RIMS Kôkyûroku Bessatsu
Bx (202x), 000–000

On the norm of normal matrices

By

Ludovick Bouthat*, Javad Mashreghi** and Frédéric Morneau-Guérin***

Abstract

In this article, we present some recent results related to the calculation of the induced p-norm of $n \times n$ circulant matrices $A(n, a, b)$ with diagonal entries equal to $a \in \mathbb{R}$ and off-diagonal entries equal to $b \in \mathbb{R}$. For circulant matrices with nonnegative entries, an explicit formula for the induced p-norm ($1 \leq p \leq \infty$) is given, whereas for $A(n, -a, b)$, $a > 0$ the situation is no longer so simple and calls for a more subtle analysis. As a matter of fact, while the 2-norm of $A(n, -a, b)$ is precisely determined, the exact value of the induced p-norm for $1 < p < \infty$, $p \neq 2$, still remains elusive. Nevertheless, we provide a lower bound as well as two different categories of upper bounds. As an indication of not being far from the exact values, our estimates coincide at both ends points (i.e., $p = 1$ and $p = \infty$) as well as at $p = 2$ with the precise values. As an abstract approach, we also introduce the $*$-algebra generated by a normal matrix A accompanied by an axis-oriented norm, and obtain some estimations of the norm of elements of the $*$-algebra. We then exhibit the connection between the new generalized estimates and the previously obtained estimates in the special case where A is a circulant matrix. Finally, using an optimization-oriented approach, we provide insight on the nature of the maximizing vectors for $\frac{\|Ax\|_p}{\|x\|_p}$. This leads us to formulate a conjecture that, if proven valid, would make it possible to derive an exact formula for the induced p-norm of $A(n, a, b)$ whenever $a = \frac{1-n}{n}$ and $b = \frac{1}{n}$.

Received March 31, 2022. Revised ???, 2022.
2020 Mathematics Subject Classification(s): 15A60, 15B05, 47A30, 47A60
Key Words: *Circulant matrices, doubly stochastic matrices, p-norm, $*$-algebra, axis-oriented norms*
This work was partially supported by research grants from NSERC (Canada) and grants or scholarships from FRQNT (Quebec).

*Département de mathématiques et de statistique, Université Laval, Québec, QC, Canada G1K 7P4.
e-mail: Ludovick.Bouthat.1@ulaval.ca
**Département de mathématiques et de statistique, Université Laval, Québec, QC, Canada G1K 7P4.
e-mail: Javad.Mashreghi@mat.ulaval.ca
***Département Éducation, Université TÉLUQ, Québec, QC, Canada G1K 9H6.
e-mail: Frederic.Morneau-Guerin@teluq.ca

§ 1. Introduction

Given positive integers $m, n \geq 2$, let $\mathbb{C}^{m\times n}$ denote the set of $m \times n$ matrices with entries in the complex field $\mathbb{C}$. For $\alpha_0, \alpha_1, \ldots, \alpha_{n-1} \in \mathbb{C}$, the circulant matrix $\operatorname{circ}(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$ is defined as

$$\operatorname{circ}(\alpha_0, \alpha_1, \ldots, \alpha_{n-1}) \;=\; [\alpha_{i-j}]_{i,j=1}^{n} \;=\; \begin{pmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \cdots & \alpha_{n-1} \\ \alpha_{n-1} & \alpha_0 & \alpha_1 & \cdots & \alpha_{n-2} \\ \alpha_{n-2} & \alpha_{n-1} & \alpha_0 & \cdots & \alpha_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_0 \end{pmatrix}.$$

The index k in α_k is always calculated mod n, and thus runs between 0 and $n-1$. The circulant matrices are a very special type of square Toeplitz matrices [20, Chapter 4].

According to the mathematician and historian of early linear algebra Thomas Muir [39, Vol 2., Ch. 14], circulant matrices were first introduced, although somewhat implicitly, by Eugène Catalan in [9] in 1846. More than half a century elapsed, however, before the systematic study of this important class of matrices began to gain significant momentum. The first comprehensive monograph dedicated to the study of the various properties of circulant matrices, dating from 1979, was penned by Davis [16].

In recent years, circulant and block-circulant matrices have become omnipresent in many sub-disciplines of mathematics. For reasons that will become apparent later in our presentation, these matrices naturally prove themselves to be of fundamental importance in areas of mathematics where the roots of unity come into play. Additionally, they also have a wide range of applications in various parts of modern and classical mathematics. For instance, they have been used in the study of functional equations with several complex variables [46], in solving polynomial equations [30, 5], in solving various ordinary and partial differential equations [13, 17, 45, 34], in image processing [47], in signal processing [1, 25], in numerical analysis [12, 31, 48], in Wiener-Hopf equations [10, 7, 36, 6, 42], in information theory [21], and in various branches of operator theory (see [2, 27, 26, 22, 11, 18, 24, 29, 14, 28, 38, 32] and references therein). It goes without saying that, due to the abundance of contributions, our treatment is by no means exhaustive.

In order to provide a more refined definition of circulant matrices, we first need to introduce a particular operator, the *right circular shift*, that is acting on vectors of length n by operating a rearrangement of their entries consisting in moving the final entry to the first position while shifting all other entries to the next position. Formally, a right circular shift is an operator $S : \mathbb{C}^n \to \mathbb{C}^n$ defined by

$$S(\alpha_0, \alpha_1, \ldots, \alpha_{n-1}) \;:=\; (\alpha_{n-1}, \alpha_0, \ldots, \alpha_{n-2}).$$

One can view $A := \mathrm{circ}(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$ as the $n \times n$ matrix whose rows are given by iterations of the *right circular shift* acting on the vector $(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$, i.e., for $k = 1, \ldots, n$, the k-th row of A is $S^{k-1}(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$.

By identifying a circulant matrix with its first row, one can see that $\mathrm{circ}_n(\mathbb{C})$, the set of all complex-valued $n \times n$ circulant matrices, forms an n-dimensional vector space with respect to the usual operations of matrix addition and multiplication of matrices by scalars. This space can be interpreted as the space of complex-valued functions on $\mathbb{Z}/n\mathbb{Z}$, the cyclic group of order n. For more information concerning the circular shift operator and its connections with circulant matrices, we point the reader to the very detailed treatment given by Fuhrmann in [19, § 5.2 and § 5.3].

§ 2. Organization of the paper and a sketch of the main results

Recall that a function $\|\cdot\| : \mathbb{C}^{m\times n} \to [0, \infty)$ is a *matrix norm* or a *ring norm* if, for all $A, B \in \mathbb{C}^{m\times n}$, it satisfies the following five axioms:

(i) $\|A\| \geq 0$, *Nonnegativity,*

(ii) $\|A\| = 0$ if and only if $A = \mathbf{0}$, *Positivity,*

(iii) $\|\lambda A\| = |\lambda|\|A\|$ for all scalar λ, *Homogeneity,*

(iv) $\|A + B\| \leq \|A\| + \|B\|$, *Subadditivity,*

(v) $\|AB\| \leq \|A\|\|B\|$, *Submultiplicativity.*

The first four properties of a matrix norm are identical to the axioms for a norm. Since the set of $m \times n$ matrices with complex entries, can be equated to with the set of vectors of length $m \times n$ with complex entries, it does makes sense to endow $\mathbb{C}^{m\times n}$ with a norm that does not satisfy property (v). In order to avoid any confusion, such norms shall be called *vector norms*.

Suppose that a vector norm $\|\cdot\|_\alpha$ on $\mathbb{C}^n$ and a vector norm $\|\cdot\|_\beta$ on $\mathbb{C}^m$ are given. Then any $m \times n$ matrix A represents, with respect to the canonical basis, a linear operator from $\mathbb{C}^n$ to $\mathbb{C}^m$. We define the corresponding *induced matrix norm* or *operator norm* or *lub norm* (this acronym stands for the *least upper bound norm*) on the space $\mathbb{C}^{m\times n}$ of all $m \times n$ matrices as

$$\|A\|_{\alpha\to\beta} \;:=\; \sup\left\{ \frac{\|Ax\|_\beta}{\|x\|_\alpha} \,:\; x \in \mathbb{C}^n,\; x \neq \mathbf{0} \right\}.$$

The notation $\|A\|_{\alpha\to\beta}$ is a shorter replacement for $\|A\|_{(\mathbb{C}^n,\|\cdot\|_\alpha)\to(\mathbb{C}^m,\|\cdot\|_\beta)}$. Throughout this paper, the only operator norms that we shall consider are those induced by p-norms

for vectors. Recall that, for $p \geq 1$, the *p-norm* of vector $x = (x_1, \ldots, x_n)$ is defined by

$$\|x\|_p \;=\; \left(\sum_{k=1}^{n} |x_k|^p\right)^{1/p},$$

and the ∞-norm is

$$\|x\|_\infty \;=\; \max\{|x_1|, \ldots, |x_n|\}.$$

It is customary, in particular in operator theory texts, to denote the vector space $\mathbb{C}^n$ endowed by the vector p-norm by $\ell_p^n(\mathbb{C})$.

In this note, our main objective is a comprehensive study of the induced p-norm of a special class of circulant matrices, acting as operators from $\ell_p^n(\mathbb{C})$ to $\ell_p^n(\mathbb{C})$. The circulant matrices under consideration are those with the diagonal entries equal to $a \in \mathbb{R}$ and the off-diagonal entries equal to $b \in \mathbb{R}$, i.e.,

$$A(n,a,b) \;=\; \begin{pmatrix} a & b & b & \cdots & b \\ b & a & b & \cdots & b \\ b & b & a & \cdots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b & b & b & \cdots & a \end{pmatrix}, \tag{2.1}$$

where $a, b \in \mathbb{R}$. However, via a normalization process, it suffices to consider the following two cases: $A(n,a,b)$ and $A(n,-a,b)$, where $a, b \geq 0$.

Since the $A(n,a,b)$'s are matrices with real entries, it may appear that there is some ambiguity as to whether they are acting on $\mathbb{R}^n$ or on $\mathbb{C}^n$, and thus possibly ends with different p-norms for the real and complex cases. As a matter of fact, this is not an issue as it was shown by Taylor in [44] (1958) that for any $m \times n$ matrix B with real entries and all $q \geq p \geq 1$, the induced norm

$$\|B\|_{p\to q} \;:=\; \sup\left\{\frac{\|Bx\|_q}{\|x\|_p} \,:\, x \neq \mathbf{0}\right\}$$

is the same whether x runs through all non-zero vectors with complex entries or only through all non-zero vectors with real entries. Note that the author points out that a proof of this result had been sketched in an earlier paper (1927) by M. Riesz [40]. Also, we refer to Crouzeix [15] since this is, to the best of our knowledge, the most simple and direct proof of this result in the case that is considered here.

Our interest for estimating the induced p-norms of $A(n,\pm a,b)$ stems from other studies of the geometry of the set of doubly stochastic matrices. Indeed, one can verify that the *Chebyshev radius* of the n-dimensional Birkhoff polytope, i.e., the greatest lower bound of the radii of all balls containing $\mathcal{D}_n$, with respect to the induced p-norm

$(1 \le p \le \infty)$ is given by

$$R_p(\mathcal{D}_n) \;=\; \left\| I - \frac{1}{n} K \right\|_{p\to p},$$

where I is the $n \times n$ identity matrix and K is the $n \times n$ all-ones matrix. This corresponds to the induced p-norm of $-A\left(n, \frac{1-n}{n}, \frac{1}{n}\right)$.

In Section 3 of the present paper, we discuss some recent results (see [8, 41]) related to the calculation of $\|A(n, \pm a, b)\|_{p\to p}$ for $a, b \ge 0$ and $1 \le p \le \infty$. We shall see that the negative sign plays a crucial role, as the p-norms of $A(n, a, b)$ and $A(n, -a, b)$ differ significantly. Indeed,

$$\|A(n, a, b)\|_{p\to p} \;=\; a + (n-1)b, \qquad (1 \le p \le \infty),$$

whereas $\|A(n, -a, b)\|_{p\to p}$ is a non-constant function of p that is monotonically non-increasing for $1 \le p \le 2$ and monotonically non-decreasing for $2 \le p \le \infty$ (see [41]), with

$$\|A(n, -a, b)\|_{2\to 2} \;=\; \begin{cases} a + b, & \text{if } (n-2)b \le 2a, \\ -a + (n-1)b, & \text{if } (n-2)b \ge 2a. \end{cases}$$

But the exact value of the induced p-norm of $A(n, -a, b)$ for $1 < p < \infty$, $p \ne 2$, remains unknown. However, the following lower and upper bounds

$$\begin{cases} a + b \;\le\; \|A(n, -a, b)\|_{p\to p} \;\le\; n^{\left|\frac{1}{p} - \frac{1}{2}\right|}(a + b) & \text{if } (n-2)b \le 2a, \\ (n-1)b - a \;\le\; \|A(n, -a, b)\|_{p\to p} \;\le\; n^{\left|\frac{1}{p} - \frac{1}{2}\right|}((n-1)b - a) & \text{if } (n-2)b \ge 2a, \end{cases}$$

were obtained in [8]. Here, as a novel contribution, we provide a proof of the following refined estimates

$$\begin{cases} a + b \le \|A(n, -a, b)\|_{p\to p} \le (a + b)\left(\frac{(n-1)b + a}{a + b}\right)^{\left|\frac{2}{p} - 1\right|} & \text{if } (n-2)b \le 2a, \\ (n-1)b - a \le \|A(n, -a, b)\|_{p\to p} \le ((n-1)\,b - a)\left(\frac{(n-1)b + a}{(n-1)b - a}\right)^{\left|\frac{2}{p} - 1\right|} & \text{if } (n-2)b \ge 2a. \end{cases}$$

In Section 4, we take a wider perspective and investigate the induced p-norm of elements of the $*$-algebra generated by a normal matrix A in the case where $\|\cdot\|$ is an *axis-oriented matrix norm*. We then discuss how these estimates are related to those mentioned above in the case where A is a circulant matrix.

Finally, in Section 5 we open some new fronts in dealing with the question that was initially attracted our interest in estimating the induced p-norm of circulant matrices of the form given by (2.1). By leveraging the power of an optimization-oriented approach, we gain insights on the nature of the maximizing vectors for $\frac{\|Ax\|_p}{\|x\|_p}$. This leads us to

formulate a conjecture that, if proven valid, would make it possible to derive an exact formula for $\|I - \frac{1}{n}K\|_{p\to p}$.

§ 3. The induced p-norm of circulant matrices

As versatile and pervasive as circulant matrices are, a great deal of their key properties can be established using only elementary linear algebra. In this section, we will present those properties that shall prove useful later on.

§ 3.1. The DFT matrix

The following cyclic permutation matrix of order n

$$C_n \;:=\; \operatorname{circ}(0,1,0,\dots,0) \;=\; \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

is regarded as the *basic circulant matrix*. Indeed, C_n has the fundamental representation property

$$A \;=\; \alpha_0 I + \alpha_1 C_n + \alpha_2 C_n^2 + \cdots + \alpha_{n-1} C_n^{n-1} \;=\; \mathcal{P}_A(C_n), \tag{3.1}$$

where $A = \operatorname{circ}(\alpha_0, \alpha_1, \dots, \alpha_{n-1})$ and $\mathcal{P}_A$ is the polynomial

$$\mathcal{P}_A(z) \;=\; \alpha_0 + \alpha_1 z + \alpha_2 z^2 + \cdots + \alpha_{n-1} z^{n-1}.$$

This polynomial is called the *associated polynomial*, or *representer* (see [16, 33]), of the circulant matrix A.

The characteristic polynomial of C_n is $p_{C_n}(\lambda) = \lambda^n - 1$, and thus its eigenvalues are

$$\lambda_j \;:=\; \omega_n^j, \qquad 0 \le j \le n-1,$$

where $\omega_n = \exp(2\pi i/n)$ is a primitive n-th root of unity. The corresponding normalized eigenvectors are

$$x_j \;:=\; \frac{1}{\sqrt{n}}(1, \omega_n^j, \omega_n^{2j}, \dots, \omega_n^{(n-1)j})^{\intercal}, \qquad 0 \le j \le n-1.$$

Since these vectors are pairwise orthogonal, we deduce the diagonalization

$$C_n \;=\; W_n \operatorname{diag}(1, \omega_n, \omega_n^2, \dots, \omega_n^{n-1})\, W_n^*, \tag{3.2}$$

where W_n is the unitary matrix of order n whose columns are $x_0, x_1, \ldots, x_{n-1}$. More explicitly,

$$W_n \;=\; \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & \omega_n & \omega_n^2 & \cdots & \omega_n^{n-2} & \omega_n^{n-1} \\ 1 & \omega_n^2 & \omega_n^4 & \cdots & \omega_n^{2(n-2)} & \omega_n^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \omega_n^{n-2} & \omega_n^{(n-2)2} & \cdots & \omega_n^{(n-2)^2} & \omega_n^{(n-2)(n-1)} \\ 1 & \omega_n^{n-1} & \omega_n^{(n-1)2} & \cdots & \omega_n^{(n-1)(n-2)} & \omega_n^{(n-1)^2} \end{pmatrix}. \tag{3.3}$$

The sequence ω_n^k, $k \geq 0$, is periodic, and thus there are only n distinct elements in W_n. For example,

$$W_4 \;=\; \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix}.$$

As a matter of fact, W_n is a most remarkable matrix: it is easily established that it is *symmetric*, i.e. $W_n^{\intercal} = W_n$, and *unitary*, i.e., $W_n^{-1} = W_n^* = \overline{W}_n^{\intercal}$. The combination of these properties means that $W_n^{-1} = \overline{W}_n$. Moreover, it is the Vandermonde matrix for the roots of unity up to a normalization factor. What is more essential is that W_n is closely related to F_n, the unitary Discrete Fourier Transform (DFT) matrix [16, § 2.5] of order n which was introduced by Sylvester in 1867 [43]. Indeed, $W_n = F_n^* = \overline{F_n}$. Finally, we deduce from (3.1), (3.2), and $C_n^n = I$ the following essential representation. If $A \in \mathrm{circ}_n(\mathbb{C})$ then

$$A \;=\; W_n \,\mathrm{diag}\big(\mathcal{P}_A(1), \mathcal{P}_A(\omega_n), \mathcal{P}_A(\omega_n^2), \ldots, \mathcal{P}_A(\omega_n^{n-1})\big)\, W_n^*. \tag{3.4}$$

This means that the unitary matrix W_n simultaneously diagonalizes all $n \times n$ circulant matrices. For further details see [16, § 3.2].

It follows from the foregoing that, $\mathcal{P}_A(z)$, the associated polynomial of a given $A \in \mathrm{circ}_n(\mathbb{C})$ is actually closely related to $p_A(z)$, the characteristic polynomial of A. Indeed, $\mathcal{P}_A(z)$ is the unique polynomial of degree strictly less than n whose values at ω_n^j, $j = 0, 1, \ldots, n-1$, spans over each and every one of eigenvalues of A (counted with multiplicity), whereas $p_A(z)$ is the unique monic polynomial of degree n that vanishes precisely at the eigenvalues of A.

To conclude this introductory commentary, remark that, in the light of above observations, one can easily check that the product of two circulant matrices is again circulant and that, for this set of matrices, multiplication is commutative. Thus, $\mathrm{circ}_n(\mathbb{C})$ forms

a commutative $*$-algebra over $\mathbb{C}$, with involution given by the conjugate transpose. Furthermore, $\mathrm{circ}_n(\mathbb{C})$ is canonically isomorphic to $\mathrm{diag}_n(\mathbb{C})$, the set of all complex-valued diagonal matrices of order n.

§ 3.2. The induced p-norm of $A(n, \pm a, b)$, Cases $p = 1$ and $p = \infty$

As previously mentioned, we seek to find the induced p-norm of the circulant matrices $A(n, \pm a, b)$ as given by (2.1). For $p = \infty$, the situation is straightforward. Indeed, given any matrix $A \in \mathbb{C}^{n\times n}$, we have

$$\|Ax\|_\infty \;=\; \max_{1\le i\le n}\left|\sum_{j=1}^{n} a_{i,j}x_j\right| \;\le\; \|x\|_\infty \max_{1\le i\le n}\sum_{j=1}^{n}|a_{i,j}|.$$

This shows that $\|A\|_{\infty\to\infty} \le \max\limits_{1\le i\le n}\sum\limits_{j=1}^{n}|a_{i,j}|$. For the reverse inequality, we choose an index $1 \le i_0 \le n$ that realizes the maximum in the previous expression and we consider $x \in \mathbb{C}^n$ such that $x_j := \mathrm{sgn}(a_{i_0,j})$, where $\mathrm{sgn}(z)$ denotes the complex signum function defined as

$$\mathrm{sgn}(z) \;=\; \begin{cases} \frac{z}{|z|}, & z \neq 0, \\ 0, & z = 0. \end{cases}$$

Hence, $\|x\|_\infty = 1$ and

$$\|Ax\|_\infty \;=\; \max_{1\le i\le n}\left|\sum_{j=1}^{n} a_{i,j}x_j\right| \ge \left|\sum_{j=1}^{n} a_{i_0,j}x_j\right| = \sum_{j=1}^{n}|a_{i_0,j}| = \max_{1\le i\le n}\sum_{j=1}^{n}|a_{i,j}|.$$

Therefore,

$$\|A\|_{\infty\to\infty} \;=\; \max_{1\le i\le n}\sum_{j=1}^{n}|a_{i,j}|,$$

which is simply the maximum absolute row sum of the matrix. In particular, for $a, b \ge 0$ and for $A(n, \pm a, b)$ as given by (2.1), we have

$$\|A(n, \pm a, b)\|_{\infty\to\infty} \;=\; |a| + (n-1)|b|.$$

From the above calculation, we can directly also infer that

$$\|A(n, \pm a, b)\|_{1\to 1} = |a| + (n-1)|b|.$$

Indeed $\|A\|_{p\to p} = \|A\|_{q\to q}$ for Hölder conjugate indices p and q, since A is self-adjoint [23, § 5.4 and Theorem 5.3.35]. Nevertheless, we provide a different detailed proof, which is interesting in its own right. Given any $x \in \mathbb{C}^n$, consider a partition of a

matrix $A \in \mathbb{C}^{n\times n}$ according to its columns as $A = [a_1|a_2|\cdots|a_n] \in \mathbb{C}^{n\times n}$. The triangle inequality gives

$$\|Ax\|_p \;=\; \Big\| \sum_{j=1}^{n} x_j a_j \Big\|_p \;\leq\; \sum_{j=1}^{n} |x_j| \cdot \|a_j\|_p \;=\; \|x\|_1 \max_{1\leq j\leq n} \|a_j\|_p$$

This shows that $\|A\|_{1\to p} \leq \max\limits_{1\leq j\leq n} \|a_j\|_p$. For the reverse inequality, we choose an index $1 \leq j_0 \leq n$ that realizes the maximum in the previous expression and we consider the canonical basis vector e_{j_0}, i.e., the vector whose components are all zero, except the j_0-th that equals one. Then

$$\|Ae_{j_0}\|_p \;=\; \|a_{j_0}\|_p = \max_{1\leq j\leq n} \|a_j\|_p.$$

Hence,

$$\|A\|_{1\to p} \;=\; \max_{1\leq j\leq n} \|a_j\|_p.$$

For $p = 1$, this is simply the maximum absolute column sum of the matrix. In particular, for $a, b \geq 0$ and for $A(n, \pm a, b)$ as given by (2.1), we have

$$\|A(n, \pm a, b)\|_{1\to 1} \;=\; |a| + (n-1)|b|.$$

As a by-product, we also deduce

$$\|A(n, \pm a, b)\|_{1\to p} \;=\; \big(|a|^p + (n-1)|b|^p\big)^{1/p}.$$

§ 3.3. The induced p-norm of $A(n, \pm a, b)$, Case $p = 2$

For $1 < p < \infty$, the situation is no longer so simple. One cannot easily calculate the induced p-norm of $A(n, \pm a, b)$ by inspection. As a matter of fact, we shall see that the sign of the real number a plays a crucial role as the p-norms of $A(n, a, b)$ and $A(n, -a, b)$ are entirely different in general. Before discussing the general case, we start by addressing the case $p = 2$. To do this, we first remark that, for a diagonal matrix, all the induced p-norms are equal to the maximum of the absolute value of the entries. In [41], Sahasranand calculated this value for diagonal matrices involved in the diagonalization of $A(n, a, b)$ and $A(n, -a, b)$ as in (3.4).

Lemma 3.1. *Let $a, b \geq 0$, and let $A = A(n, \pm a, b)$ be given by (2.1). Let $W_n \Lambda W_n^*$ be the diagonalization of A as in (3.4) and assume that $1 \leq p \leq \infty$.*

(i) If $A = A(n, a, b) = W_n \Lambda W_n^*$, then we have

$$\|\Lambda\|_{p\to p} = a + (n-1)b.$$

(ii) If $A = A(n, -a, b) = W_n \Lambda W_n^*$, then we have

$$\|\Lambda\|_{p\to p} = \begin{cases} a+b, & \text{if } (n-2)b \le 2a, \\ -a+(n-1)b, & \text{if } (n-2)b \ge 2a. \end{cases}$$

Proof. By (3.4), the matrix A can be expressed as

$$A \;=\; W_n \Lambda W_n^*,$$

where Λ is a diagonal matrix whose entries on the main diagonal are given by

$$\Lambda_{1,1} \;=\; \mathcal{P}_A(1) \;=\; a+(n-1)b$$

and, for $k = 2, \dots, n$,

$$\begin{aligned} \Lambda_{k,k} &= \mathcal{P}_A(\omega_n^{k-1}) \\ &= a + b\omega_n^{k-1} + b\omega_n^{2(k-1)} + \cdots + b\omega_n^{(n-1)(k-1)} \\ &= (a-b) + b\sum_{j=0}^{n-1} w_n^{j(k-1)} \\ &= a-b. \end{aligned}$$

The last equality results from the application of the following well-known identity (see [35, § 3, Ex. 3]),

$$\sum_{j=0}^{n-1} \omega_n^{js} \;=\; \begin{cases} n, & \text{if } s = 0 \mod n, \\ 0, & \text{otherwise.} \end{cases}$$

The result follows by calculating

$$\begin{aligned} \|\Lambda\|_{p\to p} &= \sup\left\{ \frac{\|\Lambda x\|_p}{\|x\|_p} : \; x \neq \mathbf{0} \right\} \\ &= \sup\left\{ \frac{\Big(\sum\limits_{k=1}^{n} |\Lambda_{k,k} x_k|^p \Big)^{1/p}}{\Big(\sum\limits_{k=1}^{n} |x_k|^p \Big)^{1/p}} : \; x \neq \mathbf{0} \right\} \\ &= \max_{1\le k\le n} |\Lambda_{k,k}| \end{aligned}$$

for $1 \le p \le \infty$, first for $A(n, a, b)$ and then for $A(n, -a, b)$. □

As in [41], we now use the previous lemma to derive an exact expression for $\|A(n, a, b)\|_{2\to 2}$ and $\|A(n, -a, b)\|_{2\to 2}$ for $a, b \ge 0$. In doing so, we obtain an alternative proof of the result presented in [8].

Theorem 3.2. *Let $a, b \geq 0$.*

(i) For $A = A(n, a, b)$ defined as in (2.1), we have

$$\|A\|_{2\to 2} \;=\; a + (n-1)b.$$

(ii) For $A = A(n, -a, b)$ defined as in (2.1), we have

$$\|A\|_{2\to 2} \;=\; \begin{cases} a + b, & \text{if } (n-2)b \;\leq\; 2a, \\ -a + (n-1)b, & \text{if } (n-2)b \;\geq\; 2a. \end{cases}$$

Proof. Let $W_n \Lambda W_n^*$ be the diagonalization of $A(n, \pm a, b)$ as in (3.4). We have

$$\|A\|_{2\to 2} \;=\; \|W_n \Lambda W_n^*\|_{2\to 2} \;\leq\; \|W_n\|_{2\to 2} \cdot \|\Lambda\|_{2\to 2} \cdot \|W_n^*\|_{2\to 2}.$$

Observe that

$$\|W^* x\|_2^2 \;=\; (W^* x)^* W^* x \;=\; x^* W W^* x \;=\; x^* x \;=\; \|x\|_2^2.$$

From this relation – which bears the name of *Parseval's identity* – we deduce that $\|W^*\|_{2\to 2} = 1$, and the same goes for W. The first part of Lemma 3.1 implies that

$$\|A(n, a, b)\|_{2\to 2} \;\leq\; a + (n-1)b, \tag{3.5}$$

while the second part entails that

$$\|A(n, -a, b)\|_{2\to 2} \;\leq\; a + b \tag{3.6}$$

if $(n-2)b \leq 2a$, and

$$\|A(n, -a, b)\|_{2\to 2} \;\leq\; -a + (n-1)b \tag{3.7}$$

if $(n-2)b \geq 2a$.

One can easily check that $x = (1, 1, \ldots, 1)^\intercal$ is actually a maximizing vector for (3.5) and (3.7), i.e., that we can replace $\geq$ by the equal sign in these equations. As for the inequality in (3.6), it suffices to consider the vector $y = (1, -1, 0, \ldots, 0)^\intercal$ to see that it too is saturated. □

We shall now present a generalization of the case $A(3, -a, b)$.

Proposition 3.3. *If $A = \mathrm{circ}(\alpha_1, \alpha_2, \alpha_3)$ for arbitrary $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$, then*

$$\|A\|_{2\to 2} \;=\; \begin{cases} \sqrt{\alpha_1^2 + \alpha_2^2 + \alpha_3^2 - (\alpha_1\alpha_2 + \alpha_2\alpha_3 + \alpha_3\alpha_1)}, & \text{if } \alpha_1\alpha_2 + \alpha_2\alpha_3 + \alpha_3\alpha_1 \leq 0, \\ |\alpha_1 + \alpha_2 + \alpha_3|, & \text{if } \alpha_1\alpha_2 + \alpha_2\alpha_3 + \alpha_3\alpha_1 \geq 0. \end{cases}$$

Note that the special case $b = \alpha_2 = \alpha_3 \geq 0$ and $-a = \alpha_1 < 0$ yields the case $A(3, -a, b)$. The proof of Proposition 3.3 is omitted since we shall now state and prove a more general theorem, presented in [8], that subsumes both previous results and that conceptually explains why there are two cases that arise in calculations.

Theorem 3.4. *Suppose $A = [c_1|c_2|\cdots|c_n] \in \mathbb{R}^{n\times n}$ is an arbitrary matrix whose columns $c_1, \ldots, c_n \in \mathbb{R}^n$ satisfy*

$$c_1^\intercal c_1 \,=\, \ldots \,=\, c_n^\intercal c_n \,=\, \rho, \qquad c_i^\intercal c_j \,=\, \beta, \quad (1 \leq i < j \leq n),$$

for some scalars $\rho, \beta \in \mathbb{R}$. Then

$$\|A\|_{2\to 2} \;=\; \begin{cases} \sqrt{\rho - \beta}\,, & \text{if } \beta \,\leq\, 0, \\ \sqrt{\rho + (n-1)\beta}, & \text{if } \beta \,\geq\, 0. \end{cases}$$

Proof. It is well-known that the square of the induced 2-norm (also called the *spectral norm*) of any real $n \times n$ matrix A is precisely the largest eigenvalue of the symmetric matrix $A^\intercal A$. Then, under the above-stated hypotheses,

$$A^\intercal A \;=\; \begin{pmatrix} \rho & \beta & \beta & \cdots & \beta \\ \beta & \rho & \beta & \cdots & \beta \\ \beta & \beta & \rho & \cdots & \beta \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta & \beta & \beta & \cdots & \rho \end{pmatrix} \;=\; (\rho - \beta)I + \beta K,$$

where I is the $n \times n$ identity matrix and K is the $n \times n$ all-ones matrix (i.e., the matrix where every element is equal to 1). The eigenvalues of the positive semidefinite matrix $A^\intercal A = (\rho - \beta)I + \beta K$ are $\rho - \beta$ (with multiplicity of $n - 1$) and $\rho + (n-1)\beta$ (with multiplicity 1). Taking their maximum and then its square root, the result follows. □

Note that Theorem 3.2.(ii) corresponds to the case

$$\rho \;=\; a^2 + (n-1)b^2 \quad \text{and} \quad \beta \;=\; -2ab + (n-2)b^2,$$

whereas Proposition 3.3 corresponds to the case $n = 3$, and

$$\rho \;=\; \alpha_1^2 + \alpha_2^2 + \alpha_3^2 \quad \text{and} \quad \beta \;=\; \alpha_1\alpha_2 + \alpha_2\alpha_3 + \alpha_3\alpha_1,$$

of Theorem 3.4.

§ 3.4. The p-norm of $A(n, a, b)$

In [8, Theorem 3.2], it was shown that if $a, b \geq 0$, then for all $1 \leq p \leq \infty$, $\|A\|_{p\to p} = a + (n-1)b$. Here, we present a short proof taken from [41].

Theorem 3.5. *Let $a, b \geq 0$, and let $A = A(n, a, b)$ be given by (2.1). Assume $1 \leq p \leq \infty$. Then*

$$\|A\|_{p\to p} \;=\; a + (n-1)b.$$

Proof. Using the vector $x = (1, 1, \ldots, 1)^\intercal$, we see that

$$\|A(n, a, b)\|_{p\to p} \;\geq\; \frac{\|A(n, a, b)x\|_p}{\|x\|_p} \;=\; a + (n-1)b.$$

This implies that, in order to prove the result, it suffices to show that $x = (1, 1, \ldots, 1)^\intercal$ is actually a maximizing vector.

We have seen that

$$\|A\|_{\infty\to\infty} \;=\; a + (n-1)b \;=\; \|A\|_{2\to 2}.$$

Now, by applying the Riesz–Thorin interpolation theorem [37, Theorem 1.1.1] with $p_0 = q_0 = 2$, $p_1 = q_1 = \infty$ and $p_\theta = q_\theta$, we obtain,

$$\|A\|_{p_\theta\to p_\theta} \;\leq\; \|A\|_{2\to 2}^{1-\theta}\|A\|_{\infty\to\infty}^{\theta}, \tag{3.8}$$

for (with a mild abuse of notation)

$$\frac{1}{p_\theta} \;=\; \frac{1-\theta}{2} + \frac{\theta}{\infty} \;=\; \frac{1-\theta}{2}, \qquad (0 \leq \theta \leq 1).$$

Hence,

$$\|A\|_{p_\theta\to p_\theta} \;\leq\; \|A\|_{\infty\to\infty} \;=\; a + (n-1)b,$$

for $2 \leq p_\theta \leq \infty$. Since A is self-adjoint, the result also follows for $1 \leq p \leq 2$. □

Using similar arguments as in the proof of Theorem 3.5, one can show that

$$\|\operatorname{circ}(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})\|_{p\to p} \;\leq\; \alpha_0 + \alpha_1 + \cdots + \alpha_{n-1},$$

where $\alpha_j \geq 0$. Then, considering again the same maximizing vector $x = (1, 1, \ldots, 1)^\intercal$, one concludes that

$$\|\operatorname{circ}(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})\|_{p\to p} \;=\; \alpha_0 + \alpha_1 + \cdots + \alpha_{n-1}.$$

§ 3.5. The p-norm of $A(n, -a, b)$

It turns out that the calculations of the induced p-norms of $A(n, -a, b)$ with $a, b \geq 0$ are more involved than those of $A(n, a, b)$. In fact, to the best of our knowledge, the exact value of $\|A(n, -a, b)\|_{p\to p}$ for $1 < p < \infty$, $p \neq 2$ is not known yet. Note however that lower bounds were provided in [8] as well as the two different set of upper bounds that we shall now present.

Theorem 3.6. *Let $a, b \geq 0$, and let $A = A(n, -a, b)$ be given by (2.1). Assume $1 \leq p \leq \infty$. Then*

$$\|A\|_{p\to p} \;\leq\; n^{\left|\frac{1}{p}-\frac{1}{2}\right|}\|A\|_{2\to 2}.$$

Proof. We follow [41]. Because A is self-adjoint, we will assume without any loss of generality that $2 \leq p \leq \infty$. Recall that for vectors in $\mathbb{C}^n$, we have

$$\|x\|_2^2 \;=\; \overline{x}^{\intercal}x \;=\; x^*x.$$

Moreover, for $1 \leq r \leq p$,

$$\|x\|_p \;\leq\; \|x\|_r \;\leq\; n^{\frac{1}{r}-\frac{1}{p}}\|x\|_p. \tag{3.9}$$

Both inequalities will prove useful in the special case $2 = r \leq p$.

For $x \neq \mathbf{0}$, the essential representation (3.4) implies that

$$\begin{aligned}\|Ax\|_p \;&\leq\; \|Ax\|_2\\ &=\; \|W_n\Lambda W_n^*x\|_2\\ &\leq\; \|W_n\|_{2\to 2}\cdot\|\Lambda\|_{2\to 2}\cdot\|W_n^*x\|_2.\end{aligned}$$

By the Parseval's identity and that $\|W_n\|_{2\to 2} = 1$, we deduce that

$$\|Ax\|_p \;\leq\; \|\Lambda\|_{2\to 2}\cdot\|x\|_2 \;\leq\; \|\Lambda\|_{2\to 2}\cdot n^{\frac{1}{2}-\frac{1}{p}}\|x\|_p$$

whereby

$$\|A\|_{p\to p} \;\leq\; n^{\frac{1}{2}-\frac{1}{p}}\|\Lambda\|_{2\to 2}.$$

The desired result now follows from Lemma 3.1. □

Using once again the Riesz–Thorin interpolation theorem, it is possible to obtain another set of upper bounds for $\|A(n, -a, b)\|_{p\to p}$.

Theorem 3.7. *Let $a, b \geq 0$, and let $A = A(n, -a, b)$ be given by (2.1). Assume $1 \leq p \leq \infty$. Then*

$$\|A\|_{p\to p} \;\leq\; \|A\|_{2\to 2}\left(\frac{\|A\|_{\infty\to\infty}}{\|A\|_{2\to 2}}\right)^{\left|1-\frac{2}{p}\right|}.$$

Proof. For $2 \leq p \leq \infty$, it suffices to interpolate between between $\|A\|_{2\to 2}$ and $\|A\|_{\infty\to\infty}$. As in the proof of Theorem 3.5, fix $p_0 = q_0 = 2$, $p_1 = q_1 = \infty$, and $q_\theta = p_\theta$ to obtain

$$\|A\|_{p_\theta\to p_\theta} \;\leq\; \|A\|_{2\to 2}^{1-\theta}\|A\|_{\infty\to\infty}^{\theta} \tag{3.10}$$

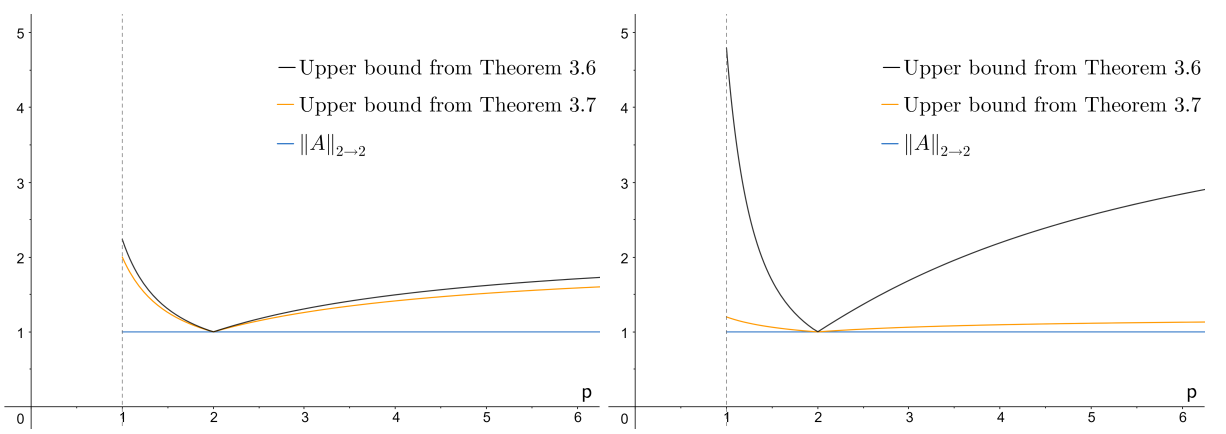
for (with a mild abuse of notation)

$$\frac{1}{p_\theta} \;=\; \frac{1-\theta}{2} + \frac{\theta}{\infty} \;=\; \frac{1-\theta}{2}, \qquad (0 \le \theta \le 1).$$

Remark that $\theta = 1 - \frac{2}{p_\theta}$. Hence (3.10) can be restated as

$$\|A\|_{p_\theta \to p_\theta} \;\le\; \|A\|_{2\to 2}^{\frac{2}{p_\theta}} \|A\|_{\infty\to\infty}^{1-\frac{2}{p_\theta}}, \qquad (2 \le p_\theta \le \infty).$$

Since A is self-adjoint, the result follows for $1 \le p \le 2$ with a bit of simplification. □



(a) $n = 5$, $a = \frac{2}{3}$ and $b = \frac{1}{3}$. (b) $n = 23$, $a = 0.1$ and $b = 0.05$.

Figure 1: Upper and lower bounds of $\|A\|_{p\to p}$ for different values of n, a, b.

Even though the estimations provided in Theorem 3.7 are more complex than those given in Theorem 3.6, they are more accurate. In fact, when n is small, the difference between the two estimations is marginal. However, as n grows, the first estimation becomes significantly better than the latter (see Figure 1, where the parameters a, b have been normalized so that $\|A\|_{2\to 2} = 1$). Note, however, that neither of them provide the precise value of $\|A(n, -a, b)\|_{p\to p}$.

Open question 3.8. *Evaluate* $\|A(n, -a, b)\|_{p\to p}$.

Using similar arguments as in the proof of Theorem 3.7, Sahasranand [41, Theorem 6] showed the following.

Theorem 3.9. *Let* $a, b \ge 0$*, and let* $A = A(n, -a, b)$ *be given by* (2.1)*. For* $1 \le p \le 2$*,* $\|A\|_{p\to p}$ *is monotonically non-increasing in* p*, whereas for* $2 \le p \le \infty$*,* $\|A\|_{p\to p}$ *is monotonically non-decreasing in* p*.*

Proof. Fix $p \ge 2$ and $\beta > 0$. Let $\alpha \ge 0$ be such that $\frac{1}{p-\alpha} + \frac{1}{p+\beta} = 1$. This time, we apply the Riesz–Thorin interpolation theorem with $p_0 = q_0 = p - \alpha$, $p_1 = q_1 = p + \beta$

and $p_\theta = q_\theta$ to obtain

$$\|A\|_{p_\theta \to p_\theta} \ \leq \ \|A\|^{1-\theta}_{(p-\alpha)\to(p-\alpha)} \|A\|^{\theta}_{(p+\beta)\to(p+\beta)}, \tag{3.11}$$

for

$$\frac{1}{p_\theta} \ = \ \frac{1-\theta}{p-\alpha} + \frac{\theta}{p+\beta}, \qquad (0 \leq \theta \leq \ 1).$$

Since $p-\alpha$ and $p+\beta$ are Hölder conjugates and since A is self-adjoint, we have $\|A\|_{(p-\alpha)\to(p-\alpha)} = \|A\|_{(p+\beta)\to(p+\beta)}$ and (3.11) can be restated as

$$\|A\|_{p\to p} \ \leq \ \|A\|_{(p+\beta)\to(p+\beta)},$$

which implies that $\|A\|_{p\to p}$ is monotonically non-decreasing in p for $2 \leq p \leq \infty$. The case $1 \leq p \leq 2$ is dealt with similarly. □

Finally, as Sahasranand [41] pointed out, conditional upon knowing the exact value of $\|A(n,-a,b)\|_{r\to r}$ for some $r > p > 2$, yet another use of Riesz–Thorin interpolation theorem allows us to derive an upper bound $\|A\|_{p\to p}$ that is even more precise than the one given by Theorem 3.7.

Corollary 3.10. *Let $a, b \geq 0$, and let $A = A(n,-a,b)$ be given by* (2.1)*. Assume $2 \leq p \leq \infty$. Then, for $\beta \geq \ 0$,*

$$\|A\|_{p\to p} \ \leq \ \|A\|^{\frac{\frac{1}{p}-\frac{1}{p+\beta}}{\frac{1}{2}-\frac{1}{p+\beta}}}_{2\to 2} \|A\|^{\frac{\frac{1}{2}-\frac{1}{p}}{\frac{1}{2}-\frac{1}{p+\beta}}}_{(p+\beta)\to(p+\beta)}.$$

Obviously, an analogous result holds for $1 \leq p \leq 2$.

§ 4. An operator theoretic approach

In this section, we adopt an abstract approach to study the norm. Our setting is general enough and in the special case of circulants, leads us to some of our previous results as well as some new estimations.

§ 4.1. Axis-oriented norms

Suppose we endow $\mathbb{C}^{n\times n}$, the ring of complex $n \times n$ matrices, with a norm $\|\cdot\|$ which satisfies

$$\|\mathrm{diag}(z_1,\ldots,z_n)\| \ = \ \max\{|z_1|,\ldots,|z_n|\}. \tag{4.1}$$

Recall that the spectral radius of a given matrix $A \in \mathbb{C}^{n\times n}$ whose eigenvalues are $\lambda_1,\ldots,\lambda_n$ is defined as

$$\rho(A) \ := \ \max\{|\lambda_1|,\ldots,|\lambda_n|\}.$$

The spectral radius, though not itself a matrix norm, is some sort of infimum over all matrix norms, in that $\rho(A) \leq \|A\|$ for all $A \in \mathbb{C}^{n\times n}$ and all matrix norm $\|\cdot\|$. Therefore, a norm satisfying (4.1) is optimal for all diagonal matrices.

Matrix norms induced by a vector norm verifying (4.1) were said to be *axis-oriented* by Bauer and Fike in [3]. As it turns out, using the following equivalence relations (see [4, Theorems 2 and 3]), one can easily verify that axis-orientedness is not an uncommon property of induced matrix norms.

Theorem 4.1. *Let $n \geq 2$, and let $\|\cdot\|$ be a vector norm on $\mathbb{C}^n$. The following are equivalent.*

(i) The matrix norm induced by $\|\cdot\|$ satisfies (4.1).

(ii) The vector norm $\|\cdot\|$ is absolute, i.e., $\|x\| = \big\||x|\big\|$ for all $x \in \mathbb{C}^n$, where $|x|$ denote the vector the components of which are the moduli of the components of x.

(iii) The vector norm $\|\cdot\|$ is monotonic, i.e., $\|x\| \leq \|y\|$ whenever $|x_i| \leq |y_i|$ for all $1 \leq i \leq n$.

One can easily verify from their definition that the p-norms ($1 \leq p \leq \infty$) are absolute. Therefore, their respective induced matrix norms satisfy (4.1).

The extension of the concept of axis-orientedness to all matrix norms (that is, even to the matrix norms that are not induced by any vector norm) and to vector norms on $\mathbb{C}^{n\times n}$ is a natural step forward since there are vector norms and non-induced matrix norms satisfying (4.1). Consider for instance the element-wise (vector) norm

$$\|A\|_{\max} := \max_{1\leq i,j\leq n} |a_{ij}|,$$

which is not sub-multiplicative and thus not a matrix norm, or the sub-multiplicatives matrix norms

$$\|A\|_{\max\{p,q\}} := \max\{\|A\|_{p\to p}, \|A\|_{q\to q}\}, \qquad (1 \leq p, q \leq \infty,\ p \neq q),$$

which are not *minimal*, hence non-induced (see [23, Theorem 5.6.32]).

§ 4.2. The $*$-algebra $\mathcal{T}_A$

Let $\mathcal{P}$ denote the set of all analytic polynomials in two independent variables, i.e.,

$$p(z, w) := \sum_{m,n=0}^{N} a_{mn} z^m w^n.$$

Note that, being in the world of commutative polynomial, there is no difference between zw and wz. Hence, for an arbitrary polynomial p, the combination $p(A, B)$ will be

meaningful if the matrices A and B commute. In particular, $p(A, A^*)$ makes sense whenever A is normal, i.e., $AA^* = A^*A$.

Given a normal matrix A, let $\mathcal{T}_A$ be the $*$-algebra generated by A, i.e.,

$$\mathcal{T}_A \;=\; \{p(A, A^*) : p \in \mathcal{P}\}.$$

We may even generalize the above definition by considering the functions $f(z, w)$ which are defined at least on the set $\sigma(A) \times \overline{\sigma(A)}$. However, we stick to polynomials to make the presentation free of some technical difficulties.

By the Spectral Theorem for normal matrices [23, Theorem 2.5.3], there exists a unitary matrix U such that $A = U\Lambda U^*$, where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix whose entries are the eigenvalues of A. Therefore, an easy algebraic computation shows

$$p(A, A^*) \;=\; Up(\Lambda, \Lambda^*)U^* \;=\; U\operatorname{diag}\big(p(\lambda_1, \overline{\lambda_1}), \ldots, p(\lambda_n, \overline{\lambda_n})\big)U^*. \tag{4.2}$$

In subsequent subsections, we seek to calculate (or, failing that, estimate) the value of $\|p(A, A^*)\|$, where $\|\cdot\|$ is some axis-oriented matrix norm.

§ 4.3. Unitarily invariant norms

A norm $\|\cdot\|$ on $\mathbb{C}^{n\times n}$ is called *unitarily invariant* if

$$\|UX\| \;=\; \|XU\| \;=\; \|X\|$$

for all $X \in \mathbb{C}^{n\times n}$ and all unitary matrices $U \in \mathbb{C}^{n\times n}$. Important examples of unitarily invariant norms include the spectral norm $\|A\| = \sqrt{\rho(A^*A)}$, the Frobenius norm

$$\|A\|_F \;:=\; \left(\sum_{i=1}^{n}\sum_{j=1}^{n} |a_{i,j}|^2\right)^{1/2},$$

and the Schatten p-norms

$$\|A\|_{\mathcal{S}_\infty} \;:=\; \max_{1\le i\le n} \sigma_i(A) \qquad \& \qquad \|A\|_{\mathcal{S}_p} \;:=\; \left(\sum_{i=1}^{n} \sigma_i^p(A)\right)^{1/p}, \quad (1 \le p < \infty),$$

where σ_i's designate the singular values of $A \in \mathbb{C}^{n\times n}$.

One needs only look to the Frobenius norm to see that not all unitarily invariant norms on $\mathbb{C}^{n\times n}$ satisfy the condition (4.1). In fact, it is not difficult to verify that the spectral norm is the *only* unitarily invariant norm on $\mathbb{C}^{n\times n}$ that is axis-oriented. By the singular value decomposition theorem [23, Theorem 2.6.3], for all $A \in \mathbb{C}^{n\times n}$, there are unitary matrices $U, V \in \mathbb{C}^{n\times n}$ such that $A = U\Sigma V^*$, where $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$

is a diagonal matrix whose entries are the singular values of A listed in non-increasing order. Hence, if $\|\cdot\|$ is a unitarily invariant norm satisfying (4.1), then

$$\|A\| = \|U\Sigma V^*\| = \|\Sigma\| = \|\mathrm{diag}(\sigma_1, \ldots, \sigma_n)\| = \max\{|\sigma_1|, \ldots, |\sigma_n|\} = \sigma_1 = \|A\|_{2\to 2}.$$

The following result is an immediate consequence of (4.2) and the assumption (4.1).

Proposition 4.2. *For all normal matrix $A \in \mathbb{C}^{n\times n}$ and all polynomial in two variables p,*

$$\|p(A, A^*)\|_{2\to 2} \;=\; \max\left\{|p(\lambda_1, \overline{\lambda_1})|, \ldots, |p(\lambda_n, \overline{\lambda_n})|\right\}.$$

§ 4.4. More general norms

If $\|\cdot\|$ is not unitarily invariant, then some upper and lower estimates come into play. A given matrix U acts as an operator on $\mathbb{C}^{n\times n}$ by left or right multiplication. Hence, we naturally consider the *left operator norm induced by* $\|\cdot\|$

$$\|U\|_\ell \;:=\; \sup\left\{\frac{\|UX\|}{\|X\|} : X \in \mathbb{C}^{n\times n}, X \neq \mathbf{0}\right\}, \tag{4.3}$$

and the *right operator norm induced by* $\|\cdot\|$

$$\|U\|_r \;:=\; \sup\left\{\frac{\|XU\|}{\|X\|} : X \in \mathbb{C}^{n\times n}, X \neq \mathbf{0}\right\}. \tag{4.4}$$

In other words, $\|U\|_\ell$ and $\|U\|_r$ are respectively the best constant such that

$$\|UX\| \;\le\; \|U\|_\ell \|X\|, \qquad X \in \mathbb{C}^{n\times n}, \tag{4.5}$$

and

$$\|XU\| \;\le\; \|U\|_r \|X\|, \qquad X \in \mathbb{C}^{n\times n}. \tag{4.6}$$

If U is a unitary matrix, we can replace U by U^* in the above inequalities and obtain the corresponding bounds for U^*. Then, depending on the case, replacing X by UX or XU, we obtain the lower bounds

$$\|UX\| \;\ge\; \frac{1}{\|U^*\|_\ell}\|X\|, \qquad X \in \mathbb{C}^{n\times n}, \tag{4.7}$$

and

$$\|XU\| \;\ge\; \frac{1}{\|U^*\|_r}\|X\|, \qquad X \in \mathbb{C}^{n\times n}. \tag{4.8}$$

In the important case where $\|\cdot\|$ is submultiplicative, we have $\|A\|_\ell \le \|A\|$ and $\|A\|_r \le \|A\|$ for every $A \in \mathbb{C}^{n\times n}$. The reverse inequalities trivially hold if $\|\cdot\|$ satisfies $\|I\| = 1$. From these observations, we can deduce the following result.

Proposition 4.3. *If $\|\cdot\|$ is a matrix norm on $\mathbb{C}^{n\times n}$ satisfying $\|I\| = 1$, then the left operator norm and the right operator norm coincide with the norm itself.*

Although there is a vast class of norms for which $\|\cdot\|_\ell \equiv \|\cdot\|_r \equiv \|\cdot\|$, and chief among them will be the induced matrix norms, this equivalence does not always occur. One only needs to took at the left and right operator norms induced by the Frobenius norm. For the identity matrix I, for instance, we have

$$\|I\|_\ell = \|I\|_r = 1 \ < \ \sqrt{n} = \|I\|_F.$$

It should furthermore be noted that if $\|X\| = \|X^*\|$ for all $X \in \mathbb{C}^{n\times n}$, then $\|U\|_\ell = \|U^*\|_r$ and $\|U^*\|_\ell = \|U\|_r$. And if $\|\cdot\|$ is a unitarily invariant norm, then clearly $\|U\|_\ell = \|U\|_r = 1$.

Equipped with these estimations, we may derive a result analogous to Proposition 4.2 that equally holds for unitarily invariant and non-unitarily invariant norms. The price to pay for having greater generality is that we no longer have an exact expression for $\|p(A, A^*)\|$, but rather some lower and upper bounds.

Proposition 4.4. *Let $\|\cdot\|$ be a norm on $\mathbb{C}^{n\times n}$ which satisfies* (4.1) *and let $A \in \mathbb{C}^{n\times n}$ be a normal matrix which admits the decomposition $U\,\mathrm{diag}\big(\lambda_1, \ldots, \lambda_n\big)U^*$. Then*

$$\|p(A, A^*)\| \ \leq \ \|U\|_\ell \|U^*\|_r \max\big\{|p(\lambda_1, \overline{\lambda_1})|, \ldots, |p(\lambda_n, \overline{\lambda_n})|\big\}$$

and

$$\|p(A, A^*)\| \ \geq \ \frac{1}{\|U\|_r \|U^*\|_\ell} \max\big\{|p(\lambda_1, \overline{\lambda_1})|, \ldots, |p(\lambda_n, \overline{\lambda_n})|\big\}.$$

Proof. The proof is based on a judicious use of estimations developed above. First, by (4.2),

$$p(A, A^*) \ = \ Up(\Lambda, \Lambda^*)U \ = \ U\mathrm{diag}\big(p(\lambda_1, \overline{\lambda_1}), \ldots, p(\lambda_n, \overline{\lambda_n})\big)U^*.$$

Then by (4.5) (applied for U) and (4.6) (applied for U^*), we obtain

$$\|p(A, A^*)\| \ \leq \ \|U\|_\ell \|U^*\|_r \|\mathrm{diag}\big(p(\lambda_1, \overline{\lambda_1}), \ldots, p(\lambda_n, \overline{\lambda_n})\big)\|.$$

The result now follows upon using (4.1). The other part is proved similarly. □

Note that the decomposition $A = U\mathrm{diag}\big(\lambda_1, \ldots, \lambda_n\big)U^*$ is not unique, and some decomposition may lead to more precise estimations than others in Proposition 4.4. It is easily verified that these estimations are sharp in general since we can replace both inequalities by an equality when $\|\cdot\|$ is the spectral norm (see Proposition 4.2). Moreover, it is trivial to show that both estimations becomes an equality if A is a diagonal matrix. However, it is unknown in general when it is possible to replace the inequalities of Proposition 4.4 by equalities.

Open question 4.5. *Characterize the cases of equality in both estimations of Proposition 4.4.*

Remark. In (4.3) and (4.4), we defined $\|U\|_\ell$ and $\|U\|_r$ as

$$\sup\left\{\frac{\|UX\|}{\|X\|} : X \in \mathbb{C}^{n\times n}, X \neq \mathbf{0}\right\}, \qquad \& \qquad \sup\left\{\frac{\|XU\|}{\|X\|} : X \in \mathbb{C}^{n\times n}, X \neq \mathbf{0}\right\},$$

respectively. These definitions were deemed natural, since we considered the matrix U *acting as an operator on* $\mathbb{C}^{n\times n}$ by left or right multiplication. We then restricted ourselves to the cases where U is a unitary matrix. With this assumption, it is in many ways more natural to view U as an operator acting on the unitary group $\mathcal{U}(n)$, i.e., the group of $n \times n$ unitary matrices. The associated norms would thus be

$$\|U\|_{\ell *} \;:=\; \sup\left\{\frac{\|UX\|}{\|X\|} : X \in \mathcal{U}_n\right\}$$

and

$$\|U\|_{r *} \;:=\; \sup\left\{\frac{\|XU\|}{\|X\|} : X \in \mathcal{U}_n\right\}.$$

Of course, these norms are not appropriate for our study and they do not help us obtain a better result than Proposition 4.4. Nonetheless, we propose the following question which is interesting in its own right.

Open question 4.6. *Find necessary and sufficient conditions such that* $\|U\|_{\ell *} = \|U\|_\ell$.

§ 4.5. Application to circulant matrices

With the help of the tools presented in the previous subsections, we now turn our attention to the circulant matrices. Let A be an arbitrary circulant matrix and recall the fundamental representation (3.4) of A, i.e.,

$$A \;=\; W_n \operatorname{diag}\big(\mathcal{P}_A(1), \mathcal{P}_A(\omega_n), \mathcal{P}_A(\omega_n^2), \ldots, \mathcal{P}_A(\omega_n^{n-1})\big)\, W_n^*,$$

where W_n is the adjoint of the Discrete Fourier Transform matrix and $\mathcal{P}_A$ is the associated polynomial of A. Then, by Proposition 4.4,

$$\|A\| \;=\; \|\mathcal{P}_A(C_n)\| \;\leq\; \|W_n\|_\ell \|W_n^*\|_r \max\big\{|\mathcal{P}_A(1)|, \ldots, |\mathcal{P}_A(\omega_n^{n-1})|\big\},$$

and

$$\|A\| \;=\; \|\mathcal{P}_A(C_n)\| \;\geq\; \frac{1}{\|W_n\|_r \|W_n^*\|_\ell} \max\big\{|\mathcal{P}_A(1)|, \ldots, |\mathcal{P}_A(\omega_n^{n-1})|\big\}.$$

In particular, if the norm is unitarily invariant (i.e., if it is the spectral norm) then

$$\|A\|_{2\to 2} \;=\; \max\big\{|\mathcal{P}_A(1)|,\dots,|\mathcal{P}_A(\omega_n^{n-1})|\big\}, \tag{4.9}$$

which corresponds to a special case of Proposition 4.2. More generally, if the matrix norm is the induced p-norm, then Proposition 4.3 reveals that the above estimations reduce to

$$\|A\|_{p\to p} \;\leq\; \|W_n\|_{p\to p}\|W_n\|_{q\to q} \max_{1\leq k\leq n} |\mathcal{P}_A(\omega_n^k)| \tag{4.10}$$

and

$$\|A\|_{p\to p} \;\geq\; \frac{1}{\|W_n\|_{p\to p}\|W_n\|_{q\to q}} \max_{1\leq k\leq n} |\mathcal{P}_A(\omega_n^k)|. \tag{4.11}$$

It is not immediately clear how precise these estimations are and naturally wonders if they are sharper than those presented in Theorem 3.7, in the particular case where $A = A(n,-a,b)$. To answer this question, we first need to compute the value of $\|W_n\|_{p\to p}$.

Lemma 4.7. *Let W_n be defined as in* (3.3) *and assume that $1\leq p\leq\infty$. Then*

$$\|W_n\|_{p\to p} = \|W_n^*\|_{p\to p} = n^{|1/p-1/2|}.$$

Proof. Let us suppose that $1\leq p\leq 2$. Since W_n is unitary, we have

$$\|W_n x\|_p \;\leq\; n^{\frac{1}{p}-\frac{1}{2}}\|W_n x\|_2 \;=\; n^{\frac{1}{p}-\frac{1}{2}}\|x\|_2 \;\leq\; n^{\frac{1}{p}-\frac{1}{2}}\|x\|_p. \tag{4.12}$$

It thus follows that $\|W_n\|_{p\to p}\leq n^{1/p-1/2}$.

Considering the vector $x=(1,0,\dots,0)^\intercal$, we have

$$\frac{\|W_n x\|_p^p}{\|x\|_p^p} \;=\; \frac{\sum_{i=0}^{n-1}\left|\sum_{j=0}^{n-1} x_j\omega^{-ij}\right|^p}{n^{p/2}\sum_{i=0}^{n-1}|x_i|^p} \;=\; \frac{\sum_{i=0}^{n-1}|1|^p}{n^{p/2}} \;=\; \frac{n}{n^{p/2}} \;=\; n^{1-\frac{p}{2}}.$$

Hence, we conclude $\|W_n\|_{p\to p}=n^{1/p-1/2}$. The same method can then be used to show that $\|W_n^*\|_{p\to p}=n^{\frac{1}{p}-\frac{1}{2}}$. Therefore, the conclusion follows after a bit of simplification by noticing that

$$\|W_n^*\|_{p\to p} \;=\; \|W_n\|_{q\to q} \qquad \& \qquad \|W_n\|_{p\to p} \;=\; \|W_n^*\|_{q\to q},$$

where q denote the Hölder conjugate of p. □

Remark. Let c_p and C_p be the the best constants such that

$$c_p\|x\|_p \;\leq\; \|Ux\|_p \;\leq\; C_p\|x\|_p, \tag{4.13}$$

where U varies over all unitary matrices and $x \in \mathbb{C}^n$. It is not difficult to show that we have $c_p C_p = 1$. Also, observe that (4.12) and its analog for $2 \leq p \leq \infty$ are also valid for any unitary matrix U. In particular, we have

$$\|Ux\|_p \;\leq\; n^{\left|\frac{1}{p}-\frac{1}{2}\right|}\|x\|_p$$

for any $n \times n$ unitary matrix U and $x \in \mathbb{C}^n$. Moreover, we have shown in the previous proof that there exist some $x \in \mathbb{C}^n$ such that $\|W_n x\|_p = n^{|1/p-1/2|}\|x\|_p$. Therefore, we cannot replace $n^{|1/p-1/2|}$ by a smaller constant in the inequality $\|Ux\|_p \leq C_p\|x\|_p$ and thus, we have shown that $C_p = n^{|1/p-1/2|} = c_p^{-1}$.

Now, we know from (4.9) that $\max_{1\leq k\leq n} |\mathcal{P}_A(\omega_n^k)| = \|A\|_{2\to 2}$ and Lemma 4.7 ensures that $\|W_n\|_p = n^{|1/p-1/2|}$. Therefore, we see that (4.10) and (4.11) become

$$n^{-\left|\frac{2}{p}-1\right|}\|A\|_{2\to 2} \;\leq\; \|A\|_{p\to p} \;\leq\; n^{\left|\frac{2}{p}-1\right|}\|A\|_{2\to 2}.$$

Compare this to the upper bound

$$\|A\|_{p\to p} \;\leq\; n^{\left|\frac{1}{p}-\frac{1}{2}\right|}\|A\|_{2\to 2}.$$

of Theorem 3.6. While the form is obviously very similar, it is interesting to note that their derivation was quite different. In fact, it is an easy exercise to show that for every $p \in [1, \infty]$, the latter (i.e., the upper bound obtained by interpolation) is sharper. Hence, the new approach detailed in this section, despite being novel and general, does not improve the previous estimations of the induced p-norm of $A = A(n, -a, b)$ for $a, b \geq 0$.

As for the lower bound, remark that since the induced p-norm is a matrix norm, we have $\rho(A) \leq \|A\|_{p\to p}$ for every $A \in \mathbb{C}^{n\times n}$. Moreover, notice that the $|\mathcal{P}_A(\omega_n^k)|$'s are in fact the eigenvalues of A. Thus, it follows that

$$n^{-\left|\frac{2}{p}-1\right|}\|A\|_{2\to 2} \;=\; n^{-\left|\frac{2}{p}-1\right|} \max_{1\leq k\leq n} |\mathcal{P}_A(\omega_n^k)| \;=\; n^{-\left|\frac{2}{p}-1\right|}\rho(A).$$

Hence, we easily see that the lower bound is not sharp, except if $n^{-\left|\frac{1}{p}-\frac{1}{2}\right|} = 1$. As $n \geq 2$, this only happens if $p = 2$.

§ 5. An optimization-oriented approach

In this section, we discuss some of our partial findings relative to the Open question 3.8: *Is it possible to provide a simple closed formula for* $\|A(n, -a, b)\|_{p\to p}$*?* Here, we adopt a more direct approach using essentially tools from elementary calculus. Accordingly, the results obtained are more limited in scope, and their proofs are quite

technical in nature. Although, once more, we derive some results that cover a wide scope, the key difference is that we eventually focus our attention on the particular case of $\|A(n,-a,b)\|_{p\to p}$.

§ 5.1. A more precise lower bound

We begin with the following proposition, which provides a more refined lower bound for $\|A(n,-a,b)\|_{p\to p}$. This estimate will be used later on.

Proposition 5.1. *Let $a,b\geq 0$, and let $A=A(n,-a,b)$ be given by (2.1). Assume $1<p<\infty$. Then*

$$\|A\|_{p\to p}^p \;\geq\; \frac{\left|a+b\,(n-1)^{\frac{2-p}{1-p}}\right|^p+(n-1)^{\frac{1}{1-p}}\left|a+b\left((n-1)^{\frac{1}{p-1}}-n+2\right)\right|^p}{1+(n-1)^{\frac{1}{1-p}}}.$$

In particular, we also have

$$\|A\|_{p\to p} \;\geq\; a+b,$$

with equality if and only if either $n=2$, or $(n-2)b\leq 2a$ and $p=2$.

Proof. Fix $\eta:=n-1$ and $\rho:=\frac{1}{p-1}$. Consider the vector $x=(-\eta^\rho,1,1,\dots,1)^\intercal$. A simple computation gives

$$\frac{\|Ax\|_p^p}{\|x\|_p^p} \;=\; \frac{\left|a+b\eta^{1-\rho}\right|^p+\eta^{-\rho}\left|a+b\,(\eta^\rho-\eta+1)\right|^p}{1+\eta^{-\rho}},$$

proving the first part of the statement. Now, let $f(z):=|z|^p$ and let $t:=(1+\eta^{-\rho})^{-1}$. Then we can rewrite the above equation as

$$\frac{\|Ax\|_p^p}{\|x\|_p^p} \;=\; tf\left(a+b\eta^{1-\rho}\right)+(1-t)f\left(a+b\,(\eta^\rho-\eta+1)\right).$$

Hence, the convexity of f ensures us that

$$\begin{aligned}\frac{\|Ax\|_p^p}{\|x\|_p^p} &\geq f\left(t\left(a+b\eta^{1-\rho}\right)+(1-t)\left(a+b\,(\eta^\rho-\eta+1)\right)\right)\\ &= f\left(a+bt\left(\eta^{-\rho}+1\right)(\eta-\eta^\rho)+b\,(\eta^\rho-\eta+1)\right)\\ &= f\left(a+b\,(\eta-\eta^\rho)+b\,(\eta^\rho-\eta+1)\right)\\ &= f\left(a+b\right)\\ &= (a+b)^p.\end{aligned}$$

It follows that $\|A\|_{p\to p}\geq a+b$. Moreover, since f is never linear, the inequality is in fact an equality if and only if $a+b\eta^{1-\rho}=a+b\,(\eta^\rho-\eta+1)$. This occurs if and only if $(\eta^\rho+1)\left(\eta^{1-\rho}-1\right)=0$, if and only if $n=2$ or $p=2$. It is trivial to verify that $\|A\|_{p\to p}=a+b$ for every $p\in[1,\infty]$ if $n=2$ and we know from Theorem 3.2 that $\|A\|_{2\to 2}=a+b$ if and only if $(n-2)b\leq 2a$. This concludes the proof. □

Remark. If p tends to 1 (resp. to ∞), then the inequality given in Proposition 5.1 becomes $a + (n-1)b$, which is precisely the value of $\|A\|_{1\to 1}$ (resp. of $\|A\|_{\infty\to\infty}$).

As usual, let $a, b \geq 0$, and let $A = A(n, -a, b)$ be given by (2.1). The maximizing real vectors of $\frac{\|Ax\|_p}{\|x\|_p}$ are the vectors x for which

$$\frac{\|Ax\|_p}{\|x\|_p} \;=\; \|A\|_{p\to p}.$$

As a direct application to the previous proposition, we obtain the following result. This corollary will be used in the following section, which provides further details on the maximizing real vectors of $\frac{\|Ax\|_p}{\|x\|_p}$.

Corollary 5.2. *Let $a, b \geq 0$, and let $A = A(n, -a, b)$ be given by (2.1). Assume that $x \in \mathbb{R}^n$ is a maximizing real vector for $\frac{\|Ax\|_p}{\|x\|_p}$ which satisfies $x_1 + \cdots + x_n = 0$. Then either $n = 2$, or $(n-2)b \leq 2a$ and $p = 2$.*

Proof. If $x_1 + \cdots + x_n = 0$, a direct computation reveals that $\frac{\|Ax\|_p}{\|x\|_p} = a + b$. Hence, since x is a maximizing real vector for $\frac{\|Ax\|_p}{\|x\|_p}$, we find $\|A\|_{p\to p} = a + b$. Proposition 5.1 then ensures that this occurs if and only if either $n = 2$, or $(n-2)b \leq 2a$ and $p = 2$. □

§ 5.2. The maximizing vectors for $\frac{\|Ax\|_p}{\|x\|_p}$

Let $a, b \geq 0$, and let $A = A(n, -a, b)$ be given by (2.1). We now focus on the maximizing real vectors for $\frac{\|Ax\|_p}{\|x\|_p}$. It is proved in [8] that, in the particular case where $p = 2$ and $(n-2)b \leq 2a$, the maximizing real unit vectors are those whose entries verify $x_1 + x_2 + \cdots + x_n = 0$. Whereas in the case where $p = 2$ but $(n-2)b \geq 2a$, the maximizing real unit vectors are $x = \pm\frac{1}{\sqrt{n}}(1, 1, \ldots, 1)^\intercal$.

Finding the maximizing real unit vectors for $\frac{\|Ax\|_p}{\|x\|_p}$ when $1 \leq p \leq \infty$ and $p \neq 2$ is a daunting task. However, in what follows, we show that the entries of a given maximizing real vector always form a set of cardinality at most *three*. The proof of this result is somewhat convoluted. Prior to presenting a detailed demonstration, we need to establish the following two technical lemmas. We also need to define $x^{[p]} := \operatorname{sgn}(x)|x|^p = x|x|^{p-1}$, which is the derivative of $\frac{1}{p+1}|x|^{p+1}$ when $1 < p < \infty$.

Lemma 5.3. *Assume $1 \leq p \leq \infty$. Let $a, b, c_1, c_2, d \in \mathbb{R}$, and consider the real-valued function*

$$f(x) \;:=\; a\,(c_1 x - c_2)^{[p-1]} - b x^{[p-1]} + d.$$

Then either f is identically zero or it has at most *three distinct roots.*

Proof. Suppose that f is not identically zero. A computation reveals that

$$f'(x) \;=\; (p-1)\left(ac_1\,|c_1x-c_2|^{p-2}-b\,|x|^{p-2}\right).$$

If $abc_1 \leq 0$, then f' is either always non-negative or always non-positive. In the special case where $f' \equiv 0$, we find that f has no roots since $f \not\equiv 0$. Otherwise, f is either strictly increasing or strictly decreasing, and thus f has exactly one root.

If $abc_1 > 0$, then f' changes sign precisely at the following *two* points:

$$x_1 \;=\; \frac{\left(\frac{ac_1}{b}\right)^{\frac{1}{p-2}}c_2}{\left(\frac{ac_1}{b}\right)^{\frac{1}{p-2}}c_1+1} \qquad \& \qquad x_2 \;=\; \frac{\left(\frac{ac_1}{b}\right)^{\frac{1}{p-2}}c_2}{\left(\frac{ac_1}{b}\right)^{\frac{1}{p-2}}c_1-1}.$$

Hence, f is decreasing on $(-\infty, \min\{x_1,x_2\})$, increasing on $(\min\{x_1,x_2\}, \max\{x_1,x_2\})$ and decreasing on $(\max\{x_1,x_2\},\infty)$, or *vice versa*. It follows immediately that f has at most three distinct roots, as stated. □

Lemma 5.4. *Let $a,b \geq 0$, with at least one of them non-zero. Let $A = A(n,-a,b)$ be given by (2.1). Suppose $n > 2$. Assume $1 < p < \infty$. If $x \in \mathbb{R}^n$ is a maximizing real vector for $\frac{\|Ax\|_p}{\|x\|_p}$, then, for every $1 \leq j,k,l \leq n$,*

$$\left(w_k^{[p-1]} - w_j^{[p-1]}\right)\left(x_k^{[p-1]} - x_l^{[p-1]}\right) \;=\; \left(w_k^{[p-1]} - w_l^{[p-1]}\right)\left(x_k^{[p-1]} - x_j^{[p-1]}\right) \tag{5.1}$$

where $w_i = (a+b)x_i - b(x_1+x_2+\cdots+x_n)$.

Proof. We have

$$\begin{aligned}\|A\|_{p\to p}^p \;&=\; \sup_{\|x\|_p=1}\|Ax\|_p^p \\ &=\; \sup_{\|x\|_p=1}\sum_{i=1}^n |(a+b)x_i - b(x_1+\cdots+x_n)|^p \\ &=\; \sup_{\|x\|_p=1}\sum_{i=1}^n |w_i|^p.\end{aligned}$$

Since the unit circle is compact, we can define the Lagrange multiplier

$$\mathcal{L}(x,\lambda) \;:=\; \|Ax\|_p^p - \lambda\left(\|x\|_p^p - 1\right)$$

and since for $1 < p < \infty$, the derivative of $|x|^p$ exist everywhere, we find that the maximum of $\|Ax\|_p^p$ on the set of unit vectors is obtained if and only if we have $\frac{\partial \mathcal{L}}{\partial \lambda} = 0$ and $\frac{\partial \mathcal{L}}{\partial x_k} = 0$, for every $k \in \{1,2,\ldots,n\}$. It is then easily verified that we have $\frac{\partial \mathcal{L}}{\partial x_k} = 0$ if and only if

$$(a+b)\,w_k^{[p-1]} - \sum_{i=1}^n b w_i^{[p-1]} \;=\; \lambda x_k^{[p-1]}. \tag{5.2}$$

Now, if any two of x_j, x_l and x_k are equal ($1 \leq j,k,l \leq n$) equation (5.1) is directly satisfied. Otherwise, choose any j,k,l such that $x_j \neq x_k \neq x_l$. Subtracting the equation (5.2) associated with the coefficient j (resp. l) to the one associated with the coefficient k and simplifying, we get

$$\frac{w_k^{[p-1]} - w_j^{[p-1]}}{x_k^{[p-1]} - x_j^{[p-1]}} \;=\; \frac{\lambda}{a+b} \qquad \& \qquad \frac{w_k^{[p-1]} - w_l^{[p-1]}}{x_k^{[p-1]} - x_l^{[p-1]}} \;=\; \frac{\lambda}{a+b}.$$

Note that we can safely divide by $a+b$ since $a,b \geq 0$, with at least one of them non-zero. Hence, equalling the left hand side of both equations and simplifying yield

$$\big(w_k^{[p-1]} - w_j^{[p-1]}\big)\big(x_k^{[p-1]} - x_l^{[p-1]}\big) \;=\; \big(w_k^{[p-1]} - w_l^{[p-1]}\big)\big(x_k^{[p-1]} - x_j^{[p-1]}\big)$$

and we are done. □

We now have everything in hand to demonstrate that the following holds true.

Proposition 5.5. *Let $a,b \geq 0$, with at least one of them non-zero. Let $A = A(n,-a,b)$ be given by (2.1). Suppose $n > 2$. Assume $1 < p < \infty$, with $p \neq 2$. If $x \in \mathbb{R}^n$ is a maximizing real vector for $\frac{\|Ax\|_p}{\|x\|_p}$, then the entries of x form a set of cardinality at most* three.

Proof. Suppose without any loss of generality, even if it means rearranging the coefficients of x, that $x_1 \neq x_2$ and for simplicity, define $\rho := p-1$. Then Lemma 5.4 ensures us that

$$\big(w_k^{[\rho]} - w_1^{[\rho]}\big)\big(x_k^{[\rho]} - x_2^{[\rho]}\big) \;=\; \big(w_k^{[\rho]} - w_2^{[\rho]}\big)\big(x_k^{[\rho]} - x_1^{[\rho]}\big)$$

for any k, where $w_i = (a+b)x_i - b(x_1 + x_2 + \cdots + x_n)$. Since $n > 2$ and $p \neq 2$, Corollary 5.2 ensures us that $x_1 + \cdots + x_n \neq 0$. Hence, we can define $y_i := x_i/(x_1 + \cdots + x_n)$ and $z_i := w_i/(x_1 + \cdots + x_n) = (a+b)y_i - b$. A simple division by $x_1 + \cdots + x_n \neq 0$ then allows us to show that

$$\big(z_k^{[\rho]} - z_1^{[\rho]}\big)\big(y_k^{[\rho]} - y_2^{[\rho]}\big) \;=\; \big(z_k^{[\rho]} - z_2^{[\rho]}\big)\big(y_k^{[\rho]} - y_1^{[\rho]}\big), \qquad (1 \leq k \leq n).$$

This can be rewritten using the function

$$f(x) \;:=\; \big(y_2^{[\rho]} - y_1^{[\rho]}\big)\big((a+b)x - b\big)^{[\rho]} + \big(z_1^{[\rho]} - z_2^{[\rho]}\big)x^{[\rho]} + \big(y_1^{[\rho]} z_2^{[\rho]} - y_2^{[\rho]} z_1^{[\rho]}\big) \tag{5.3}$$

as $f(y_k) = 0$ for every $1 \leq k \leq n$. Now, let $\alpha := y_1$ and $\beta := y_2$. An application of Lemma 5.3 reveals that the function f has *at most* three distinct roots. Moreover, we easily verify that $f(\alpha) = f(\beta) = 0$. Therefore, the roots of f are α, β and possibly another value, call it γ. Since $f(y_k) = 0$ for every $1 \leq k \leq n$, it follows that for any $1 \leq k \leq n$, we either have $y_k = \alpha$, β or γ. Hence, y_k can take at most three distinct values and thus, x_k can also take at most three distinct values. □

Remark. Note that, even though the proof given in the previous proposition does not apply for $p = 1, 2, \infty$, it is not actually a problem since the norm of A for all of these values are known and are almost trivial problems. Moreover, in all of these cases (with the exception of $p = 2$ and $(n-2)b \leq 2a$), the entries of the maximizing vectors actually only have at most two distinct values.

If x is a maximizing real vector for $\frac{\|Ax\|_p}{\|x\|_p}$, then Proposition 5.5 tells us that x has at most three distinct coefficients. Suppose that x does indeed have exactly three distinct coefficient and define once again $y_i := x_i/(x_1 + \cdots + x_n)$ and $z_i := w_i/(x_1 + \cdots + x_n) = ((a+b)y_i - b)$. Moreover, assume that two of the three distinct coefficients of y are α, occurring m times, and β, occurring k times. Then the third distinct coefficient of y, say γ, is precisely the unique root of f which is not equal to α or β, where f is given by (5.3). Moreover, by construction, the sum of the coefficients of y must be equal to 1, i.e., they must satisfy the equation $m\alpha + k\beta + (n-m-k)\gamma = 1$ and thus we also find that $\gamma = \frac{1-m\alpha-k\beta}{n-m-k}$. Remark that this is two completely different way of determining the value of γ. However, it is natural to assume that the two functions determining the value of γ are independent and thus that we have no reason to expect the two methods to give the same result. This observation motivated the following conjecture, which is strongly backed up with numerical evidence for several values of a, b, n and p.

Conjecture 5.6. *Let $a, b \geq 0$, with at least one of them non-zero. Let $A = A(n, -a, b)$ be given by (2.1). Assume $1 \leq p \leq \infty$, with $p \neq 2$. If $x \in \mathbb{R}^n$ is a maximizing real vector for $\frac{\|Ax\|_p}{\|x\|_p}$, then the entries of x form a set of cardinality at most two.*

§ 5.3. The special case of $I - \frac{1}{n}K$

Our interest in the matrices $A(n, a, b)$ stems from other studies on the geometry of the Birkhoff polytope and originated with the matrix $I - \frac{1}{n}K$ which corresponds to the matrix $-A(n, \frac{1-n}{n}, \frac{1}{n})$. Here, we provide a description of $\|I - \frac{1}{n}K\|_{p\to p}$, assuming the validity of Conjecture 5.6. We then show that the conjecture is valid in the special case $n = 3$.

Proposition 5.7. *Let $1 \leq p \leq \infty$, with $p \neq 2$. Let x_p be the unique root of the function*

$$x \longmapsto (p-1)\left(1 + (x-1)^{\frac{1}{p-1}}\right)\left(1 - (x-1)^{p-2}\right) + \left(1 - (x-1)^{\frac{2-p}{p-1}}\right)\left(1 + (x-1)^{p-1}\right)$$

in the interval $[1, 2]$, and let $m_1 := \left\lfloor \frac{n}{x_p} \right\rfloor$ and $m_2 := \left\lceil \frac{n}{x_p} \right\rceil$. Suppose that Conjecture 5.6 is valid. Then

$$\left\|I - \tfrac{1}{n}K\right\|_{p\to p} = \max_{m \in \{m_1, m_2\}} \frac{\left(\left(\frac{n}{m} - 1\right)^{p-1} + 1\right)^{\frac{1}{p}} \left(\left(\frac{n}{m} - 1\right)^{\frac{1}{p-1}} + 1\right)^{1-\frac{1}{p}}}{\frac{n}{m}}. \tag{5.4}$$

Proof. Let x be the maximizing real vector with coefficients α, occurring m times, and β, occurring $(n-m)$ times. We know from Corollary 5.2 that the sum of the coefficients of x cannot be 0. Hence, without any loss of generality, suppose that $m\alpha + (n-m)\beta = n$, i.e., that $\beta = \frac{n-m\alpha}{n-m}$. In this case, a direct computation reveals that

$$\frac{\|(I - \frac{1}{n}K)x\|_p^p}{\|x\|_p^p} \;=\; \frac{m\,(\alpha-1)^p + (n-m)\left(\frac{m-m\alpha}{n-m}\right)^p}{m\alpha^p + (n-m)\left(\frac{n-m\alpha}{n-m}\right)^p} \;=:\; f_m(\alpha).$$

Taking the derivative relative to α, we get

$$f_m'(\alpha) \;=\; 0 \quad\Longleftrightarrow\quad \alpha \;=\; \frac{n\,(n-m)^{\frac{2-p}{p-1}}}{m\,(n-m)^{\frac{2-p}{p-1}} - m^{\frac{1}{p-1}}}, \tag{5.5}$$

and we easily verify that this critical point is, in fact, a local maximum. Putting this value in $f_m(\alpha)$ and simplifying, we get

$$f_m(\alpha) \;=\; \begin{cases} \dfrac{\left(\left(\frac{n}{m}-1\right)^{p-1}+1\right)\left(\left(\frac{n}{m}-1\right)^{\frac{1}{p-1}}+1\right)^{p-1}}{\left(\frac{n}{m}\right)^p} & \text{if} \quad m \neq 0, \\[2ex] 1 & \text{if} \quad m = 0. \end{cases}$$

Therefore,

$$\left\|I - \tfrac{1}{n}K\right\|_{p\to p}^p \;=\; \max_{1\le m\le n} \frac{\left(\left(\frac{n}{m}-1\right)^{p-1}+1\right)\left(\left(\frac{n}{m}-1\right)^{\frac{1}{p-1}}+1\right)^{p-1}}{\left(\frac{n}{m}\right)^p}. \tag{5.6}$$

Remark that, since α and β are interchangeable, we can consider only the maximum for $\frac{n}{2} \le m \le n$. Moreover, defining

$$g(x) \;:=\; \frac{\left((x-1)^{p-1}+1\right)\left((x-1)^{\frac{1}{p-1}}+1\right)^{p-1}}{x^p},$$

we have that the right hand side of (5.6) correspond to $\max_{1\le m\le n} g(n/m)$. Since $\frac{n}{2} \le m \le n$, finding the maximum of $g(n/m)$ is closely related to finding the maximum of g in $[1,2]$. Let x_p be the unique root of the function

$$x \longmapsto (p-1)\left(1+(x-1)^{\frac{1}{p-1}}\right)\left(1-(x-1)^{p-2}\right) + \left(1-(x-1)^{\frac{2-p}{p-1}}\right)\left(1+(x-1)^{p-1}\right)$$

in $[1,2]$. Then, an analysis of g reveal that it is increasing on $[1,x_p]$ and decreasing on $[x_p,2]$. Therefore, the value of $m \ge \frac{n}{2}$ for which the maximum of $g(n/m)$ is attained must be *one of the two closest* values of n/m to x_p. More precisely, it can be verified that it must be either $\left\lfloor \frac{n}{x_p} \right\rfloor$ or $\left\lceil \frac{n}{x_p} \right\rceil$. □

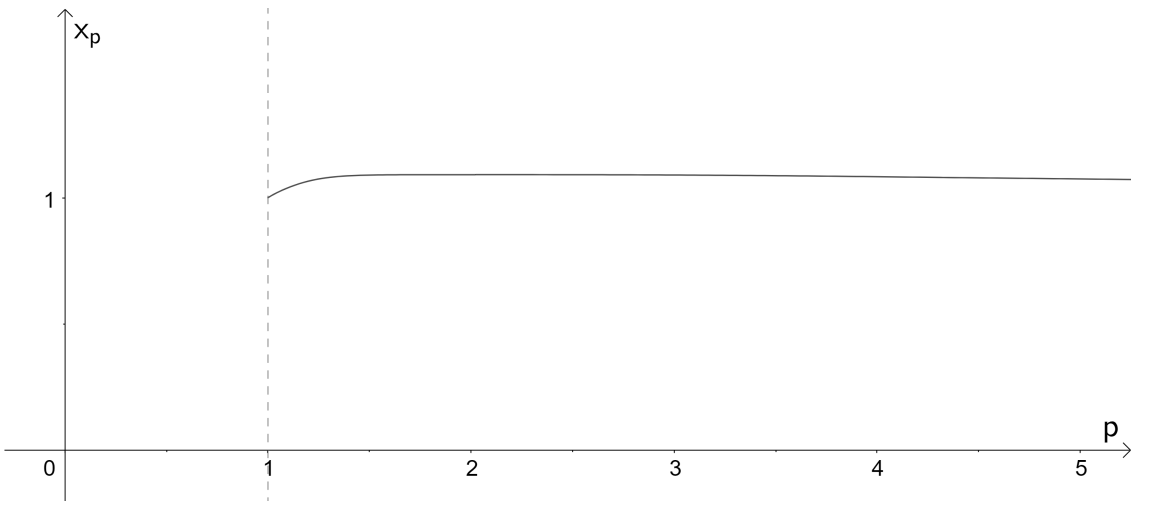


Figure 2: Values of x_p between $p = 1$ and $p = 5.25$.

Despite its complicated description, we may use various methods to obtain approximations of x_p and evaluate it numerically. It is also possible to show that x_p varies between 1 (when $p \to 1, \infty$) and ≈ 1.090776, which is the unique root of $\ln(x-1) - \frac{2x}{x-2}$ in $[1, 2]$ (when $p \to 2$). This observation alone can greatly reduce the number of cases needed to verify in (5.6) to compute the norm of $I - \frac{1}{n}K$ (still assuming the validity of Conjecture 5.6), even without precisely knowing x_p for every value of the parameter p.

§ 5.4. A special case of the Conjecture 5.6

We now turn our attention to proving that Conjecture 5.6 is valid in the particular case $n = 3$. To do this, we first need to prove the following two technical lemmas.

Lemma 5.8. *Let $J(t) := \left(t + \frac{\gamma-1}{3}\right)^{p-1} - t^{p-1}$, where $t \in (\frac{1}{2}, 1]$ and $2^{\frac{2-p}{p-1}} \leq \gamma < 1$. If $p > 2$ is a rational number with an even numerator and an odd denominator, then $J(t) < J(1-t)$.*

Proof. We have

$$J'(t) \;=\; (p-1)\left(\left(t + \frac{\gamma-1}{3}\right)^{p-2} - t^{p-2}\right).$$

Therefore, $J'(t) < 0$ if and only $\left(t + \frac{\gamma-1}{3}\right)^{p-2} < t^{p-2}$. Because of our hypothesis on the form of p, this inequality is satisfied if and only if $\left|t + \frac{\gamma-1}{3}\right| < |t|$, which is equivalent to

$$0 \;<\; \frac{1-\gamma}{6} \;<\; t.$$

Hence, since $0 < \frac{1-\gamma}{6}$, we find that $J(t)$ is strictly decreasing on $(\frac{1-\gamma}{6}, 1]$ and strictly increasing on $[0, \frac{1-\gamma}{6})$. In particular, since

$$\frac{1-\gamma}{6} \leq \frac{1-1/2}{6} < \frac{1}{2},$$

$J(t)$ is strictly decreasing for $t \in (\frac{1}{2}, 1]$. Therefore, for each $t \in (\frac{1}{2}, 1]$, we have $J(t) < J(\frac{1}{2})$ and $\min\{J(\frac{1}{2}), J(0)\} \leq J(1-t)$. Now, notice that $J(\frac{1}{2}) \leq J(0)$ if and only if

$$\left(\frac{2\gamma-2}{3}+1\right)^{p-1} - \left(\frac{2\gamma-2}{3}\right)^{p-1} \;\leq\; 1.$$

Setting $z := \frac{2\gamma-2}{3}$, we find that $J(1/2) \leq J(0)$ if and only if

$$(z+1)^{p-1} - z^{p-1} \;\leq\; 1.$$

It is then a simple matter to show that $(z+1)^{p-1} - z^{p-1}$ is increasing for every $z \geq -1/2$. Noticing that $\frac{-2}{3} \leq \frac{2\gamma-2}{3} < 0$, we find that $(z+1)^{p-1} - z^{p-1}$ attains its maximum at $z = 0$. Hence,

$$(z+1)^{p-1} - z^{p-1} \;\leq\; (0+1)^{p-1} - 0^{p-1} \;=\; 1,$$

and we indeed have $J(1/2) \leq J(0)$. Therefore, $\min\{J(1/2), J(0)\} = J(1/2)$ and we finally find, for $t \in (1/2, 1]$,

$$J(t) \;<\; J(1/2) \;=\; \min\{J(1/2), J(0)\} \;\leq\; J(1-t),$$

which is what we wanted to show. □

Lemma 5.9. *Let $t \in [0,1]$, $s = 2^{\frac{2-p}{p-1}}$, $\gamma \in [s, 1]$ and*

$$F_\gamma(t) \;:=\; \frac{\left(t+\frac{\gamma-1}{3}\right)^p + \left(1-t+\frac{\gamma-1}{3}\right)^p + \left(\frac{1+2\gamma}{3}\right)^p}{t^p + (1-t)^p + \gamma^p}. \tag{5.7}$$

If $p > 2$ is a rational number with an even numerator and an odd denominator, then

$$\max_{s\leq\gamma\leq 1} \max_{0\leq t\leq 1} F_\gamma(t) \;=\; \frac{\left(2^{p-1}+1\right)\left(2^{\frac{1}{p-1}}+1\right)^{p-1}}{3^p},$$

where the equality is attained only if $t = \frac{1}{2}$.

Proof. **Case 1 ($\gamma = 1$):** If $\gamma = 1$, then $F_\gamma(t) = 1$ for each $t \in [0,1]$.
Case 2 ($\gamma \neq 1$): Computing the derivative of F_γ, we find

$$F'_\gamma(t) \;=\; p\left(\frac{\left(t+\frac{\gamma-1}{3}\right)^{p-1} - \left(1-t+\frac{\gamma-1}{3}\right)^{p-1}}{t^p + (1-t)^p + \gamma^p} - F_\gamma(t)\frac{t^{p-1} - (1-t)^{p-1}}{t^p + (1-t)^p + \gamma^p}\right).$$

By direct verification, $F'_\gamma(t) = 0$ if $t = 1/2$. If $t \neq 1/2$, then $t^{p-1} - (1-t)^{p-1} \neq 0$ and we have

$$F'_\gamma(t) \;=\; p\frac{t^{p-1} - (1-t)^{p-1}}{t^p + (1-t)^p + \gamma^p}\left(\frac{\left(t + \frac{\gamma-1}{3}\right)^{p-1} - \left(1 - t + \frac{\gamma-1}{3}\right)^{p-1}}{t^{p-1} - (1-t)^{p-1}} - F_\gamma(t)\right),$$

which vanishes if and only if

$$\frac{\left(t + \frac{\gamma-1}{3}\right)^{p-1} - \left(1 - t + \frac{\gamma-1}{3}\right)^{p-1}}{t^{p-1} - (1-t)^{p-1}} \;=\; F_\gamma(t).$$

We now show that we always have

$$\frac{\left(t + \frac{\gamma-1}{3}\right)^{p-1} - \left(1 - t + \frac{\gamma-1}{3}\right)^{p-1}}{t^{p-1} - (1-t)^{p-1}} \;<\; 1. \tag{5.8}$$

It will then follow that if $F'_\gamma(t)$ vanishes at a point (γ, t), with $t \neq 1/2$, then we must have $F_\gamma(t) < 1$. Therefore, by Case 1, the point (γ, t) cannot be a maximum, and we will be able to reject these points.

First notice that by the symmetry along the $x = 1/2$ axis, we can suppose without any loss of generality that $1/2 < t \leq 1$ (since we supposed that $t \neq 1/2$). In that case, both

$$\left(t + \frac{\gamma-1}{3}\right)^{p-1} - \left(1 - t + \frac{\gamma-1}{3}\right)^{p-1} \quad \& \quad t^{p-1} - (1-t)^{p-1}$$

are positive and thus (5.8) is equivalent to having

$$\left(t + \frac{\gamma-1}{3}\right)^{p-1} - t^{p-1} \;<\; \left(1 - t + \frac{\gamma-1}{3}\right)^{p-1} - (1-t)^{p-1}, \tag{5.9}$$

for $1/2 < t \leq 1$. Now, if we define

$$J(t) := \left(t + \frac{\gamma-1}{3}\right)^{p-1} - t^{p-1},$$

the above inequality can be restated as $J(t) < J(1-t)$. Lemma 5.8 then ensures us that this inequality is satisfied and thus, (5.9) is valid and so is (5.8). Therefore, the only points (γ, t) which can give rise to a maximum of $F_\gamma(t)$ occurs when $t = \frac{1}{2}$. Moreover, it is easily verified that for a fixed $\gamma \in [s, 1]$, $t = \frac{1}{2}$ is indeed a maximum of F_γ, and in particular that $F_\gamma(t) < F_\gamma(1/2)$ for every $t \in [0, \frac{1}{2}) \cup (\frac{1}{2}, 1]$. Hence, for $t \in [0, 1]$ and $\gamma \neq 1$, we have

$$\begin{aligned}
\max_{s\leq\gamma<1}\max_{0\leq t\leq 1} F_\gamma(t) \;&=\; \max_{s\leq\gamma<1} F_\gamma(1/2) \\
&=\; \max_{s\leq\gamma<1} \frac{2\left(\frac{1}{2} + \frac{\gamma-1}{3}\right)^p + \left(\frac{1+2\gamma}{3}\right)^p}{2^{1-p} + \gamma^p} \\
&=\; \frac{2^{1-p}+1}{3^p}\max_{s\leq\gamma<1}\frac{(1+2\gamma)^p}{2^{1-p}+\gamma^p} \;=:\; \frac{2^{1-p}+1}{3^p}\max_{s\leq\gamma<1} H(\gamma),
\end{aligned}$$

where the first equality is attained if and only if $t = 1/2$.

Now, we have

$$H'(\gamma) \;=\; p\frac{\left(2^{2-p} - \gamma^{p-1}\right)(2\gamma+1)^{p-1}}{\left(2^{1-p}+\gamma^p\right)^2}.$$

Since γ is positive, we have $H'(\gamma) \le 0$ if and only if $2^{2-p} - \gamma^{p-1} \le 0$, i.e., if and only if $\gamma \ge 2^{\frac{2-p}{p-1}} = s$. Therefore, $H(\gamma)$ is decreasing for $s \le \gamma < 1$ and thus,

$$\begin{aligned}
\max_{s\le\gamma<1}\max_{0\le t\le 1} F_\gamma(t) \;&=\; \frac{2^{1-p}+1}{3^p}\max_{s\le\gamma<1} H(\gamma)\\
&=\; \frac{2^{1-p}+1}{3^p}H(s)\\
&=\; \frac{2^{1-p}+1}{3^p}\frac{(2s+1)^p}{2^{1-p}+s^p}\\
&=\; \frac{\left(2^{p-1}+1\right)\left(2^{\frac{1}{p-1}}+1\right)^{p-1}}{3^p},
\end{aligned}$$

where the equalities are attained simultaneously only if $t = 1/2$. Combining Case 1 and Case 2 and noticing that $\frac{\left(2^{p-1}+1\right)\left(2^{\frac{1}{p-1}}+1\right)^{p-1}}{3^p} \ge 1$ for every $p \in [1,\infty]$ yield the desired result. □

Equipped with these lemmas, we are now ready to prove that Conjecture 5.6 is true for $A = -A(3, -\frac{2}{3}, \frac{1}{3})$. The proof is in three parts. First, we show that if x is a critical point of the function $\frac{\|Ax\|_p}{\|x\|_p}$, then x must be of a particular form which only depends on one parameter t (up to a constant). Next, we show that if it is also a maximizing real vector, then $\frac{\|Ax\|_p}{\|x\|_p} \le F_\gamma(t)$, where $F_\gamma(t)$ is defined as in (5.7). We then use Lemma 5.9 to prove that the maximizing real vector of $\frac{\|Ax\|_p}{\|x\|_p}$ must have at most two distinct coefficients, and we provide the operator p-norm of A at the same time.

Theorem 5.10. *Let $A = -A\left(3, -\frac{2}{3}, \frac{1}{3}\right) = I_3 - \frac{1}{3}K_3$, and let $1 < p < \infty$ with $p \ne 2$. If $x \in \mathbb{R}^3$ is a maximizing real vector of the function $\frac{\|Ax\|_p}{\|x\|_p}$, then x has at most two distinct coefficients. Moreover,*

$$\|A\|_{p\to p} \;=\; \frac{\left(2^{p-1}+1\right)^{\frac{1}{p}}\left(2^{\frac{1}{p-1}}+1\right)^{1-\frac{1}{p}}}{3}. \tag{5.10}$$

Proof. Suppose without loss of generality that $p > 2$ is a rational number with an even numerator and an odd numerator. Let $x := (x_1, x_2, x_3)^\intercal$, $\mu := \frac{x_1+x_2+x_3}{3}$ and

$$f(x) \;:=\; \frac{\|(I_3 - \frac{1}{3}K_3)x\|_p^p}{\|x\|_p^p} \;=\; \frac{(x_1-\mu)^p + (x_2-\mu)^p + (x_3-\mu)^p}{x_1^p + x_2^p + x_3^p}.$$

If $f(x)$ is a maximum, then each partial derivative $\frac{\partial f}{\partial x_k}$ vanishes. Since $\|x\|_p \neq 0$, it is an easy exercise to verify that this occurs if and only if

$$(x_j-\mu)^{p-1} - \frac{(x_1-\mu)^{p-1}+(x_2-\mu)^{p-1}+(x_3-\mu)^{p-1}}{3} \;=\; x_j^{p-1} f(x),$$

for $j=1,2,3$. Summing these equations over each $j=1,2,3$, we get

$$f(x)(x_1^{p-1}+x_2^{p-1}+x_3^{p-1}) \;=\; 0. \tag{5.11}$$

Therefore, for x to be a maximum of f, we need to have $x_1^{p-1}+x_2^{p-1}+x_3^{p-1}=0$. Because of the hypothesis on the form of p, we can thus suppose without any loss of generality that $x_3 = -(x_1^{p-1}+x_2^{p-1})^{\frac{1}{p-1}}$.

It is obvious that x_1, x_2 and x_3 are not all of the same sign, otherwise we would not have $x_1^{p-1}+x_2^{p-1}+x_3^{p-1}=0$. Hence, without any loss of generality, suppose that $x_1, x_2 \geq 0$ and $x_3 \leq 0$ (if it is not the case, rearrange the index and/or take $-x$ instead of x). Notice that we then have $x_1+x_2>0$. Hence, we can furthermore suppose that $x_1+x_2=1$ (take $\tilde{x} := x/(x_1+x_2)$ instead of x), i.e., that $x_2 = 1-x_1$. Because of the positivity of x_1, x_2, it follows that $x_1 \in [0,1]$. Therefore, to find the maximum of $f(x)$, it suffice to consider the vectors of the form

$$x \;=\; \left(t, 1-t, -(t^{p-1}+(1-t)^{p-1})^{\frac{1}{p-1}}\right)^{\mathsf{T}}, \qquad (0 \leq t \leq 1).$$

For simplicity, we will write from now on (with a mild abuse of notation) $f(t)$ instead of $f(x)$. Moreover, we will also denote $(t^{p-1}+(1-t)^{p-1})^{\frac{1}{p-1}}$ by Γ. Note that Γ attains its maximum of 1 at $t=0,1$ and its minimum of $s := 2^{\frac{2-p}{p-1}}$ at $t=1/2$. Therefore, using Lemma 5.9, we have

$$\begin{aligned}
\max_{0\leq t\leq 1} f(t) \;&=\; \max_{0\leq t\leq 1} \frac{\left(t+\frac{\Gamma-1}{3}\right)^p + \left(1-t+\frac{\Gamma-1}{3}\right)^p + \left(\frac{1+2\Gamma}{3}\right)^p}{t^p+(1-t)^p+\Gamma^p} \\
&\leq\; \max_{s\leq\gamma\leq 1} \max_{0\leq t\leq 1} \frac{\left(t+\frac{\gamma-1}{3}\right)^p + \left(1-t+\frac{\gamma-1}{3}\right)^p + \left(\frac{1+2\gamma}{3}\right)^p}{t^p+(1-t)^p+\gamma^p} \\
&=:\; \max_{s\leq\gamma\leq 1} \max_{0\leq t\leq 1} F_\gamma(t) \;=\; \frac{\left(2^{p-1}+1\right)\left(2^{\frac{1}{p-1}}+1\right)^{p-1}}{3^p},
\end{aligned}$$

where the last equality is attained only if $t=\frac{1}{2}$. Furthermore, we have

$$f(1/2) \;=\; \frac{\left(2^{p-1}+1\right)\left(2^{\frac{1}{p-1}}+1\right)^{p-1}}{3^p}.$$

Hence, it follows that

$$\max_{0\leq t\leq 1} f(t) \;=\; \frac{\left(2^{p-1}+1\right)\left(2^{\frac{1}{p-1}}+1\right)^{p-1}}{3^p},$$

where the equality is attained only if $t = 1/2$. Therefore, the unique (up to a constant) maximizing real vector of $\frac{\|(I_3-\frac{1}{3}K_3)x\|_p}{\|x\|_p}$ is

$$x \;=\; \left(\tfrac{1}{2}, 1-\tfrac{1}{2}, -((\tfrac{1}{2})^{p-1}+(1-\tfrac{1}{2})^{p-1})^{\frac{1}{p-1}}\right)^{\mathsf{T}} \;=\; \left(\tfrac{1}{2}, \tfrac{1}{2}, -s\right)^{\mathsf{T}},$$

which has at most two distinct coefficient. Moreover, we also find that

$$\left\|I_3 - \tfrac{1}{3}K_3\right\|_{p\to p} \;=\; \frac{\left(2^{p-1}+1\right)^{\frac{1}{p}}\left(2^{\frac{1}{p-1}}+1\right)^{1-\frac{1}{p}}}{3},$$

which correspond to the cases $m = 1$ and $m = 2$ in (5.4). Finally, since we supposed that $p > 2$, we note that the result follows for $1 \le p \le 2$ because $I_3 - \frac{1}{3}K_3$ is self-adjoint. □

Note that an analogue of (5.11) actually holds true for any $n > 1$. This provides even further evidence for the validity of Conjecture 5.6, as it gives another independent method to determine the third value of the maximizing real vector x, given that the first two values are known. Finally, we note that what was shown in this section cannot readily be generalized to the case $A(n, -a, b)$. In particular, there is not an easy way to extend Proposition 5.7 to the general setting $A(n, -a, b)$. This is due to the fact that we cannot solve simply the analogue of the equation $f'_m(\alpha) = 0$ in (5.5) in the more general case. Hence, to provide a complete description of the operator p-norm of $A(n, -a, b)$ (assuming the validity of Conjecture 5.6), we will need some stronger tools and some new ideas.

§ 6. Concluding remarks

(i) As mentioned earlier, the particular case $A = -A(n, \frac{1-n}{n}, \frac{1}{n}) = I - \frac{1}{n}K$, the norm $\|A\|_{p\to p}$ represents the Chebyshev radius of the n-dimensional Birkhoff polytope. But there is also another description of $\|A\|_{p\to p}$ in this special case, which is closely related to various important measures of statistical dispersion or variability. Indeed, observe that

$$\|Ax\|_p^p \;=\; \|x - \tfrac{1}{n}Kx\|_p^p \;=\; \sum_{k=1}^{n}\left|x_k - \frac{x_1+x_2+\cdots+x_n}{n}\right|^p.$$

For $p = 1$, we see that

$$n^{-1}\|Ax\|_1 \;=\; \frac{1}{n}\sum_{k=1}^{n}|x_k - \overline{x}|,$$

corresponds to the *mean absolute deviation from the arithmetic mean* of the data set x. For $p = 2$, we have that

$$n^{-1/2}\|Ax\|_2 \;=\; \sqrt{\frac{1}{n}\sum_{k=1}^{n}|x_k - \overline{x}|^2},$$

which is the *standard deviation* of x. Although it is less robust than the mean absolute deviation from the arithmetic mean, the standard deviation remains the most commonly used measure of the amount of variation or dispersion of a set of values in various fields of application.

For $p = \infty$,

$$\|Ax\|_\infty = \max\left\{|x_k - \overline{x}| : \ k = 1, \ldots, n\right\}$$

corresponds to the *maximum absolute deviation from the arithmetic mean* of x. This maximum is realized either by the sample maximum or by the sample minimum, and it cannot be less than half the range of the data set. Being highly influenced by outliers, the maximum absolute deviation from the arithmetic mean is a highly non-robust estimator of dispersion that is seldom used.

For any value of p other than 1, 2 and ∞, one can define the *mean p-deviation from the arithmetic mean* of the vector x by

$$\frac{1}{n^{1/p}}\|Ax\|_p \ = \ \left(\frac{1}{n}\sum_{k=1}^{n}|x_k - \overline{x}|^p\right)^{1/p}.$$

Note that the greater the value of p, the heavier large deviations are weighted and thus the more heavily outliers can influence the measure of statistical dispersion.

In light of the above considerations, the value of the norm $\|A\|_{p\to p}$ for some fixed $1 \leq p < \infty$ represents the largest possible p-deviation of an n-dimensional vector lying on the unit sphere of radius $n^{1/p}$ with respect to the p-norm.

While this interpretation fosters interest in the resolution of Open question 3.8, it also adds weight to Conjecture 5.6.

Open question 6.1. *Does this interpretation lead to a precise formula for* $\|A\|_{p\to p}$ *in the special case* $A = -A(n, \frac{1-n}{n}, \frac{1}{n}) = I - \frac{1}{n}K$*?*

(ii) We encountered the following description of $A(n, -a, b)$ which comes from harmonic analysis. This observation provides another upper bound for $\|A\|_{p\to p}$. Let $\mathcal{P}_n$ denote the space of polynomials of degree at most n. We can write

$$A(n, -a, b) \ = \ -(a+b)I + bK,$$

where K is the $n \times n$ all-ones matrix. We may interpret K as an operator on $\mathcal{P}_{n-1}$. More explicitly, for each polynomial $f(z) = a_0 + a_1 z + \cdots + a_{n-1}z^{n-1} \in \mathcal{P}_{n-1}$, we have

$$(Kf)(z) \ = \ (a_0 + a_1 + \cdots + a_{n-1})\varphi(z),$$

where

$$\varphi(z) \;=\; 1 + z + \cdots + z^{n-1}.$$

Note that as a consequence of this interpretation, i.e., making a correspondence between $f \in \mathcal{P}_{n-1}$ and the vector $(a_0, a_1, \ldots, a_{n-1}) \in \mathbb{R}^n$, we have

$$\|f\|_p \;=\; (|a_0|^p + |a_1|^p + \cdots + |a_{n-1}|^p)^{1/p}.$$

Moreover, as another integral representation, it is also straightforward to see that

$$(Kf)(z) \;=\; \varphi(z) \int_0^{2\pi} f(e^{i\theta})\overline{\varphi(e^{i\theta})}\,\frac{d\theta}{2\pi}.$$

Therefore, we immediately see that

$$\begin{aligned}\|K\|_{p\to p} \;&=\; \sup_{f\in\mathcal{P}_{n-1}} \frac{\|Kf\|_p}{\|f\|_p} \\ &=\; \sup_{f\in\mathcal{P}_{n-1}} \frac{\|\varphi\|_p}{\|f\|_p} \left| \int_0^{2\pi} f(e^{i\theta})\overline{\varphi(e^{i\theta})}\,\frac{d\theta}{2\pi} \right|.\end{aligned}$$

Now, on one hand, $\|\varphi\|_p = n^{1/p}$ and, on the other hand,

$$\begin{aligned}\left| \int_0^{2\pi} f(e^{i\theta})\overline{\varphi(e^{i\theta})}\,\frac{d\theta}{2\pi} \right| \;&\le\; \int_0^{2\pi} |f(e^{i\theta})\varphi(e^{i\theta})|\,\frac{d\theta}{2\pi} \\ &\le\; \int_0^{2\pi} \|f\|_p n^{1/q} |\varphi(e^{i\theta})|\,\frac{d\theta}{2\pi} \\ &=\; n^{1/q}\|f\|_p \|\varphi\|_{L^1(\mathbb{T})},\end{aligned}$$

where

$$\|\varphi\|_{L^1(\mathbb{T})} \;=\; \int_0^{2\pi} |\varphi(e^{i\theta})|\,\frac{d\theta}{2\pi}.$$

Therefore,

$$\|K\|_{p\to p} \;\le\; n\|\varphi\|_{L^1(\mathbb{T})},$$

which gives the upper estimate

$$\|A(n,-a,b)\|_{p\to p} \;\le\; a + b + nb\|\varphi\|_{L^1(\mathbb{T})}. \tag{6.1}$$

In the light of Theorem 3.9, the estimation (6.1), which does not depend on p cannot entail an accurate upper bound for $\|A(n,-a,b)\|_{p\to p}$ for $a, b \ge 0$.

Open question 6.2. *Can variations along the lines that were used to obtain the estimation given by* (6.1) *lead to a precise formula for* $\|A\|_{p\to p}$*?*

References

[1] Mircea Andrecut. Applications of left circulant matrices in signal and image processing. *Modern Physics Letters B*, 22(04):231–241, 2008.

[2] Wathiq Bani-Domi and Fuad Kittaneh. Norm equalities and inequalities for operator matrices. *Linear Algebra Appl.*, 429(1):57–67, 2008.

[3] Friedrich L. Bauer and C. T. Fike. Norms and exclusion theorems. *Numer. Math.*, 2:137–141, 1960.

[4] Friedrich L. Bauer, Joseph. Stoer, and Christoph Witzgall. Absolute and monotonic norms. *Numer. Math.*, 3:257–264, 1961.

[5] Dario A. Bini and Albrecht Böttcher. Polynomial factorization through Toeplitz matrix computations. volume 366, pages 25–37. 2003. Special issue on structured matrices: analysis, algorithms and applications (Cortona, 2000).

[6] Albrecht Böttcher. Orthogonal symmetric Toeplitz matrices. *Complex Anal. Oper. Theory*, 2(2):285–298, 2008.

[7] Albrecth Böttcher, Sergei M. Grudsky, and Enrique Ramírez de Arellano. Approximating inverses of Toeplitz matrices by circulant matrices. *Methods Appl. Anal.*, 11(2):211–220, 2004.

[8] Ludovick Bouthat, Apoorva Khare, Javad Mashreghi, and Frédéric Morneau-Guérin. The p-norm of circulant matrices. *Linear and Multilinear Algebra*, pages 1–13, 2021.

[9] Eugène Catalan. Recherches sur les déterminants. *Bulletins de l'Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique*, 13(1):534–555, 1846.

[10] Raymond H. Chan, Xiao-Qing Jin, and Michael K. Ng. Circulant integral operators as preconditioners for Wiener-Hopf equations. *Integral Equations Operator Theory*, 21(1):12–23, 1995.

[11] Raymond H. Chan, Xiao-Qing Jin, and Man-Chung Yeung. The circulant operator in the Banach algebra of matrices. *Linear Algebra Appl.*, 149:41–53, 1991.

[12] Raymond Hon-Fu Chan and Xiao-Qing Jin. *An introduction to iterative Toeplitz solvers*, volume 5 of *Fundamentals of Algorithms*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2007.

[13] Wen Chen, Ji Lin, and Ching S. Chen. The method of fundamental solutions for solving exterior axisymmetric Helmholtz problems with high wave-number. *Adv. Appl. Math. Mech.*, 5(4):477–493, 2013.

[14] Dariusz Chruściński. On Kossakowski construction of positive maps on matrix algebras. *Open Syst. Inf. Dyn.*, 21(3):1450001, 12, 2014.

[15] Michel Crouzeix. A note on the complex and real operator norms of real matrices. *RAIRO Modél. Math. Anal. Numér.*, 20(3):427–428, 1986.

[16] Philip J. Davis. *Circulant matrices*. John Wiley & Sons, New York-Chichester-Brisbane, 1979. A Wiley-Interscience Publication, Pure and Applied Mathematics.

[17] Jorge Delgado, Neptalí Romero, Alvaro Rovella, and Francesc Vilamajó. Bounded solutions of quadratic circulant difference equations. *J. Difference Equ. Appl.*, 11(10):897–907, 2005.

[18] Victor D. Didenko and Bernd Silbermann. On the stability of some operator sequences and the approximate solution of singular integral equations with conjugation. *Integral Equations Operator Theory*, 16(2):224–243, 1993.

[19] Paul A. Fuhrmann. *A polynomial approach to linear algebra*. Universitext. Springer-Verlag, New York, 1996.

[20] Stephan Ramon Garcia, Javad Mashreghi, and William T. Ross. *Introduction to model spaces and their operators*, volume 148 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2016.
[21] Robert M. Gray. Toeplitz and circulant matrices: a review. *Communications and Information Theory*, 2(3):155–239, 2005.
[22] Qiang Guo and Man W. Wong. Analysis of matrices of pseudo-differential operators with separable symbols on $\mathbb{Z}_N$. *J. Pseudo-Differ. Oper. Appl.*, 7(2):249–259, 2016.
[23] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, second edition, 2013.
[24] Marko Huhtanen and Olavi Nevanlinna. The real linear resolvent and cosolvent operators. *J. Operator Theory*, 58(2):229–250, 2007.
[25] Thang Huynh and Rayan Saab. Fast binary embeddings and quantized compressed sensing with structured matrices. *Comm. Pure Appl. Math.*, 73(1):110–149, 2020.
[26] Xiaoyu Jiang and Kicheon Hong. Equalities and inequalities for norms of block imaginary circulant operator matrices. *Abstr. Appl. Anal.*, pages Art. ID 521214, 5, 2015.
[27] Zhao Lin Jiang, Yun Cheng Qiao, and Shu Dong Wang. Norm equalities and inequalities for three circulant operator matrices. *Acta Math. Sin. (Engl. Ser.)*, 33(4):571–590, 2017.
[28] Danko R. Jocić. Clarkson-McCarthy inequalities for several operators and related norm inequalities for p-modified unitarily invariant norms. *Complex Anal. Oper. Theory*, 13(3):583–613, 2019.
[29] Jacek Jurkowski. On numerical ranges of operators. In *Quantum bio-informatics V*, volume 30 of *QP–PQ: Quantum Probab. White Noise Anal.*, pages 217–228. World Sci. Publ., Hackensack, NJ, 2013.
[30] Dan Kalman and James E. White. Polynomial equations and circulant matrices. *Amer. Math. Monthly*, 108(9):821–840, 2001.
[31] Wen-Fong Ke, King-Fai Lai, Tsung-Lin Lee, and Ngai-Ching Wong. Preconditioning random Toeplitz systems. *J. Nonlinear Convex Anal.*, 17(4):757–770, 2016.
[32] Fuad Kittaneh and Satyajit Sahoo. On $\mathbb{A}$-numerical radius equalities and inequalities for certain operator matrices. *Ann. Funct. Anal.*, 12(4):Paper No. 52, 23, 2021.
[33] Irwin Kra and Santiago R. Simanca. On circulant matrices. *Notices Amer. Math. Soc.*, 59(3):368–377, 2012.
[34] Wieslaw Krawcewicz, Shiwang Ma, and Jianhong Wu. Multiple slowly oscillating periodic solutions in coupled lossless transmission lines. *Nonlinear Anal. Real World Appl.*, 5(2):309–354, 2004.
[35] Walter Ledermann. *Complex Numbers*. Springer Science & Business Media, 2013.
[36] Jongrak Lee. Hyponormality of Toeplitz operators on the weighted Bergman space with matrix-valued circulant symbols. *Linear Algebra Appl.*, 576:35–50, 2019.
[37] Jorgen Lofstrom and Joran Bergh. Interpolation spaces. *SpringereVerlag, Newe*, 1976.
[38] Joachim M. Mouanda. On von Neumann's inequality for complex triangular Toeplitz contractions. *Rocky Mountain J. Math.*, 50(1):213–224, 2020.
[39] Thomas Muir. *The Theory of Determinants In The Historical Order Of Development (4 volumes)*. Macmillan and Company, limited, 1906.
[40] Marcel Riesz. Sur les maxima des formes bilinéaires et sur les fonctionnelles linéaires. *Acta Math.*, 49(3-4):465–497, 1927.
[41] K. R. Sahasranand. The p-norm of circulant matrices via Fourier analysis. *Concr. Oper.*, 9(1):1–5, 2022.
[42] Thomas Strohmer. Four short stories about Toeplitz matrix calculations. volume 343/344,

pages 321–344. 2002. Special issue on structured and infinite systems of linear equations.

[43] James J. Sylvester. LX. Thoughts on inverse orthogonal matrices, simultaneous signsuccessions, and tessellated pavements in two or more colours, with applications to Newton's rule, ornamental tile-work, and the theory of numbers. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 34(232):461–475, 1867.

[44] Angus E. Taylor. The norm of a real linear transformation in Minkowski space. *Enseign. Math. (2)*, 4:101–107, 1958.

[45] Alan C. Wilde. Differential equations involving circulant matrices. *Rocky Mountain J. Math.*, 13(1):1–13, 1983.

[46] Alan C. Wilde. Algebras of operators isomorphic to the circulant algebra. *Proc. Amer. Math. Soc.*, 105(4):808–816, 1989.

[47] H-J Wittsack, Afra M Wohlschläger, Eva K Ritzl, Raimund Kleiser, Mathias Cohnen, Rüdiger J Seitz, and Ulrich Mödder. Ct-perfusion imaging of the human brain: advanced deconvolution analysis using circulant singular value decomposition. *Computerized Medical Imaging and Graphics*, 32(1):67–77, 2008.

[48] Shu-Lin Wu and Tao Zhou. Parallel implementation for the two-stage SDIRK methods via diagonalization. *J. Comput. Phys.*, 428:Paper No. 110076, 18, 2021.