021.
- [20] J. Delgado, H. Orera, and J. M. Peña. High relative accuracy with matrices of q -integers. *Numer. Linear Algebra Appl.*, 28(5):e2383, 2021.
- [21] J. Delgado, H. Orera, and J. M. Peña. Characterizations and accurate computations for tridiagonal Toeplitz matrices. *Linear Multilinear Algebra*, pages 1–20, 2021, Published online, DOI: 10.1080/03081087.2021.1884180.
- [22] J. Delgado, G. Peña, and J. M. Peña. Accurate and fast computations with positive extended Schoenmakers-Coffey matrices. *Numer. Linear Algebra Appl.*, 23(6):1023–1031, 2016.
- [23] J. Delgado and J. M. Peña. A shape preserving representation with an evaluation algorithm of linear complexity. *Comput. Aided Geom. Design*, 20(1):1–10, 2003.
- [24] J. Delgado and J. M. Peña. Accurate computations with collocation matrices of rational bases. *Appl. Math. Comput.*, 219(9):4354–4364, 2013.
- [25] J. Delgado and J. M. Peña. Fast and accurate algorithms for Jacobi-Stirling matrices. *Appl. Math. Comput.*, 236:253–259, 2014.
- [26] J. Delgado and J. M. Peña. Accurate computations with collocation matrices of q -Bernstein polynomials. *SIAM J. Matrix Anal. Appl.*, 36(2):880–893, 2015.
- [27] J. Delgado and J. M. Peña. Accurate computations with Lupaş matrices. *Appl. Math. Comput.*, 303:171–177, 2017.
- [28] J. Delgado and J. M. Peña. Extremal and optimal properties of B-bases collocation matrices. *Numer. Math.*, 146(1):105–118, 2020.
- [29] J. Demmel. *Applied Numerical Linear Algebra*. SIAM, 1997.

- [30] J. Demmel, I. Dumitriu, O. Holtz, and P. Koev. Accurate and efficient expression evaluation and linear algebra. *Acta Numer.*, 17:87–145, 2008.
- [31] J. Demmel, M. Gu, S. Eisenstat, I. Slapničar, K. Veselić, and Z. Drmač. Computing the singular value decomposition with high relative accuracy. *Linear Algebra Appl.*, 299:21–80, 1999.
- [32] J. Demmel and P. Koev. Accurate SVDs of weakly diagonally dominant M-matrices. *Numer. Math.*, 98:99–104, 2004.
- [33] J. Demmel and P. Koev. The accurate and efficient solution of a totally positive generalized Vandermonde linear system. *SIAM J. Matrix Anal. Appl.*, 27(1):142–152, 2005.
- [34] W. Ding, Z. Luo, and L. Qi. P-tensors, P_0 -tensors, and their applications. *Linear Algebra Appl.*, 555:336–354, 2018.
- [35] T. Ernst. q -Stirling numbers, an umbral approach. *Adv. Dyn. Syst. Appl.*, 3(2):251–282, 2008.
- [36] S. M. Fallat and C. R. Johnson. *Totally nonnegative matrices*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2011.
- [37] F. P. Gantmacher and M. G. Krein. *Oscillation matrices and kernels and small vibrations of mechanical systems*. AMS Chelsea Publishing, Providence, RI, revised edition, 2002. Translation based on the 1941 Russian original.
- [38] L. Gao, C. Li, and Y. Li. Parameterized error bounds for linear complementarity problems of B_π^R -matrices and their optimal values. *Calcolo*, 56(3):Paper No. 31, 24, 2019.
- [39] M. García-Esnaola and J. M. Peña. Error bounds for linear complementarity problems for B -matrices. *Appl. Math. Lett.*, 22(7):1071–1075, 2009.
- [40] M. García-Esnaola and J. M. Peña. A comparison of error bounds for linear complementarity problems of H -matrices. *Linear Algebra Appl.*, 433(5):956–964, 2010.
- [41] M. García-Esnaola and J. M. Peña. Error bounds for linear complementarity problems of Nekrasov matrices. *Numer. Algorithms*, 67(3):655–667, 2014.
- [42] M. García-Esnaola and J. M. Peña. B_π^R -matrices and error bounds for linear complementarity problems. *Calcolo*, 54(3):813–822, 2017.
- [43] M. García-Esnaola and J. M. Peña. B-nekrasov matrices and error bounds for linear complementarity problems. *Numer. Algorithms*, 72(2):435–445, 2016.
- [44] M. Gasca and C. A. Micchelli, editors. *Total positivity and its applications*, volume 359 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

- [45] M. Gasca and J. M. Peña. Total positivity and Neville elimination. *Linear Algebra Appl.*, 165:25–44, 1992.
- [46] M. Gasca and J. M. Peña. A matricial description of Neville elimination with applications to total positivity. *Linear Algebra Appl.*, 202:33–53, 1994.
- [47] M. Gasca and J. M. Peña. On factorizations of totally positive matrices. In *Total positivity and its applications (Jaca, 1994)*, volume 359 of *Math. Appl.*, pages 109–130. Kluwer Acad. Publ., Dordrecht, 1996.
- [48] E. Grosswald. *Bessel polynomials*, volume 698 of *Lecture Notes in Mathematics*. Springer, Berlin, 1978.
- [49] H. Han and S. Seo. Combinatorial proofs of inverse relations and log-concavity for Bessel numbers. *European J. Combin.*, 29(7):1544–1554, 2008.
- [50] N. Higham. *Accuracy and Stability of Numerical Algorithms*. SIAM, 2002.
- [51] A. J. Hoffman. On the nonsingularity of real matrices. *Math. Comp.*, 19:56–61, 1965.
- [52] M. E. H. Ismail. *Classical and quantum orthogonal polynomials in one variable*, volume 98 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2005.
- [53] C. R. Johnson and R. A. Horn. *Topics in Matrix Analysis*. Cambridge University Press, 1991.
- [54] V. Kac and P. Cheung. *Quantum calculus*. Universitext. Springer-Verlag, New York, 2002.
- [55] S. Karlin. *Total positivity. Vol. I*. Stanford University Press, Stanford, Calif., 1968.
- [56] O. M. Katkova and A. M. Vishnyakova. On sufficient conditions for the total positivity and for the multiple positivity of matrices. *Linear Algebra Appl.*, 416(2-3):1083–1097, 2006.
- [57] I.-P. Kim. LDU decomposition of an extension matrix of the Pascal matrix. *Linear Algebra Appl.*, 434(10):2187–2196, 2011.
- [58] P. Koev. TNTool library. <http://www.math.sjsu.edu/~koev/software/TNTool.html>. [Online; accessed 01-May-2022].
- [59] P. Koev. Accurate eigenvalues and SVDs of totally nonnegative matrices. *SIAM J. Matrix Anal. Appl.*, 27(1):1–23, 2005.
- [60] P. Koev. Accurate computations with totally nonnegative matrices. *SIAM J. Matrix Anal. Appl.*, 29(3):731–751, 2007.
- [61] L. Y. Kolotilina. On bounding inverses to Nekrasov matrices in the infinity norm. *Journal of Mathematical Sciences*, 199(4):432–437, 2014.

- [62] T. H. Koornwinder, R. Wong, R. Koekoek, and R. F. Swarttouw. Orthogonal polynomials. In *NIST handbook of mathematical functions*, pages 435–484. U.S. Dept. Commerce, Washington, DC, 2010.
- [63] C. Li and Y. Li. Double B -tensors and quasi-double B -tensors. *Linear Algebra Appl.*, 466:343–356, 2015.
- [64] C. Li, L. Qi, and Y. Li. MB -tensors and MB_0 -tensors. *Linear Algebra Appl.*, 484:141–153, 2015.
- [65] H.-B. Li, T.-Z. Huang, and H. Li. On some subclasses of P -matrices. *Numer. Linear Algebra Appl.*, 14(5):391–405, 2007.
- [66] X.-G. Lv, T.-Z. Huang, and Z.-G. Ren. A new algorithm for linear systems of the Pascal type. *J. Comput. Appl. Math.*, 225(1):309–315, 2009.
- [67] E. Mainar and J. M. Peña. Accurate computations with collocation matrices of a general class of bases. *Numer. Linear Algebra Appl.*, 25(5):e2184, 12, 2018.
- [68] E. Mainar, J. M. Peña, and B. Rubio. Accurate bidiagonal decomposition of collocation matrices of weighted φ -transformed systems. *Numer. Linear Algebra Appl.*, 27(3):e2295, 16, 2020.
- [69] E. Mainar, J. M. Peña, and B. Rubio. Accurate computations with collocation and Wronskian matrices of Jacobi polynomials. *J. Sci. Comput.*, 87(3):Paper No. 77, 30, 2021.
- [70] E. Mainar, J. M. Peña, and B. Rubio. Accurate computations with Wronskian matrices. *Calcolo*, 58(1):Paper No. 1, 15, 2021.
- [71] A. Marco and J.-J. Martínez. A fast and accurate algorithm for solving Bernstein-Vandermonde linear systems. *Linear Algebra Appl.*, 422(2-3):616–628, 2007.
- [72] A. Marco and J.-J. Martínez. Accurate computations with totally positive Bernstein-Vandermonde matrices. *Electron. J. Linear Algebra*, 26:357–380, 2013.
- [73] A. Marco and J.-J. Martínez. A total positivity property of the Marchenko-Pastur law. *Electron. J. Linear Algebra*, 30:106–117, 2015.
- [74] A. Marco and J.-J. Martínez. Bidiagonal decomposition of rectangular totally positive Said-Ball-Vandermonde matrices: error analysis, perturbation theory and applications. *Linear Algebra Appl.*, 495:90–107, 2016.
- [75] A. Marco, J.-J. Martínez, and J. M. Peña. Accurate bidiagonal decomposition of totally positive Cauchy-Vandermonde matrices and applications. *Linear Algebra Appl.*, 517:63–84, 2017.

- [76] A. Marco, J.-J. Martínez, and R. Viaña. Accurate bidiagonal decomposition of totally positive h-Bernstein-Vandermonde matrices and applications. *Linear Algebra Appl.*, 579:320–335, 2019.
- [77] A. Marco and J.-J. Martínez. Accurate computation of the moore–penrose inverse of strictly totally positive matrices. *Journal of Computational and Applied Mathematics*, 350:299–308, 2019.
- [78] M. Neumann, J. M. Peña, and O. Pryporova. Some classes of nonsingular matrices and applications. *Linear Algebra Appl.*, 438(4):1936–1945, 2013.
- [79] H. Orera and J. M. Peña. Accurate inverses of Nekrasov Z -matrices. *Linear Algebra Appl.*, 574:46–59, 2019.
- [80] H. Orera and J. M. Peña. B_{π}^R -tensors. *Linear Algebra Appl.*, 581:247–259, 2019.
- [81] H. Orera and J. M. Peña. Infinity norm bounds for the inverse of Nekrasov matrices using scaling matrices. *Appl. Math. Comput.*, 358:119–127, 2019.
- [82] H. Orera and J. M. Peña. Accurate determinants of some classes of matrices. *Linear Algebra Appl.*, 630:1–14, 2021.
- [83] H. Orera and J. M. Peña. Error bounds for linear complementarity problems of B_{π}^R -matrices. *Comput. Appl. Math.*, 40(3):Paper No. 94, 13, 2021.
- [84] J. M. Peña. M -matrices whose inverses are totally positive. *Linear Algebra Appl.*, 221:189–193, 1995.
- [85] J. M. Peña. A class of P -matrices with applications to the localization of the eigenvalues of a real matrix. *SIAM J. Matrix Anal. Appl.*, 22(4):1027–1037, 2001.
- [86] J. M. Peña. A stable test to check if a matrix is a nonsingular M -matrix. *Math. Comp.*, 73(247):1385–1392, 2004.
- [87] J. M. Peña. LDU decompositions with L and U well conditioned. *Electron. Trans. Numer. Anal.*, 18:198–208, 2004.
- [88] J. M. Peña. *Shape Preserving Representations in Computer-aided Geometric Design*. Nova Science Publishers, 1999.
- [89] A. Pinkus. *Totally positive matrices*, volume 181 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2010.
- [90] L. Qi and Z. Luo. *Tensor analysis*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2017. Spectral theory and special tensors.
- [91] H. B. Said. A generalized Ball curve and its recursive algorithm. *ACM Trans. Graph.*, 8(4):360–371, oct 1989.

- [92] G. D. Smith. *Numerical solution of partial differential equations*. Oxford Applied Mathematics and Computing Science Series. The Clarendon Press, Oxford University Press, New York, third edition, 1985. Finite difference methods.
- [93] Y. Song and L. Qi. Properties of some classes of structured tensors. *J. Optim. Theory Appl.*, 165(3):854–873, 2015.
- [94] T. Szulc. Some remarks on a theorem of Gudkov. *Linear Algebra Appl.*, 225:221–235, 1995.
- [95] M. J. Tsatsomeros and L. Li. A recursive test for P -matrices. *BIT*, 40(2):410–414, 2000.
- [96] C. F. Van Loan. The ubiquitous Kronecker product. *J. Comput. Appl. Math.*, 123(1-2):85–100, 2000.
- [97] J. M. Varah. A lower bound for the smallest singular value of a matrix. *Linear Algebra Appl.*, 11:3–5, 1975.
- [98] S. L. Yang and Z. K. Qiao. The Bessel numbers and Bessel matrices. *J. Math. Res. Exposition*, 31(4):627–636, 2011.
- [99] L. Zhang, L. Qi, and G. Zhou. M -tensors and some applications. *SIAM J. Matrix Anal. Appl.*, 35(2):437–452, 2014.
- [100] Z. Zhang. The linear algebra of the generalized Pascal matrix. *Linear Algebra Appl.*, 250:51–60, 1997.
- [101] Z. Zhang and M. Liu. An extension of the generalized Pascal matrix and its algebraic properties. *Linear Algebra Appl.*, 271:169–177, 1998.

Index of abbreviations

CAGD, Computer Aided Geometric Design,
207

DD, diagonally dominant, 17

HRA, high relative accuracy, 27

LCP, linear complementarity problem, 16

NE, Neville elimination, 21

RRD, rank revealing decomposition, 28

SDD, strictly diagonally dominant, 17

SF, subtraction-free, 28

STP, strictly totally positive, 20

TP, totally positive, 20