ON CHEBYSHEV POLYNOMIALS OF MATRICES

VANCE FABER*, JÖRG LIESEN†, AND PETR TICHÝ‡

Abstract. The mth Chebyshev polynomial of a square matrix A is the monic polynomial that minimizes the matrix 2-norm of $p(A)$ over all monic polynomials $p(z)$ of degree m. This polynomial is uniquely defined if m is less than the degree of the minimal polynomial of A. We study general properties of Chebyshev polynomials of matrices, which in some cases turn out to be generalizations of well known properties of Chebyshev polynomials of compact sets in the complex plane. We also derive explicit formulas of the Chebyshev polynomials of certain classes of matrices, and explore the relation between Chebyshev polynomials of one of these matrix classes and Chebyshev polynomials of lemniscatic regions in the complex plane.

Key words. matrix approximation problems, Chebyshev polynomials, complex approximation theory, Krylov subspace methods, Arnoldi's method

AMS subject classifications. 41A10, 15A60, 65F10

1. Introduction.

Let $A \in \mathbb{C}^{n\times n}$ be a given matrix, let $m \geq 1$ be a given integer, and let $\mathcal{M}_m$ denote the set of complex *monic* polynomials of degree m. We consider the approximation problem

$$\min_{p\in\mathcal{M}_m} \|p(A)\|, \tag{1.1}$$

where $\|\cdot\|$ denotes the matrix 2-norm (or spectral norm). As shown by Greenbaum and Trefethen [11, Theorem 2] (also cf. [13, Theorem 2.2]), the problem (1.1) has a uniquely defined solution when m is smaller than $d(A)$, the degree of the minimal polynomial of A. This is a nontrivial result since the matrix 2-norm is not strictly convex, and approximation problems in such norms are in general not guaranteed to have a unique solution; see [13, pp. 853–854] for more details and an example. In this paper we assume that $m < d(A)$, which is necessary and sufficient so that the value of (1.1) is positive, and we denote the unique solution of (1.1) by $T_m^A(z)$. Note that if $A \in \mathbb{R}^{n\times n}$, then the Chebyshev polynomials of A have real coefficients, and hence in this case it suffices to consider only real monic polynomials in (1.1).

It is clear that (1.1) is unitarily invariant, i.e., that $T_m^A(z) = T_m^{U^*AU}(z)$ for any unitary matrix $U \in \mathbb{C}^{n\times n}$. In particular, if the matrix A is normal, i.e., unitarily diagonalizable, then

$$\min_{p\in\mathcal{M}_m} \|p(A)\| \;=\; \min_{p\in\mathcal{M}_m} \max_{\lambda\in\Lambda(A)} |p(\lambda)|,$$

where $\Lambda(A)$ denotes the set of the eigenvalues of A. The (uniquely defined) mth degree monic polynomial that deviates least from zero on a compact set Ω in the

*CloudPak, Seattle, WA, USA (vance.faber@gmail.com).

†Institute of Mathematics, Technical University of Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany (liesen@math.tu-berlin.de). The work of this author was supported by the Heisenberg Program of the Deutsche Forschungsgemeinschaft.

‡Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 18207 Prague, Czech Republic (tichy@cs.cas.cz). The work of this author is a part of the Institutional Research Plan AV0Z10300504, it was supported by the GAAS grant IAA100300802 and by the project M100300901 of the institutional support of the Academy of Sciences of the Czech Republic.

complex plane is called the *mth Chebyshev polynomial*[1] *of the set* Ω. We denote this polynomial by $T_m^{\Omega}(z)$.

The last equation shows that for a normal matrix A the matrix approximation problem (1.1) is equal to the scalar approximation problem of finding $T_m^{\Lambda(A)}(z)$, and in fact $T_m^A(z) = T_m^{\Lambda(A)}(z)$. Because of these relations, the problem (1.1) can be considered a generalization of a classical problem of mathematics from scalars to matrices. As a consequence, Greenbaum and Trefethen [11] as well as Toh and Trefethen [25] have called the solution $T_m^A(z)$ of (1.1) the *mth Chebyshev polynomial of the matrix A* (regardless of A being normal or not).

A motivation for studying the problem (1.1) and the Chebyshev polynomials of matrices comes from their connection to Krylov subspace methods, and in particular the Arnoldi method for approximating eigenvalues of matrices [2]. In a nutshell, after m steps of this method a relation of the form $AV_m = V_m H_m + r_m e_m^T$ is computed, where $H_m \in \mathbb{C}^{m\times m}$ is an upper Hessenberg matrix, $r_m \in \mathbb{C}^n$ is the mth "residual" vector, e_m is the mth canonical basis vector of $\mathbb{C}^m$, and the columns of $V_m \in \mathbb{C}^{n\times m}$ form an orthonormal basis of the Krylov subspace $\mathcal{K}_m(A, v_1) = \text{span}\{v_1, Av_1, \ldots, A^{m-1}v_1\}$. The vector $v_1 \in \mathbb{C}^n$ is an arbitrary (nonzero) initial vector. The eigenvalues of H_m are used as approximations for the eigenvalues of A. Note that $r_m = 0$ if and only if the columns of V_m span an invariant subspace of A, and if this holds, then each eigenvalue of H_m is an eigenvalue of A.

As shown by Saad [15, Theorem 5.1], the characteristic polynomial φ_m of H_m satisfies

$$\|\varphi_m(A)v_1\| \;=\; \min_{p\in\mathcal{M}_m} \|p(A)v_1\|. \tag{1.2}$$

An interpretation of this result is that the characteristic polynomial of H_m solves the Chebyshev approximation problem for A *with respect to the given starting vector* v_1. Saad pointed out that (1.2) "seems to be the only known optimality property that is satisfied by the [Arnoldi] approximation process in the nonsymmetric case" [16, p. 171]. To learn more about this property, Greenbaum and Trefethen [11, p. 362] suggested "to disentangle [the] matrix essence of the process from the distracting effects of the initial vector", and hence study the "idealized" problem (1.1) instead of (1.2). They referred to the solution of (1.1) as the *mth ideal Arnoldi polynomial of A* (in addition to the name mth Chebyshev polynomial of A).

Greenbaum and Trefethen [11] seem to be the first who studied existence and uniqueness of Chebyshev polynomials of matrices. Toh and Trefethen [24] derived an algorithm for computing these polynomials based on semidefinite programming; see also Toh's PhD thesis [21, Chapter 2]. This algorithm is now part of the SDPT3 Toolbox [23]. The paper [24] as well as [21] and [26, Chapter 29] give numerous computed examples for the norms, roots, and coefficients of Chebyshev polynomials of matrices. It is shown numerically that the lemniscates of these polynomials tend to approximate pseudospectra of A. In addition, Toh has shown that the zeros of $T_m^A(z)$ are contained in the field of values of A [21, Theorem 5.10]. This result is part of his interesting analysis of Chebyshev polynomials of linear operators in infinite dimensional Hilbert spaces [21, Chapter 5]. The first explicit solutions for the problem

[1] Pafnuti Lvovich Chebyshev (1821–1894) determined the polynomials $T_m^{\Omega}(z)$ of $\Omega = [-a, a]$ (a real interval symmetric to zero) in his 1859 paper [5], which laid the foundations of modern approximation theory. We recommend Steffens' book [18] to readers who are interested in the historical development of the subject.

(1.1) for a nonnormal matrix A we are aware of have been given in [13, Theorem 3.4]. It is shown there that $T_m^A(z) = (z-\lambda)^m$ if $A = J_\lambda$, a Jordan block with eigenvalue $\lambda \in \mathbb{C}$. Note that in this case the Chebyshev polynomials of A are independent of the size of A.

The above remarks show that the problem (1.1) is a mathematically interesting generalization of the classical Chebyshev problem, which has an important application in the area of iterative methods. Yet, our survey of the literature indicates that there has been little theoretical work on Chebyshev polynomials of matrices (in particular when compared with the substantial work on Chebyshev polynomials for compact sets). The main motivation for writing this paper was to extend the existing theory of Chebyshev polynomials of matrices. Therefore we considered a number of known properties of Chebyshev polynomials of compact sets, and tried to find matrix analogues. Among these are the behavior of $T_m^A(z)$ under shifts and scaling of A, a matrix analogue of the "alternation property", as well as conditions on A so that $T_m^A(z)$ is even or odd (Section 2). We also give further explicit examples of Chebyshev polynomials of some classes of matrices (Section 3). For a class of block Toeplitz matrices, we explore the relation between their Chebyshev polynomials and Chebyshev polynomials of lemniscatic regions in the complex plane (Section 4).

All computations in this paper have been performed using MATLAB [20]. For computing Chebyshev polynomials of matrices we have used the DSDP software package for semidefinite programming [3] and its MATLAB interface.

2. General results.

In this section we state and prove results on the Chebyshev polynomials of a general matrix A. In later sections we will apply these results to some specific examples.

2.1. Chebyshev polynomials of shifted and scaled matrices.

In the following we will write a complex (monic) polynomial of degree m as a function of the variable z and its coefficients. More precisely, for $x = [x_0, \dots, x_{m-1}]^T \in \mathbb{C}^m$ we write

$$p(z;x) \;\equiv\; z^m - \sum_{j=0}^{m-1} x_j\, z^j \;\in\; \mathcal{M}_m. \tag{2.1}$$

Let two complex numbers, α and β, be given, and define $\delta \equiv \beta - \alpha$. Then

$$\begin{aligned} p(\beta+z;x) &= p((\beta-\alpha)+(\alpha+z);x) = (\delta+(\alpha+z))^m - \sum_{j=0}^{m-1} x_j(\delta+(\alpha+z))^j \\ &= \sum_{j=0}^{m}\binom{m}{j}\delta^{m-j}(\alpha+z)^j \;-\; \sum_{j=0}^{m-1} x_j \sum_{\ell=0}^{j}\binom{j}{\ell}\delta^{j-\ell}(\alpha+z)^\ell \\ &= (\alpha+z)^m + \sum_{j=0}^{m-1}\left(\binom{m}{j}\delta^{m-j}(\alpha+z)^j - x_j\sum_{\ell=0}^{j}\binom{j}{\ell}\delta^{j-\ell}(\alpha+z)^\ell\right) \\ &= (\alpha+z)^m - \sum_{j=0}^{m-1}\left(\sum_{\ell=j}^{m-1}\binom{\ell}{j}\delta^{\ell-j}x_\ell - \binom{m}{j}\delta^{m-j}\right)(\alpha+z)^j \qquad (2.2) \\ &\equiv (\alpha+z)^m - \sum_{j=0}^{m-1} y_j(\alpha+z)^j \\ &\equiv p(\alpha+z;y). \end{aligned}$$

A closer examination of (2.2) shows that the two vectors y and x in the identity $p(\alpha + z; y) = p(\beta + z; x)$ are related by

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{m-1} \end{bmatrix} = \begin{bmatrix} \binom{0}{0}\delta^0 & \binom{1}{0}\delta^1 & \binom{2}{0}\delta^2 & \cdots & \binom{m-1}{0}\delta^{m-1} \\ & \binom{1}{1}\delta^0 & \binom{2}{1}\delta^1 & \cdots & \binom{m-2}{1}\delta^{m-2} \\ & & \ddots & & \vdots \\ & & & & \binom{m-1}{m-1}\delta^0 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{m-1} \end{bmatrix} - \begin{bmatrix} \binom{m}{0}\delta^m \\ \binom{m}{1}\delta^{m-1} \\ \vdots \\ \binom{m}{m-1}\delta^1 \end{bmatrix}.$$

We can write this as

$$y = h_\delta(x), \quad \text{where} \quad h_\delta(x) \equiv M_\delta x - v_\delta. \tag{2.3}$$

The matrix $M_\delta \in \mathbb{C}^{m\times m}$ is an invertible upper triangular matrix; all its diagonal elements are equal to 1. Thus, for any $\delta \in \mathbb{C}$,

$$h_\delta \,:\, x \mapsto M_\delta x - v_\delta$$

is an invertible affine linear transformation on $\mathbb{C}^m$. Note that if $\delta = 0$, then $M_\delta = I$ (the identity matrix) and $v_\delta = 0$, so that $y = x$.

The above derivation can be repeated with αI, βI, and A replacing α, β, and z, respectively. This yields the following result.

LEMMA 2.1. *Let $A \in \mathbb{C}^{n\times n}$, $x \in \mathbb{C}^m$, $\alpha \in \mathbb{C}$, and $\beta \in \mathbb{C}$ be given. Then for any monic polynomial p of the form (2.1),*

$$p(\beta I + A; x) = p(\alpha I + A; h_\delta(x)), \tag{2.4}$$

where $\delta \equiv \beta - \alpha$, and h_δ is defined as in (2.3).

The assertion of this lemma is an ingredient in the proof of the following theorem.

THEOREM 2.2. *Let $A \in \mathbb{C}^{n\times n}$, $\alpha \in \mathbb{C}$, and a positive integer $m < d(A)$ be given. Denote by $T_m^A(z) = p(z; x_*)$ the mth Chebyshev polynomial of A. Then the following hold:*

$$\min_{p\in\mathcal{M}_m} \|p(A+\alpha I)\| = \min_{p\in\mathcal{M}_m} \|p(A)\|, \qquad T_m^{A+\alpha I}(z) = p(z; h_{-\alpha}(x_*)), \tag{2.5}$$

where $h_{-\alpha}$ is defined as in (2.3), and

$$\min_{p\in\mathcal{M}_m} \|p(\alpha A)\| = |\alpha|^m \min_{p\in\mathcal{M}_m} \|p(A)\|, \qquad T_m^{\alpha A}(z) = p(z; D_\alpha x_*). \tag{2.6}$$

where $D_\alpha \equiv \mathrm{diag}(\alpha^m, \alpha^{m-1}, \ldots, \alpha)$.

Proof. We first prove (2.5). Equation (2.4) with $\beta = 0$ shows that $p(A; x) = p(A + \alpha I; h_{-\alpha}(x))$ holds for any $x \in \mathbb{C}^m$. This yields

$$\begin{aligned} \min_{p\in\mathcal{M}_m} \|p(A+\alpha I)\| &= \min_{x\in\mathbb{C}^m} \|p(A+\alpha I; x)\| = \min_{x\in\mathbb{C}^m} \|p(A+\alpha I; h_{-\alpha}(x))\| \\ &= \min_{x\in\mathbb{C}^m} \|p(A; x)\| = \min_{p\in\mathcal{M}_m} \|p(A)\| \end{aligned}$$

(here we have used that the transformation $h_{-\alpha}$ is invertible). To see that the polynomial $p(z; h_{-\alpha}(x_*))$ is indeed the mth Chebyshev polynomial of $A + \alpha I$, we note that

$$\|p(A+\alpha I; h_{-\alpha}(x_*))\| = \|p(A; x_*)\| = \min_{p\in\mathcal{M}_m} \|p(A)\| = \min_{p\in\mathcal{M}_m} \|p(A+\alpha I)\|.$$

The equations in (2.6) are trivial if $\alpha = 0$, so we can assume that $\alpha \neq 0$. Then the matrix D_α is invertible, and a straightforward computation yields

$$\begin{aligned}\min_{p\in\mathcal{M}_m} \|p(\alpha A)\| &= \min_{x\in\mathbb{C}^m} \|p(\alpha A; x)\| = |\alpha|^m \min_{x\in\mathbb{C}^m} \|p(A; D_\alpha^{-1}x)\| = |\alpha|^m \min_{x\in\mathbb{C}^m} \|p(A; x)\| \\ &= |\alpha|^m \min_{p\in\mathcal{M}_m} \|p(A)\|.\end{aligned}$$

Furthermore,

$$\|p(\alpha A; D_\alpha x_*)\| = |\alpha|^m \|p(A; x_*)\| = |\alpha|^m \min_{p\in\mathcal{M}_m} \|p(A)\| = \min_{p\in\mathcal{M}_m} \|p(\alpha A)\|,$$

so that $p(z; D_\alpha x_*)$ is the mth Chebyshev polynomial of the matrix αA. □

The fact that the "true" Arnoldi approximation problem, i.e., the right hand side of (1.2), is translation invariant has been mentioned previously in [11, p. 361]. Hence the translation invariance of the problem (1.1) shown in (2.5) is not surprising. The underlying reason is that the monic polynomials are normalized "at infinity".

The result for the scaled matrices in (2.6), which also may be expected, has an important consequence that is easily overlooked: Suppose that for some given $A \in \mathbb{C}^{n\times n}$ we have computed the sequence of norms of the problem (1.1), i.e., the quantities

$$\|T_1^A(A)\|, \ \|T_2^A(A)\|, \ \|T_3^A(A)\|, \ldots.$$

If we scale A by $\alpha \in \mathbb{C}$, then the norms of the Chebyshev approximation problem *for the scaled matrix* αA are given by

$$|\alpha|\,\|T_1^A(A)\|, \ |\alpha|^2\,\|T_2^A(A)\|, \ |\alpha|^3\,\|T_3^A(A)\|, \ldots.$$

A suitable scaling can therefore turn any given sequence of norms for the problem with A into a strictly monotonically decreasing (or, if we prefer, increasing) sequence for the problem with αA. For example, the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

yields

$$\|T_0^A(A)\| = 1, \ \|T_1^A(A)\| \approx 11.4077, \ \|T_2^A(A)\| = 9,$$

cf. [26, p. 280] (note that by definition $T_0^A(z) \equiv 1$ for any matrix A). The corresponding norms for the scaled matrices $\frac{1}{12}\cdot A$ and $12 \cdot A$ are then (approximately) given by

$$1, \ 0.95064, \ 0.0625, \quad \text{and} \quad 1, \ 136.8924, \ 1296,$$

respectively. In general we expect that the behavior of an iterative method for solving linear systems or for approximating eigenvalues is invariant under scaling of the given matrix. In particular, by looking at the sequence of norms of the problem (1.1) *alone* we cannot determine how fast a method "converges". In practice, we always have to measure "convergence" in some *relative* (rather than absolute) sense. Note that the quantity $\min_{p\in\mathcal{M}_m} \|p(A)\|/\|A^m\|$ is independent of a scaling of the matrix A, and hence in our context it may give relevant information. We have not explored this topic further.

2.2. Alternation property for block diagonal matrices. It is well known that Chebyshev polynomials of compact sets Ω are characterized by an *alternation property*. For example, if $\Omega = [a, b]$ is a finite real interval, then $p(z) \in \mathcal{M}_m$ is the unique Chebyshev polynomial of degree m on Ω if and only if $p(z)$ assumes its extreme values $\pm \max_{z\in\Omega} |p(z)|$ with successively alternating signs on at least $m + 1$ points (the "alternation points") in Ω; see, e.g., [4, Section 7.5]. There exist generalizations of this classical result to complex as well as to finite sets Ω; see, e.g., [6, Chapter 3] and [4, Section 7.5]. The following is a generalization to block-diagonal matrices.

THEOREM 2.3. *Consider a block-diagonal matrix $A = \mathrm{diag}(A_1, \ldots, A_h)$, let $k \equiv \max_{1\leq j\leq h} d(A_j)$, and let ℓ be a given positive integer such that $k \cdot \ell < d(A)$. Then the matrix $T^A_{k\cdot\ell}(A) = \mathrm{diag}(T^A_{k\cdot\ell}(A_1), \ldots, T^A_{k\cdot\ell}(A_h))$ has at least $\ell + 1$ diagonal blocks $T^A_{k\cdot\ell}(A_j)$ with norm equal to $\|T^A_{k\cdot\ell}(A)\|$.*

Proof. The assumption that $k \cdot \ell < d(A)$ implies that $T^A_{k\cdot\ell}(z)$ is uniquely defined. For simplicity of notation we denote $B = T^A_{k\cdot\ell}(A)$ and $B_j \equiv T^A_{k\cdot\ell}(A_j)$, $j = 1, \ldots, h$. Without loss of generality we can assume that $\|B\| = \|B_1\| \geq \cdots \geq \|B_h\|$.

Suppose that the assertion is false. Then there exists an integer i, $1 \leq i \leq \ell$, so that $\|B\| = \|B_1\| = \cdots = \|B_i\| > \|B_{i+1}\|$. Let $\delta \equiv \|B\| - \|B_{i+1}\| > 0$, and let $q_j(z) \in \mathcal{M}_k$ be a polynomial with $q_j(A_j) = 0$, $1 \leq j \leq h$. Define the polynomial

$$t(z) \equiv \prod_{j=1}^{\ell} q_j(z) \;\in\; \mathcal{M}_{k\cdot\ell}.$$

Let ϵ be a positive real number with

$$\epsilon < \frac{\delta}{\delta + \max\limits_{1\leq j\leq h} \|t(A_j)\|}.$$

Then $0 < \epsilon < 1$, and hence

$$r_\epsilon(z) \equiv (1 - \epsilon)\, T^A_{k\cdot\ell}(z) + \epsilon\, t(z) \;\in\; \mathcal{M}_{k\cdot\ell}.$$

Note that $\|r_\epsilon(A)\| = \max_{1\leq j\leq h} \|r_\epsilon(A_j)\|$.
For $1 \leq j \leq i$, we have $\|r_\epsilon(A_j)\| = (1 - \epsilon)\, \|B_j\| = (1 - \epsilon)\, \|B\| < \|B\|$.
For $i + 1 \leq j \leq h$, we have

$$\begin{aligned}
\|r_\epsilon(A_j)\| &= \|(1 - \epsilon)\, B_j + \epsilon\, t(A_j)\| \\
&\leq (1 - \epsilon)\, \|B_j\| + \epsilon\, \|t(A_j)\| \\
&\leq (1 - \epsilon)\, \|B_{i+1}\| + \epsilon\, \|t(A_j)\| \\
&= (1 - \epsilon)\, (\|B\| - \delta) + \epsilon\, \|t(A_j)\| \\
&= (1 - \epsilon)\, \|B\| + \epsilon\, (\delta + \|t(A_j)\|) - \delta.
\end{aligned}$$

Since $\epsilon\, (\delta + \|t(A_j)\|) - \delta < 0$ by the definition of ϵ, we see that $\|r_\epsilon(A_j)\| < \|B\|$ for $i + 1 \leq j \leq h$. But this means that $\|r_\epsilon(A)\| < \|B\|$, which contradicts the minimality of the Chebyshev polynomial of A. ◻

The numerical results shown in Table 2.1 illustrate this theorem. We have used a block diagonal matrix A with 4 Jordan blocks of size 3×3 on its diagonal, so that $k = 3$. Theorem 2.3 then guarantees that $T^A_{3\ell}(A)$, $\ell = 1, 2, 3$, has at least $\ell+1$ diagonal blocks with the same maximal norm. This is clearly confirmed for $\ell = 1$ and $\ell = 2$

TABLE 2.1

Numerical illustration of Theorem 2.3: Here $A = \mathrm{diag}(A_1, A_2, A_3, A_4)$, where each $A_j = J_{\lambda_j}$ is a 3×3 Jordan block. The four eigenvalues are -3, -0.5, 0.5, 0.75.

m	$\|T_m^A(A_1)\|$	$\|T_m^A(A_2)\|$	$\|T_m^A(A_3)\|$	$\|T_m^A(A_4)\|$
1	<u>2.6396</u>	1.4620	2.3970	<u>2.6396</u>
2	<u>4.1555</u>	<u>4.1555</u>	3.6828	<u>4.1555</u>
3	<u>9.0629</u>	5.6303	7.6858	<u>9.0629</u>
4	<u>14.0251</u>	<u>14.0251</u>	11.8397	<u>14.0251</u>
5	<u>22.3872</u>	20.7801	17.6382	<u>22.3872</u>
6	<u>22.6857</u>	<u>22.6857</u>	20.3948	<u>22.6857</u>

(it also holds for $\ell = 3$). For these ℓ we observe that *exactly* $\ell + 1$ diagonal blocks achieve the maximal norm. Hence in general the lower bound of $\ell+1$ blocks attaining the maximal norm in step $m = k \cdot \ell$ cannot be improved. In addition, we see in this experiment that the number of diagonal blocks with the same maximal norm is not necessarily a monotonically increasing function of the degree of the Chebyshev polynomial.

Now consider the matrix

$$A = \mathrm{diag}(A_1, A_2) = \left[\begin{array}{cc|cc} 1 & 1 & & \\ & 1 & & \\ \hline & & -1 & 1 \\ & & & -1 \end{array}\right].$$

Then for $p(z) = z^2 - \alpha z - \beta \in \mathcal{M}_2$ we get

$$p(A) = \left[\begin{array}{cc|cc} 1-(\alpha+\beta) & 2-\alpha & & \\ & 1-(\alpha+\beta) & & \\ \hline & & 1-(\alpha+\beta) & -2-\alpha \\ & & & 1-(\alpha+\beta) \end{array}\right].$$

For $\alpha = 0$ and *any* $\beta \in \mathbb{R}$ we will have $\|p(A)\| = \|p(A_1)\| = \|p(A_2)\|$. Hence there are infinitely many polynomials $p \in \mathcal{M}_2$ for which the two diagonal blocks have the same maximal norm. One of these polynomials is the Chebyshev polynomial $T_2^A(z) = z^2+1$. As shown by this example, the condition in Theorem 2.3 on a polynomial $p \in \mathcal{M}_{k\cdot\ell}$ that at least $\ell + 1$ diagonal blocks of $p(A)$ have equal maximal norm is in general necessary but *not* sufficient so that $p(z) = T_{k\cdot\ell}^A(z)$.

Finally, as a special case of Theorem 2.3 consider a matrix $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ with distinct diagonal elements $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$. If $m < n$, then there are at least $m + 1$ diagonal elements λ_j with $|T_m^A(\lambda_j)| = \|T_m^A(A)\| = \max_{1\leq i\leq n} |T_m^A(\lambda_i)|$. Note that $T_m^A(z)$ in this case is equal to the mth Chebyshev polynomial of the finite set $\{\lambda_1, \ldots, \lambda_n\} \subset \mathbb{C}$. This shows that the Chebyshev polynomial of degree m of a set in the complex plane consisting of $n \geq m + 1$ points attains its maximum value at least at $m + 1$ points.

2.3. Chebyshev polynomials with known zero coefficients. In this section we study properties of a matrix A so that its Chebyshev polynomials have known zero coefficients. An extreme case in this respect is when $T_m^A(z) = z^m$, i.e., when all coefficients of $T_m^A(z)$, except the leading one, are zero. This happens if and only if

$$\|A^m\| = \min_{p\in\mathcal{M}_m} \|p(A)\|.$$

Equivalently, this says that the zero matrix is the best approximation of A^m from the linear space $\text{span}\{I, A, \ldots, A^{m-1}\}$ (with respect to the matrix 2-norm). To characterize this property, we recall that the dual norm to the matrix 2-norm $\|\cdot\|$ is the trace norm (also called energy norm or c_1-norm),

$$|||\, M \,||| \equiv \sum_{j=1}^{r} \sigma_j(M)\,, \tag{2.7}$$

where $\sigma_1(M), \ldots, \sigma_r(M)$ denote the singular values of the matrix $M \in \mathbb{C}^{n\times n}$ with $\text{rank}(M) = r$. For $X \in \mathbb{C}^{n\times n}$ and $Y \in \mathbb{C}^{n\times n}$ we define the inner product $\langle X, Y\rangle \equiv \text{trace}(Y^*X)$. We can now adapt a result of Ziętak [28, p. 173] to our context and notation.

Theorem 2.4. *Let $A \in \mathbb{C}^{n\times n}$ and a positive integer $m < d(A)$ be given. Then*

$$T_m^A(z) = z^m$$

if and only if there exists a matrix $Z \in \mathbb{C}^{n\times n}$ with $|||\, Z \,||| = 1$, such that

$$\langle Z, A^k\rangle = 0, \quad k = 0, \ldots, m-1, \qquad \textit{and} \qquad \text{Re}\,\langle Z, A^m\rangle = \|A^m\|\,. \tag{2.8}$$

As shown in [13, Theorem 3.4], the mth Chebyshev polynomial of a Jordan block J_λ with eigenvalue $\lambda \in \mathbb{C}$ is given by $(z-\lambda)^m$. In particular, the mth Chebyshev polynomial of J_0 is of the form z^m. A more general class of matrices with $T_m^A(z) = z^m$ is studied in Section 3.1 below.

It is well known that the Chebyshev polynomials of real intervals that are symmetric with respect to the origin are alternating between even and odd, i.e., every second coefficient (starting from the highest one) of $T_m^{[-a,a]}(z)$ is zero, which means that $T_m^{[-a,a]}(z) = (-1)^m T_m^{[-a,a]}(-z)$. We next give an analogue of this result for Chebyshev polynomials of matrices.

Theorem 2.5. *Let $A \in \mathbb{C}^{n\times n}$ and a positive integer $m < d(A)$ be given. If there exists a unitary matrix P such that either $P^*AP = -A$, or $P^*AP = -A^T$, then*

$$T_m^A(z) = (-1)^m T_m^A(-z). \tag{2.9}$$

Proof. We only prove the assertion in case $P^*AP = -A$; the other case is similar. If this relation holds for a unitary matrix P, then

$$\|(-1)^m T_m^A(-A)\| = \|T_m^A(P^*AP)\| = \|P^*T_m^A(A)P\| = \|T_m^A(A)\| = \min_{p\in\mathcal{M}_m} \|p(A)\|,$$

and the result follows from the uniqueness of the mth Chebyshev polynomial of A. □

As a special case consider a normal matrix A and its unitary diagonalization $U^*AU = D$. Then $T_m^A(z) = T_m^D(z)$, so we may only consider the Chebyshev polynomial of the diagonal matrix D. Since $D = D^T$, the conditions in Theorem 2.5 are satisfied if and only if there exists a unitary matrix P such that $P^*DP = -D$. But this means that the set of the diagonal elements of D (i.e., the eigenvalues of A) must be symmetric with respect to the origin (i.e., if λ_j is an eigenvalue, $-\lambda_j$ is an eigenvalue as well). Therefore, whenever a discrete set $\Omega \subset \mathbb{C}$ is symmetric with respect to the origin, the Chebyshev polynomial $T_m^\Omega(z)$ is even (odd) if m is even (odd).

As an example of a nonnormal matrix, consider

$$A = \begin{bmatrix} 1 & \epsilon & & \\ & -1 & 1/\epsilon & \\ & & 1 & \epsilon \\ & & & -1 \end{bmatrix}, \quad \epsilon > 0\,,$$

which has been used by Toh [22] in his analysis of the convergence of the GMRES method. He has shown that $P^T AP = -A$, where

$$P = \begin{bmatrix} & & & 1 \\ & & -1 & \\ & 1 & & \\ -1 & & & \end{bmatrix}$$

is an orthogonal matrix.

For another example consider

$$C = \begin{bmatrix} J_\lambda & \\ & J_{-\lambda} \end{bmatrix}, \qquad J_\lambda,\, J_{-\lambda} \in \mathbb{C}^{n\times n}, \quad \lambda \in \mathbb{C}\,. \tag{2.10}$$

It is easily seen that

$$J_{-\lambda} = -I^{\pm} J_\lambda I^{\pm}, \quad \text{where} \quad I^{\pm} = \mathrm{diag}(1, -1, 1, \ldots, (-1)^{n-1}) \in \mathbb{R}^{n\times n}. \tag{2.11}$$

Using the symmetric and orthogonal matrices

$$P = \begin{bmatrix} & I \\ I & \end{bmatrix}, \qquad Q = \begin{bmatrix} I^{\pm} & \\ & I^{\pm} \end{bmatrix},$$

we receive $QPCPQ = -C$.

The identity (2.11) implies that

$$\|T_m^C(J_{-\lambda})\| \;=\; \|T_m^C(-I^{\pm} J_\lambda I^{\pm})\| \;=\; \|T_m^C(J_\lambda)\|,$$

i.e., the Chebyshev polynomials of C attain the same norm on each of the two diagonal blocks. In general, we can shift and rotate any matrix consisting of two Jordan blocks of the same size and with respective eigenvalues $\lambda, \mu \in \mathbb{C}$ into the form (2.10). It then can be shown that the Chebyshev polynomials $T_m^A(z)$ of $A = \mathrm{diag}(J_\lambda, J_\mu)$ satisfy the "norm balancing property" $\|T_m^A(J_\lambda)\| = \|T_m^A(J_\mu)\|$. The proof of this property is rather technical and we skip it for brevity.

2.4. Linear Chebyshev polynomials. In this section we consider the linear Chebyshev problem

$$\min_{\alpha\in\mathbb{C}} \|A - \alpha I\|.$$

Work related to this problem has been done by Friedland [8], who characterized solutions of the problem $\min_{\alpha\in\mathbb{R}} \|A - \alpha B\|$, where A and B are two complex, and possibly rectangular matrices. This problem in general does not have a unique solution. More recently, Afanasjev et al. [1] have studied the restarted Arnoldi method with restart length equal to 1. The analysis of this method involves approximation problems of the form $\min_{\alpha\in\mathbb{C}} \|(A - \alpha I)v_1\|$ (cf. (1.2)), whose unique solution is $\alpha = v_1^* A v_1$.

THEOREM 2.6. *Let $A \in \mathbb{C}^{n\times n}$ be any (nonzero) matrix, and denote by $\Sigma(A)$ the span of the right singular vectors of A corresponding to the maximal singular value of A. Then $T_1^A(z) = z$ if and only if there exists a vector $w \in \Sigma(A)$ with $w^*Aw = 0$.*

Proof. If $T_1^A(z) = z$, then $\|A\| = \min_{\alpha\in\mathbb{C}} \|A - \alpha I\|$. According to a result of Greenbaum and Gurvits [10, Theorem 2.5], there exists a unit norm vector $w \in \mathbb{C}^n$, so that[2]

$$\min_{\alpha\in\mathbb{C}} \|A - \alpha I\| = \min_{\alpha\in\mathbb{C}} \|(A - \alpha I)w\|.$$

The unique solution of the problem on the right hand side is $\alpha_* = w^*Aw$. Our assumption now implies that $w^*Aw = 0$, and the equations above yield $\|A\| = \|Aw\|$, which shows that $w \in \Sigma(A)$.

On the other hand, if there exists a vector $w \in \Sigma(A)$ such that $w^*Aw = 0$. Without loss of generality we can assume that $\|w\| = 1$. Then

$$\|A\| \geq \min_{\alpha\in\mathbb{C}} \|A - \alpha I\| \geq \min_{\alpha\in\mathbb{C}} \|Aw - \alpha w\| = \min_{\alpha\in\mathbb{C}} (\|Aw\| + \|\alpha w\|) = \|Aw\|.$$

In the first equality we have used that $w^*Aw = 0$, i.e., that the vectors w and Aw are orthogonal. The assumption $w \in \Sigma(A)$ implies that $\|Aw\| = \|A\|$, and thus equality must hold throughout the above relations. In particular, $\|A\| = \min_{\alpha\in\mathbb{C}} \|A - \alpha I\|$, and hence $T_1^A(z) = z$ follows from the uniqueness of the solution. □

An immediate consequence of this result is that if zero is outside the field of values of A, then $\|T_1^A(A)\| < \|A\|$. Note that this also follows from [21, Theorem 5.10], which states that the zeros of $T_m^A(z)$ are contained in the field of values of A.

We will now study the relation between Theorem 2.4 for $m = 1$ and Theorem 2.6. Let $w \in \Sigma(A)$ and let $u \in \mathbb{C}^n$ be a corresponding left singular vector, so that $Aw = \|A\|u$ and $A^*u = \|A\|w$. Then the condition $w^*Aw = 0$ in Theorem 2.6 implies that $w^*u = 0$. We may assume that $\|w\| = \|u\| = 1$. Then the rank-one matrix $Z \equiv uw^*$ satisfies $|||Z||| = 1$,

$$0 = w^*u = \sum_{i=1}^{n} \overline{w}_i u_i = \text{trace}(Z) = \langle Z, I\rangle = \langle Z, A^0\rangle\,,$$

and

$$\langle Z, A\rangle = \text{trace}(A^*uw^*) = \|A\|\,\text{trace}(ww^*) = \|A\| \sum_{i=1}^{n} w_i\overline{w}_i = \|A\|\,.$$

Hence Theorem 2.6 shows that $T_1^A(z) = z$ if and only if there exists a rank-one matrix Z satisfying the conditions (2.8).

Above we have already mentioned that $T_1^A(z) = z$ holds when A is a Jordan block with eigenvalue zero. It is easily seen that, in the notation of Theorem 2.6, the vector w in this case is given by the last canonical basis vector. Furthermore, $T_1^A(z) = z$ holds for any matrix A that satisfies the conditions of Theorem 2.5, i.e., $P^*AP = -A$ or $P^*AP = -A^T$ for some unitary matrix P.

[2]Greenbaum and Gurvits have stated this result for real matrices only, but since its proof mainly involves singular value decompositions of matrices, a generalization to the complex case is straightforward.

An interesting special case of Theorem 2.5 arises when the matrix A is normal, so that

$$\min_{\alpha\in\mathbb{C}} \|A-\alpha I\| = \min_{\alpha\in\mathbb{C}} \max_{\lambda_i\in\Lambda(A)} |\lambda_i-\alpha|.$$

It is well known that the unique solution α_* of this problem is given by the center of the (closed) disk of smallest radius in the complex plane that contains all the complex numbers $\lambda_1,\ldots,\lambda_n$[3].

For nonnormal matrices this characterization of α_* is not true in general. For example, if

$$A = \left[\begin{array}{c|c} J_1 & \\ \hline & -1 \end{array}\right], \quad J_1 \in \mathbb{R}^{4\times 4},$$

then the smallest circle that encloses all eigenvalues of A is centered at zero, but the solution of $\min_{\alpha\in\mathbb{C}} \|A-\alpha I\|$ is given by $\alpha_* \approx -0.4545$, and we have $\|T_1^A(A)\| \approx 1.4545 < \|A\| \approx 1.8794$.

3. Special classes of matrices. In this section we apply our previous general results to Chebyshev polynomials of special classes of matrices.

3.1. Perturbed Jordan blocks. Our first class consists of perturbed Jordan blocks of the form

$$A = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ \nu & & & 0 \end{bmatrix} = \nu(J_0^T)^{n-1} + J_0 \ \in \ \mathbb{C}^{n\times n}, \tag{3.1}$$

where $\nu\in\mathbb{C}$ is a complex parameter. Matrices of this form have recently been studied by Greenbaum in her analysis of upper and lower bounds for the norms of matrix functions [9]. Note that for $\nu=0$ the matrix A is a Jordan block with eigenvalue zero (and hence A is not diagonalizable), while for $\nu=1$ the matrix A is unitary (and hence unitarily diagonalizable), and has the nth roots of unity as its eigenvalues. We have $d(A)=n$ for any $\nu\in\mathbb{C}$.

THEOREM 3.1. *If A is as in (3.1), where $\nu\in\mathbb{C}$ is given, then, for $1\le m\le n-1$,*

$$A^m = \nu(J_0^T)^{n-m} + J_0^m, \quad \|A^m\| = \max\{1,|\nu|\}, \quad \text{and} \quad T_m^A(z) = z^m.$$

Proof. For simplicity of notation we use $J=J_0$ in this proof. Consider an integer s, $0\le s\le n-2$. Then a simple computation yields

$$\begin{aligned} (J^T)^{n-1}J^s + J(J^T)^{n-s} &= (J^T)^{n-(s+1)}\,(J^T)^sJ^s + JJ^T\,(J^T)^{n-(s+1)} \\ &= (J^T)^{n-(s+1)}\,\mathrm{diag}\big(\underbrace{0,\ldots,0}_{s},1,\ldots,1\big) + \\ &\quad\ \mathrm{diag}\,(1,\ldots,1,0)\,(J^T)^{n-(s+1)} \\ &= (J^T)^{n-(s+1)}. \end{aligned} \tag{3.2}$$

[3]The problem of finding this disk, which is uniquely determined either by two or by three of the numbers, was first posed by Sylvester in [19]. This "paper" consists solely of the following sentence: "It is required to find the least circle which shall contain a given set of points in a plane."

We prove the first identity inductively. For $m = 1$ the statement is trivial. Suppose now that the assertion is true for some m, $1 \le m \le n-2$. Then

$$\begin{aligned} A^{m+1} &= (\nu(J^T)^{n-1} + J)\,(\nu(J^T)^{n-m} + J^m) \\ &= \nu^2(J^T)^{2n-m-1} + \nu((J^T)^{n-1}J^m + J(J^T)^{n-m}) + J^{m+1} \\ &= \nu(J^T)^{n-(m+1)} + J^{m+1}, \end{aligned}$$

where in the last equality we have used (3.2).

To prove the second identity it is sufficient to realize that each row and column of A^m contains at most one nonzero entry, either ν or 1. Therefore, $\|A^m\| = \max\{1, |\nu|\}$.

Finally, note that the matrices $I, A, \ldots, A^{n-1}$ have non-overlapping nonzero patterns. Therefore, for any $p \in \mathcal{M}_m$, $1 \le m \le n-1$, at least one entry of $p(A)$ is 1 and at least one entry is ν, so $\|p(A)\| \ge \max\{1, |\nu|\}$. On the other hand, we know that $\|A^m\| = \max\{1, |\nu|\}$, and uniqueness of $T_m^A(z)$ implies that $T_m^A(z) = z^m$. $\square$

3.2. Special bidiagonal matrices. Let positive integers ℓ and h, and ℓ complex numbers $\lambda_1, \ldots, \lambda_\ell$ (not necessarily distinct) be given. We consider the matrices

$$D = \begin{bmatrix} \lambda_1 & 1 & & \\ & \lambda_2 & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_\ell \end{bmatrix} \in \mathbb{C}^{\ell\times\ell}, \quad E = (J_0^T)^{\ell-1} \in \mathbb{R}^{\ell\times\ell}, \tag{3.3}$$

and form the block Toeplitz matrix

$$B = \begin{bmatrix} D & E & & \\ & D & \ddots & \\ & & \ddots & E \\ & & & D \end{bmatrix} \in \mathbb{C}^{\ell\cdot h\times\ell\cdot h}. \tag{3.4}$$

Matrices of the form (3.4) have been used by Reichel and Trefethen [14], who related the pseudospectra of these matrices to their symbol $f_B(z) = D + zE$. Chebyshev polynomials for examples of such matrices have been studied numerically in [21, 24, 26] (cf. our examples following Theorem 3.3).

LEMMA 3.2. *In the notation established above, $\chi_D(B) = J_0^\ell$, where $\chi_D(z) = (z-\lambda_1)\cdot\ldots\cdot(z-\lambda_\ell)$ is the characteristic polynomial of D.*

Proof. Let $e_1, \ldots, e_{\ell\cdot h}$ denote the canonical basis vectors of $\mathbb{C}^{\ell\cdot h}$, and let $e_0 = e_{-1} = \cdots = e_{-\ell+1} = 0$. It then suffices to show that $\chi_D(B)e_j = e_{j-\ell}$, for $j = 1, 2, \ldots, \ell\cdot h$, or, equivalently, that

$$\chi_D(B)e_{k\cdot\ell+j} = e_{(k-1)\cdot\ell+j}, \quad k = 0, 1, \ldots, h-1, \quad j = 1, 2, \ldots, \ell. \tag{3.5}$$

To prove these relations, note that

$$\chi_D(B) = (B - \lambda_1 I)\cdot\ldots\cdot(B - \lambda_\ell I),$$

where the factors on the right hand side commute. Consider a fixed j between 1 and ℓ. Then it follows directly from the structure of the matrix $B - \lambda_j I$, that

$$(B - \lambda_j I)\, e_{k\cdot\ell+j} \;=\; e_{k\cdot\ell+j-1}, \quad k = 0, 1, \ldots, h-1.$$

Consequently, for $k = 0, 1, \dots, h-1$, and $j = 1, 2, \dots, \ell$,

$$\begin{aligned}\chi_D(B)\, e_{k\cdot\ell+j} &= (B-\lambda_{j+1}I)\cdot\ldots\cdot(B-\lambda_\ell I)\cdot(B-\lambda_1 I)\cdot\ldots\cdot(B-\lambda_j I)\, e_{k\cdot\ell+j}\\ &= (B-\lambda_{j+1}I)\cdot\ldots\cdot(B-\lambda_\ell I)\, e_{k\cdot\ell}\\ &= (B-\lambda_{j+1}I)\cdot\ldots\cdot(B-\lambda_\ell I)\, e_{(k-1)\cdot\ell+\ell}\\ &= e_{(k-1)\cdot\ell+j},\end{aligned}$$

which is what we needed to show. □

This lemma allows us to derive the following result on the Chebyshev polynomials of the matrix B.

THEOREM 3.3. *Let B be defined as (3.4), and let $\chi_D(z)$ be the characteristic polynomial of D. Then $T^B_{k\cdot\ell}(z) = (\chi_D(z))^k$ for $k = 1, 2, \dots, h-1$.*

Proof. Let M_{ij} denote the entry at position (i, j) of the matrix M. A well known property of the matrix 2-norm is $\|M\| \geq \max_{i,j} |M_{ij}|$. For any $p \in \mathcal{M}_{k\cdot\ell}$ we therefore have

$$\|p(B)\| \;\geq\; \max_{i,j} |p(B)_{ij}| \;\geq\; |p(B)_{1,k\cdot\ell+1}| \;=\; 1.$$

On the other hand, Lemma 3.2 implies that

$$\|(\chi_D(B))^k\| \;=\; \|J_0^{k\cdot\ell}\| \;=\; 1.$$

Hence the polynomial $(\chi_D(z))^m$ attains the lower bound on $\|p(B)\|$ for all $p \in \mathcal{M}_{k\cdot\ell}$. The uniqueness of the Chebyshev polynomial of B now implies the result. □

In case $\ell = 1$, i.e. $B = J_{\lambda_1} \in \mathbb{C}^{n\times n}$, the theorem shows that $(z-\lambda_1)^m$ is the mth Chebyshev polynomial of B, $m = 1, \dots, n-1$. As mentioned above, this result was previously shown in [13, Theorem 3.4]. The proof in that paper, however, is based on a different approach, namely a characterization of matrix approximation problems in the 2-norm obtained by Ziętak [27, 28].

As a further example consider a matrix B of the form (3.4) with

$$D = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}. \tag{3.6}$$

This matrix B has been studied numerically in [25, Example 6] and [21, Example 6]. The minimal polynomial of D is given by $(z-1)(z+1) = z^2 - 1$, and hence $T^B_{2k}(z) = (z^2-1)^k$ for $k = 1, 2, \dots, h-1$. However, there seems to be no simple closed formula for the Chebyshev polynomials of B of odd degree. Our numerical experiments show that these polynomials (on the contrary to those of even degree) depend on the size of the matrix. Table 3.1 shows the coefficients of $T^B_m(z)$ for $m = 1, 2, \dots, 7$ for an (8×8)-matrix B (i.e., there are four blocks D of the form (3.6) on the diagonal of B). The coefficients in the rows of the table are ordered from highest to lowest. For example, $T^B_4(z) = z^4 - 2z^2 + 1$.

It is somewhat surprising that the Chebyshev polynomials change significantly when we reorder the eigenvalues on the diagonal of B. In particular, consider

$$\widetilde{B} = \begin{bmatrix} J_1 & E \\ & J_{-1} \end{bmatrix} \;\in\; \mathbb{R}^{2\ell\times 2\ell}, \tag{3.7}$$

TABLE 3.1
Coefficients of $T_m^B(z)$ for an (8×8)-matrix B of the form (3.4) with D as in (3.6).

m								
1	1	0						
2	1	0	-1.000000					
3	1	0	0.876114	0				
4	1	0	-2.000000	0	1.000000			
5	1	0	-1.757242	0	0.830598	0		
6	1	0	-3.000000	0	3.000000	0	-1.000000	
7	1	0	-2.918688	0	2.847042	0	0.927103	0

TABLE 3.2
Coefficients of $T_m^{\widetilde{B}}(z)$ for an (8×8)-matrix $\widetilde{B}$ of the form (3.7).

m								
1	1	0						
2	1	0	-1.595438					
3	1	0	-1.975526	0				
4	1	0	-2.858055	0	2.463968			
5	1	0	-3.125673	0	2.608106	0		
6	1	0	-3.771773	0	4.945546	0	-1.863541	
7	1	0	-4.026082	0	5.922324	0	-3.233150	0

where $E = (J_0^T)^{\ell-1} \in \mathbb{R}^{\ell\times\ell}$. The coefficients of $T_m^{\widetilde{B}}(z)$, $m = 1, 2, \ldots, 7$, for an (8×8)-matrix of the form (3.7) are shown in Table 3.2.

Note that the matrices B based on (3.6) and $\widetilde{B}$ in (3.7) are similar (when they are of the same size). Another matrix similar to these two is the matrix C in (2.10) with $c = 1$. The coefficients of Chebyshev polynomials of such a matrix C of size 8×8 are shown in Table 3.3. It can be argued that the 2-norm condition number of the similarity transformations between B, $\widetilde{B}$ and C is of order 2^ℓ (we skip details for brevity of the presentation). Hence this transformation is far from being orthogonal, which indicates that the Chebyshev polynomials of the respective matrices can be very different – and in fact they are. We were unable to determine a closed formula for any of the nonzero coefficients of the Chebyshev polynomials of $\widetilde{B}$ and C (except, of course the leading one). Numerical experiments indicate that these in general depend on the sizes of the respective matrices.

In Figure 3.1 we show the roots of the Chebyshev polynomials of degrees $m = 5$ and $m = 7$ corresponding to the examples in Tables 3.1–3.3. Each figure contains three sets of roots. All the polynomials are odd, and therefore all of them have one root at the origin.

4. Matrices and sets in the complex plane. In this section we explore the relation between Chebyshev polynomials of matrices and of compact sets Ω in the complex plane. Recall that for each $m = 1, 2, \ldots$ the problem

$$\min_{p\in\mathcal{M}_m} \max_{z\in\Omega} |p(z)|$$

has a unique solution $T_m^\Omega(z)$, that is called the mth Chebyshev polynomial of Ω (cf. our Introduction). Similarly to the matrix case, Chebyshev polynomials of sets are known explicitly only in a few special cases. One of these cases is a disk in the

TABLE 3.3
Coefficients of $T_m^C(z)$ for an (8×8)-matrix C of the form (2.10) with $\lambda = 1$.

m								
1	1	0						
2	1	0	-1.763931					
3	1	0	-2.194408	0				
4	1	0	-2.896537	0	2.502774			
5	1	0	-3.349771	0	3.696082	0		
6	1	0	-3.799998	0	5.092302	0	-1.898474	
7	1	0	-4.066665	0	6.199999	0	-4.555546	0

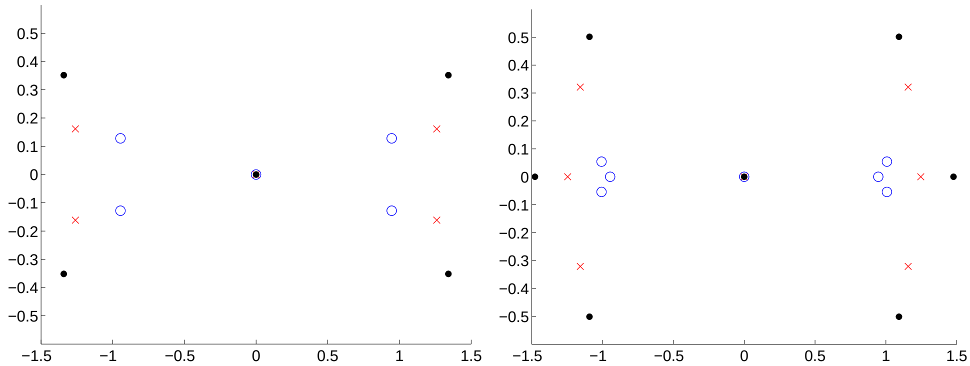

FIG. 3.1. *Roots of $T_m^B(z)$ (blue circles), $T_m^{\widetilde{B}}(z)$ (red crosses) and $T_m^C(z)$ (black points) of degrees $m = 5$ (left) and $m = 7$ (right) corresponding to the examples in Tables 3.1–3.3.*

complex plane centered at the point $\lambda \in \mathbb{C}$, for which the mth Chebyshev polynomial is $(z-\lambda)^m$; see, e.g., [17, p. 352]. Kamo and Boronin [12] allow us to generate more examples of Chebyshev polynomials.

THEOREM 4.1. *Let T_k^Ω be the kth Chebyshev polynomial of the infinite compact set $\Omega \subset \mathbb{C}$, let $p(z) = a_\ell z^\ell + \cdots + a_1 z + a_0$, $a_\ell \neq 0$, be a polynomial of degree ℓ, and let*

$$\Psi \;\equiv\; p^{-1}(\Omega) \;=\; \{z \in \mathbb{C} \,:\, p(z) \in \Omega\}$$

be the pre-image of Ω under the polynomial map p. Then $T_{k\cdot\ell}^\Psi$, the Chebyshev polynomial of degree $m = k \cdot \ell$ of the set Ψ, is given by

$$T_m^\Psi(z) \;=\; \frac{1}{a_\ell^k}\, T_k^\Omega(p(z))\,.$$

This result has been shown also by Fischer and Peherstorfer [7, Corollary 2.2], who applied it to obtain convergence results for Krylov subspace methods. Similar ideas can be used in our context. For example, let $\mathcal{S}_A = [a, b]$ with $0 < a < b$ and $p(z) = z^2$. Then

$$\mathcal{S}_B \equiv p^{-1}(\mathcal{S}_A) = [-\sqrt{a}, -\sqrt{b}] \cup [\sqrt{a}, \sqrt{b}],$$

and Theorem 4.1 implies that $T_{2k}^{\mathcal{S}_B}(z) = T_k^{\mathcal{S}_A}(z^2)$. Such relations are useful when studying two normal matrices A and B, whose spectra are contained in the sets $\mathcal{S}_A$ and $\mathcal{S}_B$, respectively.

For an application of Theorem 4.1 that to our knowledge has not been considered before, consider a given polynomial $p = (z-\lambda_1)\cdot\ldots\cdot(z-\lambda_\ell) \in \mathcal{M}_\ell$ and the *lemniscatic region*

$$\mathcal{L}(p) \equiv \{z \in \mathbb{C} : |p(z)| \leq 1\}. \tag{4.1}$$

Note that $\mathcal{L}(p)$ is the pre-image of the unit disk under the polynomial map p. Since the kth Chebyshev polynomial of the unit disk is the polynomial z^k, Theorem 4.1 implies that

$$T_{k\cdot\ell}^{\mathcal{L}(p)} = (p(z))^k.$$

Using these results and Theorem 3.3 we can now formulate the following.

THEOREM 4.2. *Let $\lambda_1, \ldots, \lambda_\ell \in \mathbb{C}$ and an integer $h > 1$ be given. Then for $p(z) = (z-\lambda_1)\cdot\ldots\cdot(z-\lambda_\ell) \in \mathcal{M}_\ell$, and each $k = 1, 2, \ldots, h-1$,*

$$(p(z))^k \;=\; T_{k\cdot\ell}^{\mathcal{L}(p)}(z) \;=\; T_{k\cdot\ell}^{B}(z),$$

where the lemniscatic region $\mathcal{L}(p)$ is defined as in (4.1), and the matrix B is of the form (3.4). Moreover,

$$\max_{z\in\mathcal{L}(p)} |T_{k\cdot\ell}^{\mathcal{L}(p)}(z)| \;=\; \|T_{k\cdot\ell}^{B}(B)\|.$$

This theorem connects Chebyshev polynomials of lemniscatic regions of the form (4.1) to Chebyshev polynomials of matrices B of the form (3.4). The key observation is the analogy between Theorems 3.3 and 4.1. We believe that it is possible to generate further examples along these lines.

5. Concluding remarks.

We have shown that Chebyshev polynomials of matrices and Chebyshev polynomials of compact sets in the complex plane have a number of common or at least related properties. Among these are the polynomials' behavior under shifts and scalings (of matrix or set), and certain "alternation" and even/odd properties. Progress on the theory of Chebyshev polynomials of matrices can certainly be made by studying other known characteristics of their counterparts of sets in the complex plane. Furthermore, we consider it promising to further explore whether the Chebyshev polynomials of a matrix can be related to Chebyshev polynomials of a set and vice versa (see Theorem 4.2 for an example). This may give additional insight into the question where a matrix "lives" in the complex plane.

Acknowledgments. We thank two anonymous referees for helpful comments and suggestions that improved the paper, and particulary led to a simplification of our original proof of Theorem 3.1.

REFERENCES

[1] M. Afanasjew, M. Eiermann, O. G. Ernst, and S. Güttel, *A generalization of the steepest descent method for matrix functions*, Electron. Trans. Numer. Anal., 28 (2007/08), pp. 206–222.

[2] W. E. Arnoldi, *The principle of minimized iteration in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., 9 (1951), pp. 17–29.

[3] Steve Benson, Yinyu Ye, and Xiong Zhang, *DSDP – Software for Semidefinite Programming, v. 5.8.* January 2006.

[4] E. K. Blum, *Numerical Analysis and Computation: Theory and Practice*, Addison-Wesley, Reading, M.A., 1972.

[5] P. L. Chebyshev, *Sur le questions de minima qui se rattachent à la représentation approximative des fonctions*, Mém. de l'Acad. de St. Pétersbourg, série VI, t. VII (1859), pp. 199–291.

[6] Ronald A. DeVore and George G. Lorentz, *Constructive Approximation*, vol. 303 of Grundlehren der mathematischen Wissenschaften, Springer, New York, 1993.

[7] Bernd Fischer and Franz Peherstorfer, *Chebyshev approximation via polynomial mappings and the convergence behaviour of Krylov subspace methods*, Electron. Trans. Numer. Anal., 12 (2001), pp. 205–215 (electronic).

[8] Shmuel Friedland, *On matrix approximation*, Proc. Amer. Math. Soc., 51 (1975), pp. 41–43.

[9] Anne Greenbaum, *Upper and lower bounds on norms of functions of matrices*, Linear Algebra Appl., 430 (2009), pp. 52–65.

[10] A. Greenbaum and L. Gurvits, *Max-min properties of matrix factor norms*, SIAM J. Sci. Comput., 15 (1994), pp. 348–358.

[11] Anne Greenbaum and Lloyd N. Trefethen, *GMRES/CR and Arnoldi/Lanczos as matrix approximation problems*, SIAM J. Sci. Comput., 15 (1994), pp. 359–368. Iterative methods in numerical linear algebra (Copper Mountain Resort, CO, 1992).

[12] S. O. Kamo and P. A. Borodin, *Chebyshev polynomials for Julia sets*, Vestnik Moskov. Univ. Ser. I Mat. Mekh., (1994), pp. 65–67.

[13] Jörg Liesen and Petr Tichý, *On best approximations of polynomials in matrices in the matrix 2-norm*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 853–863.

[14] Lothar Reichel and Lloyd N. Trefethen, *Eigenvalues and pseudo-eigenvalues of Toeplitz matrices*, Linear Algebra Appl., 162/164 (1992), pp. 153–185. Directions in matrix theory (Auburn, AL, 1990).

[15] Youcef Saad, *Projection methods for solving large sparse eigenvalue problems*, in Matrix Pencils, B. Kågström and A. Ruhe, eds., vol. 973 of Lecture Notes in Mathematics, Springer, Berlin, 1982, pp. 121–144.

[16] Yousef Saad, *Numerical Methods for Large Eigenvalue Problems*, Manchester University Press, Manchester, UK, 1992.

[17] V. I. Smirnov and N. A. Lebedev, *Functions of a complex variable: Constructive theory*, Translated from the Russian by Scripta Technica Ltd, The M.I.T. Press, Cambridge, Mass., 1968.

[18] Karl-Georg Steffens, *The History of Approximation Theory*, Birkhäuser, Boston, 2006. From Euler to Bernstein.

[19] J. J. Sylvester, *A question in the geometry of situation*, The Quarterly Journal of Pure and Applied Mathematics, 1 (1857), p. 79.

[20] The MathWorks, Inc., *MATLAB 7.9 (R2009b).* Natick, Massachusetts, USA, 2009.

[21] K.-C. Toh, *Matrix Approximation Problems and Nonsymmetric Iterative Methods*, PhD thesis, Cornell University, Ithaca, N.Y., 1996.

[22] K. C. Toh, *GMRES vs. ideal GMRES*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 30–36.

[23] K. C. Toh, M. J. Todd, and R. H. Tütüncü, *SDPT3 version 4.0 (beta) – a MATLAB software for semidefinite-quadratic-linear programming.* February 2009.

[24] Kim-Chuan Toh and L. N. Trefethen, *The Chebyshev polynomials of a matrix*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 400–419.

[25] Kim-Chuan Toh and Lloyd N. Trefethen, *The Chebyshev polynomials of a matrix*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 400–419 (electronic).

[26] Lloyd N. Trefethen and Mark Embree, *Spectra and pseudospectra*, Princeton University Press, Princeton, NJ, 2005. The behavior of nonnormal matrices and operators.

[27] K. Ziętak, *Properties of linear approximations of matrices in the spectral norm*, Linear Algebra Appl., 183 (1993), pp. 41–60.

[28] Krystyna Ziętak, *On approximation problems with zero-trace matrices*, Linear Algebra Appl., 247 (1996), pp. 169–183.