The Chebyshev Polynomials of a Matrix

Kim-Chuan Toh
Lloyd N. Trefethen

May 8, 1996

Abstract

A Chebyshev polynomial of a square matrix A is a monic polynomial of specified degree that minimizes $\|p(A)\|_2$. The study of such polynomials is motivated by the analysis of Krylov subspace iterations in numerical linear algebra. An algorithm is presented for computing these polynomials based on reduction to a semidefinite program which is then solved by a primal-dual interior point method. Examples of Chebyshev polynomials of matrices are presented, and it is noted that if A is far from normal, the lemniscates of these polynomials tend to approximate pseudospectra of A.

1 Introduction

Let A be an $N \times N$ matrix and n a nonnegative integer. The degree n *Chebyshev polynomial of* A is the unique monic polynomial p_n^* of degree n such that

$$\|p_n^*(A)\| = \text{minimum}, \tag{1}$$

where $\|\cdot\|$ denotes the matrix 2-norm. To be precise, p_n^* is unique provided that n is less than or equal to the degree of the minimal polynomial of A; otherwise we have $p_n^*(A) = 0$ and the problem ceases to be interesting.

This notion of the polynomial that minimizes $\|p(A)\|$ seems so simple and natural that one would expect it to be a standard one. We suspect it may have been considered before, perhaps decades ago in the literature of approximation theory. Nevertheless, we have been unable to find any literature on this problem before our 1994 paper with Greenbaum [7]. In that paper, Chebyshev polynomials of matrices are defined and it is proved that they exist (obvious by compactness) and that they are unique under the condition just mentioned (not obvious).

Even if they are not discussed explicitly, Chebyshev polynomials of matrices are never far away from any discussion of convergence of Krylov subspace iterations in numerical linear algebra. For these iterations, convergence depends on certain vector norms $\|p(A)q\|$ being as small or nearly as small as possible, where q is a starting vector. Most of the convergence properties of applied interest are roughly independent of q, and thus it is the near-minimality of $\|p(A)\|$ that is often the heart of the matter. For finding eigenvalues, the principal iterative method in this category is the Arnoldi iteration, which becomes the Lanczos iteration if A is real and symmetric. For solving systems of equations, the analogous methods include GMRES, biconjugate gradients, CGS, QMR, and Bi-CGSTAB in the general case and conjugate gradients if A is symmetric positive definite. (For systems of equations, the notion of a Chebyshev polynomial of A should be normalized differently by the condition $p(0) = 1$ instead of the condition that p is monic. In [7], the Chebyshev polynomial of a matrix is called the *ideal Arnoldi polynomial,* and its analogue with this other normalization is called the *ideal GMRES polynomial.*)

The motivation for the term "Chebyshev polynomials of a matrix" is as follows. All readers will be familiar with the classical Chebyshev polynomials $\{T_n\}$, which are polynomials of minimal $\|\cdot\|_\infty$-norm on the interval $[-1, 1]$. This notion was generalized by Faber in 1920 to the idea of the *Chebyshev polynomials of* S, where S is a compact set in the complex plane $\mathbb{C}$: the monic polynomials of minimal $\|\cdot\|_\infty$-norm over S [5], [24]. Now suppose that A is a hermitian or more generally a normal matrix, having a complete set of orthogonal eigenvectors. Then by a unitary reduction to diagonal form, it is easily shown that the nth Chebyshev polynomial of A as defined by (1) is precisely the nth Chebyshev polynomial of S in this latter sense, where S is the spectrum of A. Such a polynomial can be computed, for example, by generalizations of the Remez algorithm [15].

Chebyshev polynomials of normal matrices, then, are trivial; the matrix problem reduces to a scalar problem. But what if A is an arbitrary square matrix, with non-orthogonal eigenvectors or perhaps no complete set of eigenvectors? This is the subject of this paper, and our purpose is two-fold.

First, we describe an algorithm for computing Chebyshev polynomials of matrices. The optimization problem implicit in (1) is far from smooth, and unless the degree is small, these problems are usually difficult. The algorithm we describe is based on interior point methods for semidefinite programming, and we believe it is the first to have been developed for this problem. With this algorithm, we can reliably compute Chebyshev polynomials for matrices of order ≤ 50 in less than a minute on workstations available in 1996.

Second, we present computed examples, the first we know of to have been published. A few numerical coefficients are listed for possible comparison by later authors, but our main aim is to give insight into the behavior of Chebyshev polynomials of matrices, largely with the aid of pictures. A natural question is, how are the coefficients of the polynomials affected by the degree and nature of the non-normality of A? For a partial answer, we plot lemniscates $|p_n^*(z)| = \text{constant}$ of our polynomials and find that they tend to approximate pseudospectra of A, even in certain highly non-normal cases.

2 Reduction to a semidefinite program

Let $\{B_0, B_1, \ldots, B_n\}$ be a linearly independent set of matrices in $\mathbb{C}^{N\times N}$. The Chebyshev problem (1) is a special case of a norm minimization problem involving linear functions of matrices:

$$\min_{x\in\mathbb{C}^n} \|\sum_{k=1}^{n} x_k B_k \; - \; B_0\|. \tag{2}$$

For our special case, $B_0 = A^n$ and

$$B_k \; = \; A^{k-1}, \quad k = 1, \ldots, n, \tag{3}$$

and the numbers x_k are the coefficients (actually their negatives) of the Chebyshev polynomial of A.

It is well known that (2) can be expressed as a semidefinite program [11], [22]. We shall not show in detail how this is done. One difference between our work and what has been done before is that the existing literature, as far as we are aware, considers only real matrices.

Proposition 2.1 *The norm minimization problem* (2) *is equivalent to the following semidefinite program involving hermitian matrices:*

$$- \max_{x \in \mathbb{C}^n, \lambda \in R} \lambda$$

$$\text{s.t.} \quad \sum_{k=1}^{n} (\alpha_k A_k + \beta_k A_{n+k}) + \lambda A_{2n+1} + Z = A_0, \qquad Z \geq 0, \tag{4}$$

where $\alpha_k = \mathrm{Re}(x_k)$*,* $\beta_k = \mathrm{Im}(x_k)$*,*

$$A_{2n+1} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}, \quad A_0 = \begin{pmatrix} 0 & B_0 \\ B_0^* & 0 \end{pmatrix},$$

$$A_k = \begin{pmatrix} 0 & B_k \\ B_k^* & 0 \end{pmatrix}, \quad A_{n+k} = \begin{pmatrix} 0 & iB_k \\ -iB_k^* & 0 \end{pmatrix}, \tag{5}$$

$k = 1, \ldots, n$*, and* $Z \geq 0$ *means that* Z *is positive semidefinite.*

Proof. The norm minimization problem (2) is equivalent to the problem of minimizing $-\lambda$ such that

$$\left\| \sum_{k=1}^{n} x_k B_k - B_0 \right\| \leq -\lambda. \tag{6}$$

Using the fact that for any $M \in \mathbb{C}^{N \times N}$,

$$\|M\| = \lambda_{\max}\left(\begin{bmatrix} 0 & M \\ M^* & 0 \end{bmatrix} \right), \tag{7}$$

where $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue, equation (6) can be rewritten as

$$\lambda_{\max}\left(\begin{bmatrix} 0 & B(x) \\ B(x)^* & 0 \end{bmatrix} \right) \leq -\lambda, \tag{8}$$

where $B(x) = \sum_{k=1}^{n} x_k B_k - B_0$. But (8) is equivalent to

$$\begin{pmatrix} 0 & B(x) \\ B(x)^* & 0 \end{pmatrix} + \lambda \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} + Z \;=\; 0, \quad Z \geq 0. \tag{9}$$

Now, by writing (9) out in full, we get (4). □

3 Transformation to a better-conditioned basis

Before we discuss how the semidefinite program (4) can be solved by interior point methods, we must first address the issue of change of basis in (2). This is an essential point for the computation of Chebyshev polynomials of matrices. Since the power basis is usually highly ill-conditioned, it has to be replaced by a better-conditioned alternative for numerical stability.

Suppose $\{\widehat{B}_0, \widehat{B}_1, \ldots, \widehat{B}_n\}$ is another linearly independent set of matrices in $\mathbb{C}^{N\times N}$ related linearly to $\{B_0, B_1, \ldots, B_n\}$ by

$$\begin{aligned} [\, B_1 \,|\, \cdots \,|\, B_n \,] &= \left[\, \widehat{B}_1 \,|\, \cdots \,|\, \widehat{B}_n \,\right] T, \\ B_0 &= c\,\widehat{B}_0 - \left[\, \widehat{B}_1 \,|\, \cdots \,|\, \widehat{B}_n \,\right] t, \end{aligned}$$

where T is an $n \times n$ nonsingular matrix, t is an n-vector, and c is a scalar. The following proposition describes how (2) is modified by this change of basis. The proof is straightforward, and we shall omit it.

Proposition 3.1 *The minima*

$$\min_{x \in \mathbb{C}^n} \| \sum_{k=1}^{n} x_k B_k \; - \; B_0 \| \tag{10}$$

and

$$|c| \min_{\hat{x} \in \mathbb{C}^n} \| \sum_{k=1}^{n} \hat{x}_k \widehat{B}_k \; - \; \widehat{B}_0 \| \tag{11}$$

are the same, and the unique vectors x and $\hat{x}$ that achieve these minima are related by

$$\hat{x} \;=\; \frac{1}{c}(Tx + t). \tag{12}$$

We are aware of three choices of basis that are particularly attractive for practical computations.

Scaled power basis. Suppose $B_0, \ldots, B_n$ are given as in (3). A simple way to get a better conditioned basis is to scale the norm of A to 1. With such a scaling, we have $\widehat{B}_0 = B_0/\alpha_0$ and

$$\widehat{B}_k = B_k/\alpha_k, \quad k = 1, \ldots, n,$$

where $\alpha_0 = \|A\|^n$ and $\alpha_k = \|A\|^{k-1}$, $k = 1, \ldots, n$. Hence $T = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$, $t = 0$, and $c = \alpha_0$ in (12).

Faber polynomial basis. Even the best-scaled power basis is often highly ill-conditioned. A more powerful idea is to consider the basis $\{F_0(A), \ldots, F_n(A)\}$ defined by the Faber polynomials $F_0, \ldots, F_n$ associated with some region Ω in the complex plane containing the spectrum of A. The Faber polynomials $\{F_n\}$ are the natural analogues for a general region Ω in $\mathbb{C}$ of the monomials $\{z^n\}$ for the unit disk or the Chebyshev polynomials $\{T_n\}$ for $[-1, 1]$; see [2]. In most cases, $\{F_n(A)\}$ will be far better conditioned than any power basis.

For the Faber basis, the matrix T in (12) is an upper triangular matrix whose columns are the coefficients of the Faber polynomials $F_0, \ldots, F_{n-1}$. The scalar c is the positive number $\mathrm{cap}(\Omega)^n$, where $\mathrm{cap}(\Omega)$ is the logarithmic capacity of Ω. The vector t is the vector of coefficients of the expansion of the degree $n-1$ polynomial $cF_n(z) - z^n$ in terms of $F_0(z), \ldots, F_{n-1}(z)$.

Of course one must choose a region Ω for which the associated Faber polynomials can be obtained either analytically or numerically. If Ω is chosen to be an ellipse or an interval, then the Faber polynomials are simply the scaled Chebyshev polynomials $\{T_n\}$. More generally, if Ω is chosen to be a polygonal domain, the Faber polynomials can be computed numerically via Schwarz-Christoffel mapping. We have used the MATLAB Schwarz-Christoffel Toolbox for this purpose, due to Driscoll [4].

Orthonormal basis. Finally, our third idea is a more elementary one, but powerful in practice. One may simply orthonormalize the power basis $\{I, A, \ldots, A^n\}$ with respect to the "trace inner product" $\langle A, B\rangle = \mathrm{tr}(AB^*)$ in $\mathbb{C}^{N\times N}$ to obtain a basis $\{Q_1, Q_2, \ldots, Q_{n+1}\}$ that is typically well-conditioned even in the 2-norm. This can be done by a modified Gram-Schmidt procedure similar to that used in the Arnoldi iteration:

```
$Q_1 = N^{-1/2} I$
for $k = 1 : n$
    $V = A Q_k$
    for $j = 1 : k$
        $h_{jk} = \langle V, Q_j \rangle$
        $V = V - h_{jk} Q_j$
    $h_{k+1,k} = \langle V, V \rangle^{1/2}$
    $Q_{k+1} = V / h_{k+1,k}$
```

To obtain the matrix T in (12), we note that there is a unique $(n+1) \times (n+1)$ upper triangular matrix R such that

$$[\, I \,|\, A \,|\, \cdots \,|\, A^n \,]\; R \;=\; [\, Q_1 \,|\, Q_2 \,|\, \cdots \,|\, Q_{n+1} \,],$$

and the columns of R can be computed from the following recurrence relation (in MATLAB notation):

$$\begin{aligned} R(1,1) &= 1/\sqrt{N}, \\ h_{k+1,k}\, \mathbf{r_{k+1}} &= \begin{bmatrix} 0 \\ \mathbf{r_k} \end{bmatrix} - \begin{bmatrix} R(1:k, 1:k)\, \mathbf{h_k} \\ 0 \end{bmatrix}, \end{aligned}$$

where $\mathbf{r_k} = R(1:k, k)$ and $\mathbf{h_k} = (h_{1k}, \ldots, h_{kk})^T$, for $k = 1, \ldots, n$. It is now easy to see that

$$T \;=\; R^{-1}(1:n, 1:n), \quad c \;=\; R^{-1}(n+1, n+1), \quad t \;=\; R^{-1}(1:n, n+1),$$

again in MATLAB notation.

For simplicity, we use the orthonormal basis in the examples reported in this paper. Although it is more expensive to compute than the other two bases, the amount of time taken remains small compared to the time required for solving (4).

4 Solution by primal-dual interior point method

Assuming a suitable basis has been chosen, we now turn to the problem of how (4) can be solved by interior point methods similar to those in linear programming, specifically, by Mehrotra-type primal-dual predictor-corrector algorithms. Extensive research has been done on both the algorithms and the theory of SDP. We refer the reader to [1], [9], [10], [11], [12], [16], [22] and [25] for details.

A general SDP has the form:

$$
\begin{aligned}
(D): \quad & \max_{y \in \mathbb{R}^n} \; b^T y \\
\text{s.t.} \quad & \sum_{k=1}^{n} y_k A_k \; + \; Z \; = \; C, \qquad Z \; \geq \; 0,
\end{aligned}
\tag{13}
$$

where C, Z, A_k, $k = 1, \ldots, n$, are hermitian matrices and $b \in I\!R^n$. The idea behind an interior point method is to use a suitable barrier function, $-\log\det(Z)$ in the case of SDP, to transform the semidefinite constrained convex problem (D) into a parametrized family (by μ) of equality constrained convex problems whose optimal solutions $(X(\mu), y(\mu), Z(\mu))$ satisfy the optimality conditions

$$
\begin{aligned}
\sum_{k=1}^{n} y_k A_k \; + \; Z \; &= \; C, \\
\langle X, \, A_k \rangle \; &= \; b_k, \quad k = 1, \ldots, n, \\
XZ \; &= \; \mu I,
\end{aligned}
\tag{14}
$$

where X and Z are hermitian positive definite. The parameter $\mu > 0$ is to be driven explicitly to zero (as fast as possible), and in the limit $\mu \to 0$, an optimal solution of (13) is obtained.

Mehrotra-type primal-dual predictor-corrector algorithms essentially consist of a sequence of modified Newton iterations. Usually, one step of Newton's iteration is applied to (14) for each new μ.

It is readily shown that application of Newton's method to (14) gives rise to the equations

$$
\begin{aligned}
\textstyle\sum_{k=1}^{n} (\Delta y)_k \, A_k \; + \; \Delta Z \; &= \; C \; - \; Z - \textstyle\sum_{k=1}^{n} \, y_k \, A_k, \\
\langle \Delta X, \, A_k \rangle \; &= \; b_k \; - \; \langle X, \, A_k \rangle, \quad k = 1, \ldots, n,
\end{aligned}
\tag{15}
$$

$$
\Delta X \, Z \; + \; X \, \Delta Z \; = \; \mu I \; - \; XZ. \tag{16}
$$

In order to keep ΔX hermitian (this is desirable since the fundamental objects in an SDP are hermitian matrices), the equation (16) is usually symmetrized with respect to an invertible matrix P, whereupon it becomes

$$
P(\Delta X \, Z + X \, \Delta Z)P^{-1} \; + \; P^{-*}(Z \, \Delta X + \Delta Z \, X)P^{*} \; = \; R, \tag{17}
$$

where

$$
R \; = \; 2\mu I \; - \; P(XZ)P^{-1} \; - \; P^{-*}(ZX)P^{*}. \tag{18}
$$

Different choices of P give rise to different Newton steps. For example, $P = I$ gives rise to what is known as the Alizadeh-Haeberly-Overton (AHO) direction [1]; $P = Z^{1/2}$ gives rise to the Monteiro direction [10]; and $P = W^{-1/2}$, where $W = Z^{-1/2}(Z^{1/2}XZ^{1/2})^{1/2}Z^{-1/2}$, gives rise to the Nesterov-Todd (NT) direction [12].

The general algorithmic framework of a Mehrotra-type predictor-corrector method is as follows.

Algorithm. Given an initial iterate (X^0, y^0, Z^0) with X^0, Z^0 positive definite, for $k = 0, 1, \ldots,$
(Let the current and the next iterate be (X, y, Z) and (X^+, y^+, Z^+), respectively.)

- (Predictor step)
  Compute the Newton step $(\delta X, \delta y, \delta Z)$ from (15) and (17) with $\mu = 0$ in (18).

- Determine the real parameter $\mu = \sigma \langle X,\ Z\rangle / n$, where

  $$\sigma \;=\; \frac{\langle X + \alpha\,\delta X,\ Z + \beta\,\delta Z\rangle^2}{\langle X,\ Z\rangle^2}.$$

  Here α and β are suitable steplengths chosen to ensure that $X + \alpha\,\delta X$ and $Z + \beta\,\delta Z$ are positive definite. Generally, α and β have the form

  $$\alpha \;=\; \min\left(1, \frac{-\tau}{\lambda_{\min}(X^{-1}\delta X)}\right), \quad \beta \;=\; \min\left(1, \frac{-\tau}{\lambda_{\min}(Z^{-1}\delta Z)}\right), \tag{19}$$

  where $\tau < 1$ is a control parameter.

- (Corrector step)
  Compute the Newton step $(\Delta X, \Delta y, \Delta Z)$ from (15) and (17) with the right-hand side matrix R given by

  $$R \;=\; 2\mu I \;-\; P(XZ + \delta X \delta Z)P^{-1} \;-\; P^{-*}(ZX + \delta Z \delta X)P^{*}.$$

- Update (X, y, Z) to (X^+, y^+, Z^+) by

  $$X^+ \;=\; X \;+\; \alpha\,\Delta X, \quad y^+ \;=\; y \;+\; \beta\,\Delta y, \quad Z^+ \;=\; Z \;+\; \beta\,\Delta Z,$$

  where α and β are defined by (19) with δX, δZ replaced by ΔX, ΔZ.

We shall not discuss implementation details of the above algorithm — for example, how to solve efficiently for the search directions $(\delta X, \delta y, \delta Z)$ and $(\Delta X, \Delta y, \Delta Z)$ from the linear systems of $2N^2 + n$ equations (15) and (17). We refer the reader to [16] for such details. Instead, we just note that the search directions are typically computed via a Schur complement equation. For such an implementation, each iteration has a complexity of $O(nN^3) + O(n^2N^2)$, which is equal to $O(nN^3)$ for our Chebyshev approximation problem since $n < N$. Computations have shown that careful implementations of the predictor-corrector algorithm that use a Schur complement equation can typically reduce the duality gap of an SDP to about $\epsilon_{\mathrm{mach}}^{2/3}$ for the three search directions mentioned above, namely, the AHO, Monteiro and NT directions.

In all of our computations, we use the NT direction, for the following reasons. Although the orders of complexity for computing these three directions are the same, computing the AHO direction is about twice as expensive as computing the Monteiro or NT directions. Of the latter two, the NT direction has the virtue of being primal-dual symmetric. This implies that primal-dual predictor-corrector algorithms based on the NT direction are likely to be more robust than those based on the Monteiro direction, in the sense that the problems of stagnation such as taking very small steplengths are less likely to occur. It has been observed that algorithms based on the Monteiro direction often encounter such a stagnation problem for a class of SDPs known as the ETP problems [22].

5 The special case when *A* is normal

It is worth setting down the form our algorithm takes in the special case where A is normal, i.e., unitarily diagonalizable. As we have already mentioned in the Introduction, we may assume in this case that A is diagonal, so that the Chebyshev problem (1) reduces to the classical Chebyshev approximation problem on the spectrum $\Lambda(A)$ of A, i.e.,

$$\|p_n^*(A)\| \;=\; \|p_n^*\|_{\Lambda(A)} \;=\; \text{minimum}. \tag{20}$$

For this special case, the Chebyshev polynomials of A can be computed cheaply by the predictor-corrector algorithm discussed in the last section, by exploiting the block diagonal structure present in the associated SDP problem.

As in the general case, we consider the norm minimization problem (2), but the matrices B_k, $k = 0, \ldots, n$ are now diagonal: $B_k = \mathrm{diag}(d_k)$ for each k. Since the 2-norm of a diagonal matrix is the $\|\cdot\|_\infty$-norm of its diagonal vector, (2) is equivalent to the minimax problem

$$\min_{x\in\mathbb{C}^n}\ \max_{1\le l\le N}\left|\sum_{k=1}^{n} x_k d_k^{(l)} - d_0^{(l)}\right|, \tag{21}$$

where $d_k^{(l)}$ denotes the lth component of the N-vector d_k. As before, (21) can be expressed as an SDP.

Proposition 5.1 *The minimax problem* (21) *is equivalent to the following SDP involving block diagonal hermitian matrices:*

$$-\max_{x\in\mathbb{C}^n,\lambda\in R}\ \lambda$$

$$\text{s.t.}\quad \sum_{k=1}^{n}\left(\alpha_k A_k + \beta_k A_{n+k}\right) + \lambda A_{2n+1} + Z = A_0, \qquad Z \ge 0, \tag{22}$$

where $\alpha_k = \mathrm{Re}(x_k)$, $\beta_k = \mathrm{Im}(x_k)$,

$$A_{2n+1} = \mathrm{diag}\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{l=1}^{N}\right), \quad A_0 = \mathrm{diag}\left(\begin{bmatrix} 0 & d_0^{(l)} \\ \overline{d_0^{(l)}} & 0 \end{bmatrix}_{l=1}^{N}\right),$$

$$A_k = \mathrm{diag}\left(\begin{bmatrix} 0 & d_k^{(l)} \\ \overline{d_k^{(l)}} & 0 \end{bmatrix}_{l=1}^{N}\right), \quad A_{n+k} = \mathrm{diag}\left(\begin{bmatrix} 0 & i\, d_k^{(l)} \\ -i\, \overline{d_k^{(l)}} & 0 \end{bmatrix}_{l=1}^{N}\right), \tag{23}$$

$k = 1, \ldots, n$. *The matrices* A_k *consist of* N *blocks of* 2×2 *matrices on the diagonal.*

A proof of the above proposition is similar to that of Proposition 2.1, based on the observation that for any complex number a, we have

$$|a| = \lambda_{\max}\left(\begin{bmatrix} 0 & a \\ \bar{a} & 0 \end{bmatrix}\right).$$

We omit the details. Also, the process of transformation to a better-conditioned basis for (22) is exactly the same as for the general case. However, note that (22) cannot be obtained as a direct consequence of Proposition 2.1 by specializing the matrices B_k to diagonal matrices.

If the initial iterate (X^0, Z^0) is chosen to have the same block diagonal structure as the matrices A_k, then this structure is preserved throughout for (X^k, Z^k).

By exploiting this block diagonal structure, the work for each iteration of the predictor-corrector algorithm is reduced to $O(n^2N)$ flops as opposed to $O(nN^3)$ for non-normal matrices. In practice, we can compute the degree-25 Chebyshev polynomial of a normal matrix of dimension 1000 in MATLAB in about 12 minutes on a Sun Ultra Sparcstation.

6 Computed examples

We turn now to computed examples of Chebyshev polynomials of matrices. Our aim is to give some insight into the behavior of these polynomials. This is not a subject we fully understand, but the experimental observations are fascinating.

Most of our experimental results will be presented as plots. To "plot" a polynomial p_n^*, we show its roots in the complex plane and also the boundary of a region that we call the *Chebyshev lemniscate*[1] for that polynomial and the given matrix A. This region is defined by the equation

$$\mathcal{L}_n^*(A) \;=\; \{z : |p_n^*(z)| \;\leq\; \|p_n^*(A)\|\}. \tag{24}$$

The Chebyshev lemniscates characterize where in the complex plane the Chebyshev polynomials of the matrix "live", just as the spectrum or the pseudospectra characterize (though not precisely, unless A is normal) where in the complex plane the matrix A "lives". As a minimum, since $\|p_n^*\|_{\Lambda(A)} \leq \|p_n^*(A)\|$, we know that the Chebyshev lemniscate contains the spectrum

$$\Lambda(A) \;\subset\; \mathcal{L}_n^*(A). \tag{25}$$

In each example we present, the dimension of the matrix A is 48×48 or 100×100, though we typically print only its 5×5 or 6×6 analogue. For each example, we give plots showing the Chebyshev lemniscates (solid curves) of A, typically of degrees $n = 8$ and $n = 16$. The zeros of the Chebyshev polynomials are shown as small circles, and the eigenvalues of A are shown as solid dots.

For comparison with the Chebyshev lemniscate, each of our plots also shows a dotted curve. This is the boundary of an ϵ-pseudospectrum of A. The value of ϵ has been chosen "by hand" to make the match with the Chebyshev lemniscate a good one. (The ϵ-pseudospectrum of A is the set $\Lambda_\epsilon(A) = \{z : \|(zI - A)^{-1}\| \leq \epsilon^{-1}\}$ in the complex plane; see [14] and [21].)

---

[1] To be precise, the word lemniscate refers to the boundary of $\mathcal{L}_n^*$, and the region $\mathcal{L}_n^*$ should be called a *lemniscatic region*, but this expression is cumbersome and we shall avoid it.

For all of these examples, the Chebyshev polynomials were computed in MATLAB by the methods described in the previous sections.

Primal-dual predictor-corrector algorithms are highly efficient and robust for solving SDPs. For the set of examples we present here, it takes an average of 12 iterations to reduce the duality gap by a factor of 10^{10}. (This number is rather insensitive to the dimension of A; it would be essentially the same for matrices of dimensions 5×5 or 200×200. This insensitivity to problem size is one of the remarkable features of primal-dual interior point methods.) For a 48×48 real matrix, each iteration takes about 5 and 7 seconds for $n = 8$ and $n = 16$, respectively, on a Sun Ultra Sparcstation. The corresponding numbers for a 48×48 complex matrix are about 30 seconds and 45 seconds.

Here are our examples.

Example 1. Diagonal.

$$A \;=\; \mathrm{diag}(d), \qquad (100 \times 100)$$

where d is a vector whose first entry is 1 and the rest of whose entries are distributed uniformly in the interval $[-1, 0.8]$. Thus the spectrum of A consists of points that densely fill the interval $[-1, 0.8]$ and an outlier at $z = 1$.

Example 2. Bidiagonal.

$$A \;=\; \begin{pmatrix} d_1 & 0.2 & & & \\ & d_2 & 0.2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 0.2 \\ & & & & d_N \end{pmatrix}, \qquad (100 \times 100)$$

where the vector d is the same as that in Example 1. The spectrum is the same as that of Example 1.

Example 3. Grcar [20].

$$A \;=\; \begin{pmatrix} 1 & 1 & 1 & 1 & & \\ -1 & 1 & 1 & 1 & 1 & \\ & -1 & 1 & 1 & 1 & 1 \\ & & -1 & 1 & 1 & 1 \\ & & & -1 & 1 & 1 \\ & & & & -1 & 1 \end{pmatrix}. \qquad (48 \times 48)$$

Example 4. Ellipse.

$$A \;=\; \begin{pmatrix} 0 & 3 & & & \\ 2 & 0 & 3 & & \\ & 2 & 0 & 3 & \\ & & 2 & 0 & 3 \\ & & & 2 & 0 \end{pmatrix}. \qquad (48 \times 48)$$

Example 5. Bull's head [14].

$$A \;=\; \begin{pmatrix} 0 & 0 & 1 & .7 & & \\ 2i & 0 & 0 & 1 & .7 & \\ & 2i & 0 & 0 & 1 & .7 \\ & & 2i & 0 & 0 & 1 \\ & & & 2i & 0 & 0 \\ & & & & 2i & 0 \end{pmatrix}. \qquad (48 \times 48)$$

Example 6. Lemniscate1 [14].

$$A \;=\; \begin{pmatrix} 1 & 1 & & & & \\ & -1 & 1 & & & \\ & & 1 & 1 & & \\ & & & -1 & 1 & \\ & & & & 1 & 1 \\ & & & & & -1 \end{pmatrix}. \qquad (48 \times 48)$$

Example 7. Lemniscate2 [20].

$$A \;=\; \begin{pmatrix} 1 & \alpha & & & & \\ & 5 & \alpha & & & \\ & & 5 & \alpha & & \\ & & & 1 & \alpha & \\ & & & & 5 & \alpha \\ & & & & & 5 \end{pmatrix}, \qquad (48 \times 48)$$

where $\alpha = (256/27)^{1/3}$.

Example 8. Gauss-Seidel [14].

$$A = \begin{pmatrix} 0 & \frac{1}{2} & & & & \\ 0 & \frac{1}{4} & \frac{1}{2} & & & \\ 0 & \frac{1}{8} & \frac{1}{4} & \frac{1}{2} & & \\ 0 & \frac{1}{16} & \frac{1}{8} & \frac{1}{4} & \frac{1}{2} & \\ 0 & \frac{1}{32} & \frac{1}{16} & \frac{1}{8} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{64} & \frac{1}{32} & \frac{1}{16} & \frac{1}{8} & \frac{1}{4} \end{pmatrix}. \qquad (48 \times 48)$$

This is the matrix corresponds to a Gauss-Seidel iteration applied to the standard 3–point discrete Laplacian on a grid of N points.

Example 9. Beam-Warming [20].

$$A = \begin{pmatrix} -1.5 & 2.0 & -1.5 & & \\ -\frac{1}{3} & -\frac{1}{2} & 1 & -\frac{1}{6} & \\ & -\frac{1}{3} & -\frac{1}{2} & 1 & -\frac{1}{6} \\ & & -\frac{1}{3} & -\frac{1}{2} & 1 \\ & & 0.7 & -2.6 & 2.1 \end{pmatrix}. \qquad (48 \times 48)$$

Example 10. Wilkinson [20].

$$A = \begin{pmatrix} \frac{1}{N} & 1 & & & \\ & \frac{2}{N} & 1 & & \\ & & \ddots & \ddots & \\ & & & \frac{N-1}{N} & 1 \\ & & & & 1 \end{pmatrix}. \qquad (48 \times 48)$$

Example 11. Chebyshev points.

$$A = \begin{pmatrix} x_1 & \gamma_1 & & & \\ & x_2 & \gamma_2 & & \\ & & \ddots & \ddots & \\ & & & x_{N-1} & \gamma_{N-1} \\ & & & & x_N \end{pmatrix}, \qquad (48 \times 48)$$

where $\gamma_k = 0.5 - x_k$, $x_k = \cos\left(\frac{k-1}{N-1}\pi\right)$, $k = 1, \ldots, N$.

Example 12. Random [20].

$$A = \text{random}, \qquad (48 \times 48)$$

where by random we mean that the entries of A are independently drawn from the normal distribution with mean 0 and variance $1/N$.

Example 13. Random triangular [20].

$$A = \text{random triangular}, \quad (48 \times 48)$$

by which we mean that A is the strictly upper triangular part of the random matrix of Example 11.

Example 14. Convection-diffusion matrix [18].

> The matrix A is the projection of the $2N \times 2N$ Chebyshev spectral discretization matrix of a convection-diffusion operator onto the invariant subspace associated with the N eigenvalues of maximal real part ($N = 48$).

In Table 1, for later authors who may wish to compare the coefficients of some Chebyshev polynomials of matrices, we list the coefficients of p_8^* for the matrices of Examples 3 and 5. In Table 2, we give the computed values of $\|p_8^*(A)\|$.

Grcar computed p_8^*	Bull's head computed p_8^*
1	1
-7.90306320	$-0.00279600 - 0.00031731\,i$
41.3354079	$0.03375518 + 0.00745584\,i$
-150.565236	$-0.22640678 - 0.07463880\,i$
419.059092	$0.90451365 + 0.40702758\,i$
-897.405790	$-2.12812512 - 1.29015885\,i$
1464.45030	$2.67821078 + 2.33940114\,i$
-1722.68403	$-1.34847513 - 2.19132640\,i$
1271.98751	$0.05968294 + 0.74912144\,i$

Table 1: *Computed coefficients of p_8^* for the Grcar and Bull's head matrices (Examples 3 and 5). All but perhaps the last two digits printed are believed to be correct.*

The plots for our 14 examples are shown in Figures 1–14.

Example	computed $\|p_8^*(A)\|$
1. Diagonal	0.0063675408
2. Bidiagonal	0.0551494047
3. Grcar	1766.3135313
4. Ellipse	7710.2711611
5. Bull's head	1239.4186173
6. Lemniscate1	1.0000000000
7. Lemniscate2	834.73857463
8. Gauss-Seidel	0.0049251285
9. Beam-Warming	7.4348443860
10. Wilkinson	6.2747795054
11. Chebyshev points	46.395131600
12. Random	2.9537221027
13. Random triangular	0.0039633789
14. Convection-diffusion	2623904.6097

Table 2: *Norms* $\|p_8^*(A)\|$ *for Examples 1 – 14. All digits printed are believed to be correct, as the estimated relative accuracies are all less than* 10^{-11}.

Let us first consider Example 1, the special case where A is diagonal. For any Chebyshev polynomial of a matrix, we know that the Chebyshev lemniscate must contain the spectrum (25). In the present case, by the characterization theorems for the classical complex Chebyshev approximation problem [3, p. 143], we know that the nth Chebyshev lemniscate must in fact touch the spectrum $\Lambda(A)$ at no fewer than $n+1$ points. This property is evident in Figure 1, where we see that $\mathcal{L}_n^*(A)$ hugs $\Lambda(A)$ rather closely, and increasingly so as n increases. It is interesting also to note how quickly one of the roots of the polynomials p_n^*, which are analogous to the "Ritz values" often taken as eigenvalue estimates in Arnoldi or Lanczos iterations, converges to the outlier eigenvalue at $z = 1$. By $n = 6$, one of the roots of p_6^* is already very close to the outlier, and the distance between them decreases geometrically as n increases. In the remainder of the spectrum, on the other hand, no individual Ritz value is converging rapidly to any one eigenvalue of A. Rather, it is the Chebyshev lemniscate generated by these Ritz values jointly that is capturing the spectrum.

One might expect $\mathcal{L}_n^*(A)$ to more or less approximate $\Lambda(A)$ even if A is non-

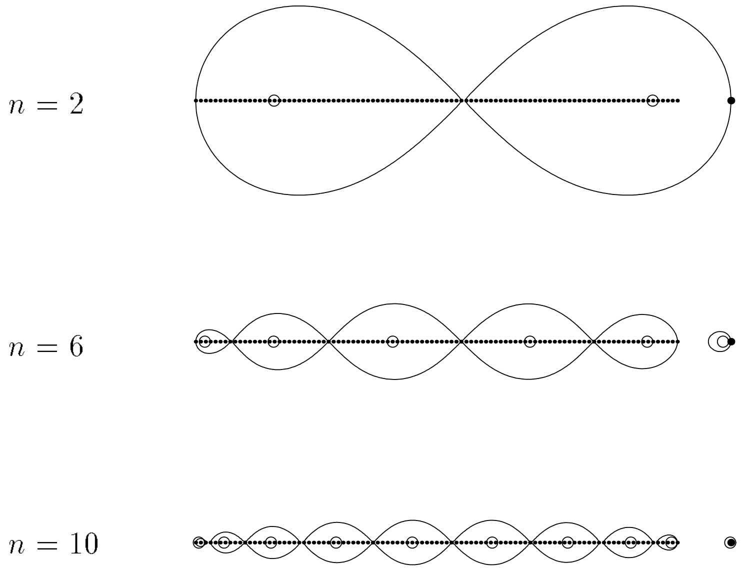

Figure 1: *Diagonal. Since* A *is normal, the Chebyshev lemniscate touches the spectrum at at least* $n + 1$ *points, and the roots of* p_n^* *lie in the convex hull of the spectrum.*

normal. But from Figures 2–14, the reader will see that this is not so. Nonetheless, though $\mathcal{L}_n^*(A)$ does not always approximate $\Lambda(A)$ very closely, it may still gain some information about A. The plots show that for these examples, to a sometimes startling degree,

$$\mathcal{L}_n^*(A) \;\approx\; \Lambda_\epsilon(A) \tag{26}$$

for some $\epsilon \geq 0$. In particular, the agreement of the Chebyshev lemniscate of p_n^* with a pseudospectrum of A is in most of these examples far closer than the agreement of the roots of p_n^* with the eigenvalues of A. For example, consider Figure 2, the bidiagonal matrix that is the non-normal analogue of Example 1 with the same spectrum. Except for the outlier eigenvalue, the roots of p_n^* bear no resemblance to individual eigenvalues of A. By contrast the Chebyshev lemniscates of these polynomials show a striking resemblance to the $\epsilon = 10^{-1.5}$ pseudospectrum of A. Clearly the Chebyshev polynomial is approximating A in a fashion that goes far beyond approximation of individual eigenvalues.

The other examples illustrate the same effect. In every case, the lemniscate of the Chebyshev polynomial shows a compelling approximation to the pseudospectrum. We do not claim that this effect is universal; these examples have been picked for their pronounced and cleanly structured non-normality. But it is certainly common.

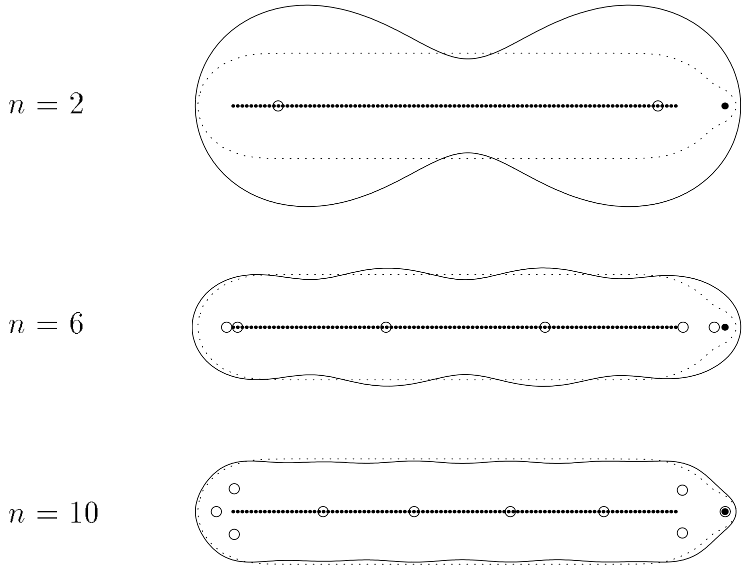

Figure 2: *Bidiagonal. A non-normal analogue of Example 1. The dotted curves are the* $\epsilon = 10^{-1.5}$ *pseudospectrum of* A.

A partial explanation of this phenomenon is as follows. It is well known that a matrix polynomial $p(A)$ can be expressed as a Cauchy integral

$$p(A) \;=\; \frac{1}{2\pi i}\int p(z)\,(zI - A)^{-1}dz, \tag{27}$$

where the integration is over any closed contour or union of contours enclosing the spectrum of A once in the counterclockwise direction [8]. Taking absolute values gives the inequality

$$\|p(A)\| \;\le\; \frac{1}{2\pi}\int |p(z)|\,\|(zI - A)^{-1}\|\,|dz|. \tag{28}$$

Now suppose we seek p such that $\|p(A)\|$ is small. When the degree of p is smaller than the dimension of A, it is impossible to achieve this in general by putting zeros of p wherever A has eigenvalues, which would make the integral zero. Instead, we must settle for making $|p(z)|$ small where $\|(zI - A)^{-1}\|$ is large. This immediately suggests a link between lemniscates of p and pseudospectra of A.

From this kind of reasoning we can derive bounds on $\|p_n^*(A)\|$. For example, to minimize $\|p(A)\|$ one might seek to minimize $\|p\|_{\Lambda_\epsilon(A)}$ for some ϵ that is not too small (hence $|p(z)|$ is small over the region where $\|(zI - A)^{-1}\|$ is larger than

ϵ^{-1}). From (28) and the minimality of $\|p_n^*(A)\|$ we conclude that

$$\|p_n^*(A)\| \;\leq\; \frac{L_\epsilon}{2\pi\epsilon}\,\min_p \|p\|_{\Lambda_\epsilon(A)}, \tag{29}$$

where L_ϵ is the arclength of the boundary of $\Lambda_\epsilon(A)$. At this point one runs into the fact that $\min_p \|p\|_{\Lambda_\epsilon(A)}$ can be huge if ϵ is not small, since the minimum typically increases geometrically with ϵ. Therefore, a compromise must be made on ϵ so that the quantity $\min_p \|p\|_{\Lambda_\epsilon(A)}/\epsilon$ on the right-hand side of (29) is as small as possible.

For some matrices A and choices of n and ϵ, the estimate just described can be quite good. It is not always very good, however, and so far, our attempts to make a more precise link between lemniscates of Chebyshev polynomials and pseudospectra of the underlying matrix have been unsuccessful except in certain limiting cases $n \to \infty$ described in [17]. Rather than present partial results that do not bring this matter to a very satisfactory state, we prefer to leave the explanation of the behavior of Figures 2–14 as an open problem.

7 Conclusions

This paper has made two contributions. The first is a fast and robust algorithm for computing the Chebyshev polynomials of a matrix, based on primal-dual interior point methods in semidefinite programming. The second is an experimental study of these polynomials that indicates that the associated lemniscates may closely approximate certain pseudospectra if the matrix A is far from normal.

We have said little about applications in iterative numerical linear algebra, though that is the field that motivates all of these developments. There are many possibilities here that might be explored now that an algorithm is available.

It was mentioned in the Introduction that for applications to iterative solution of equations rather than eigenvalue calculations, it is appropriate to minimize $\|p(A)\|$ with the normalization $p(0) = 1$ instead of $\lim_{z\to\infty} p(z)/z^n = 1$. Plots of lemniscates for these "ideal GMRES polynomials" can be found in the first author's dissertation [17]. Because this normalization gives a special status to the origin, these problems are no longer translation-invariant in the complex plane, and the lemniscates take special pains to avoid the origin. They also tend to display scallop patterns near the spectrum or pseudospectra.

Interesting connections can also be made to the notion of a *generalized Kreiss matrix theorem*. The usual Kreiss matrix theorem relates the norms $\|A^n\|$ to the behavior of the pseudospectra of A near the unit disk; generalizations are obtained by looking norms $\|p(A)\|$ of other polynomials and the behavior of the pseudospectra near other regions. These matters are investigated in [19].

We consider the idea of the Chebyshev polynomials of a matrix a natural one, suggesting many questions waiting to be explored. We hope that more will be learned about the behavior of these polynomials in the years ahead and that applications in other areas may be found.

References

[1] F. Alizadeh, J.-P.A. Haeberly and M. Overton, *Primal-dual interior-point methods for semidefinite programming*, manuscript, August, 1994.

[2] J.H. Curtiss, *Faber polynomials and the Faber series*, Amer. Math. Monthly, 78 (1971), pp. 577–596.

[3] P.J. Davis, *Interpolation and Approximation*, Dover Publications, Inc., New York, 1975.

[4] T.A. Driscoll, *A MATLAB Toolbox for Schwarz-Christoffel mapping*, ACM Transactions on Mathematical Software, to appear.

[5] G. Faber, *Über Tschebyscheffsche Polynome*, J. Reine Angew. Math. 150 (1920), 79–106.

[6] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 2nd ed., Johns Hopkins University Press, Baltimore, 1989.

[7] A. Greenbaum and L.N. Trefethen, *GMRES/CR and Arnoldi/Lanczos as matrix approximation problems*, SIAM J. Sci. Comp. 15 (1994), pp. 359–368.

[8] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin, 1966.

[9] C.-J. Lin and R. Saigal, *An infeasible start predictor-corrector method for semi-definite linear programming*, manuscript, Department of Industrial and Operations Engineering, University of Michigan, December, 1995.

[10] R.D.C. Monteiro, *Primal-dual path following algorithms for semidefinite programming*, manuscript, 1995.

[11] Y. Nesterov and A. Nemirovsky, *Interior-Point Polynomial Methods in Convex Programming*, Vol. 13 of Studies in Applied Mathematics, SIAM, Philadelphia, PA, 1994.

[12] Y. Nesterov and M.J. Todd, *Primal-dual interior-point methods for self-scaled cones*, Technical Report 1125, School of Operations Research and Industrial Engineering, Cornell University, August 1995.

[13] S.C. Reddy and L.N. Trefethen, *Pseudospectra of the convection-diffusion operator*, SIAM J. Appl. Math., 54 (1994), pp. 1634–1649.

[14] L. Reichel and L.N. Trefethen, *Eigenvalues and pseudo-eigenvalues of Toeplitz matrices*, Linear Algebra Appl., 169 (1992), pp. 153–185.

[15] P.T.P. Tang, *A fast algorithm for linear complex Chebyshev approximations*, Math. Comput., 51 (1988), pp. 721–739.

[16] M.J. Todd, K.C. Toh, and R.H. Tütüncü, *On the Nesterov-Todd direction in semidefinite programming*, submitted to SIAM J. of Optimization, 1996.

[17] K.C. Toh, *Matrix Approximation Problems and Nonsymmetric Iterative Methods*, PhD Thesis, Cornell University, May, 1996.

[18] K.C. Toh and L.N. Trefethen, *Calculation of pseudospectra by the Arnoldi iteration*, SIAM J. Sci. Comput., 17 (1996), pp. 1–15.

[19] K.C. Toh and L.N. Trefethen, *The Kreiss matrix theorem on a general complex domain*, manuscript in preparation.

[20] L.N. Trefethen, *Spectra and Pseudospectra: The Behavior of Non-Normal Matrices and Operators*, book in preparation.

[21] L.N. Trefethen, *Pseudospectra of matrices*, in D.F. Griffiths and G.A. Watson, eds., Numerical Analysis 1991, Longman Scientific and Technical, Harlow, UK, 1992.

[22] L. Vandenberghe and S. Boyd, *Semidefinite programming*, submitted to SIAM Review.

[23] L. Vandenberghe and S. Boyd, *User's guide to SP: software for semidefinite programming*, Information Systems Laboratory, Stanford University, November 1994. Available via anonymous ftp at `isl.stanford.edu` in `pub/boyd/semidef_prog`. Beta version.

[24] H. Widom, *Extremal polynomials associated with a system of curves in the complex plane*, Adv. Math., 3 (1969), pp. 127–232.

[25] Y. Zhang, *On extending primal-dual interior-point algorithms from linear programming to semidefinite programming*, TR95-20 (1995), Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore.

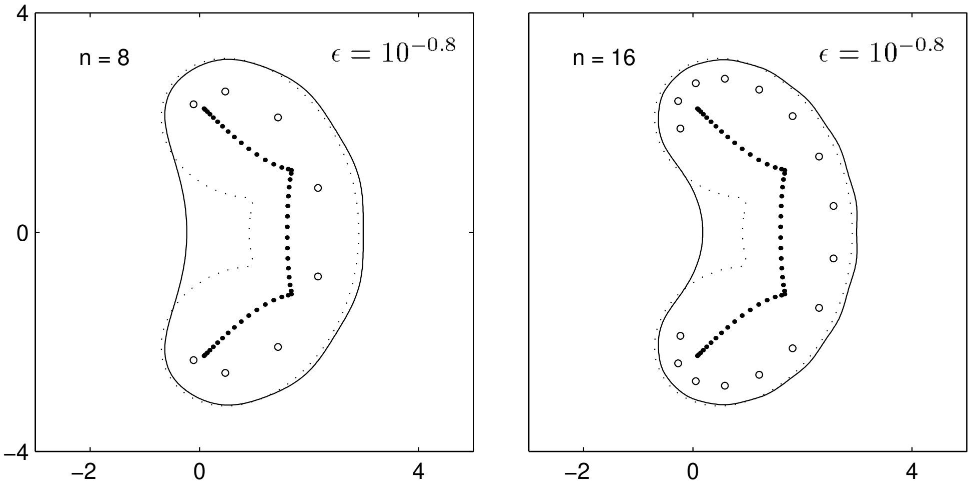

Figure 3: *Grcar — a non-normal Toeplitz matrix.*

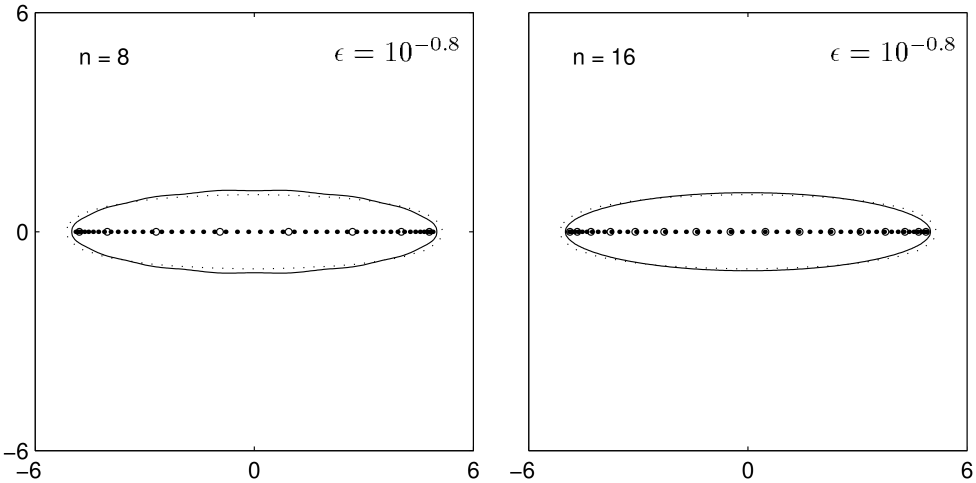

Figure 4: *Ellipse. The eigenvalues are all real, but the pseudospectra bulge into the complex plane.*

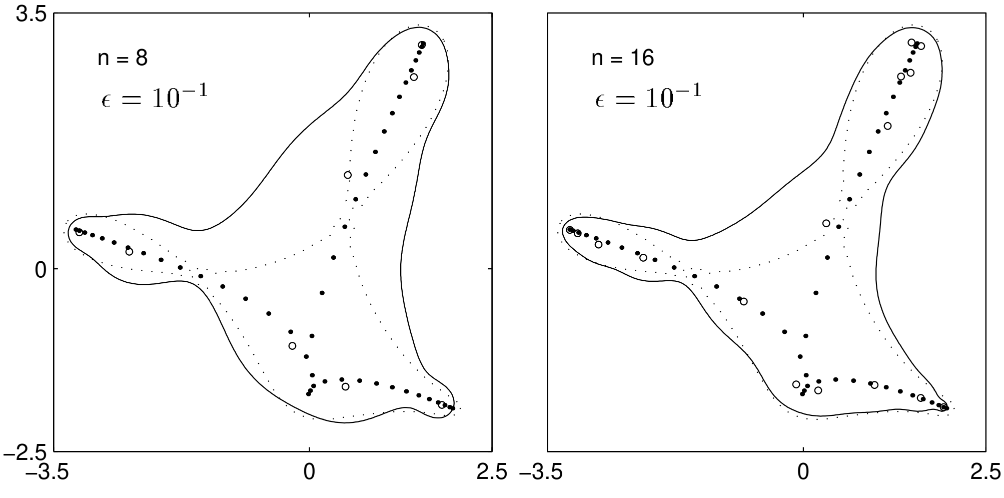

Figure 5: *Bull's head — our only example with complex entries, hence a figure with no symmetries.*

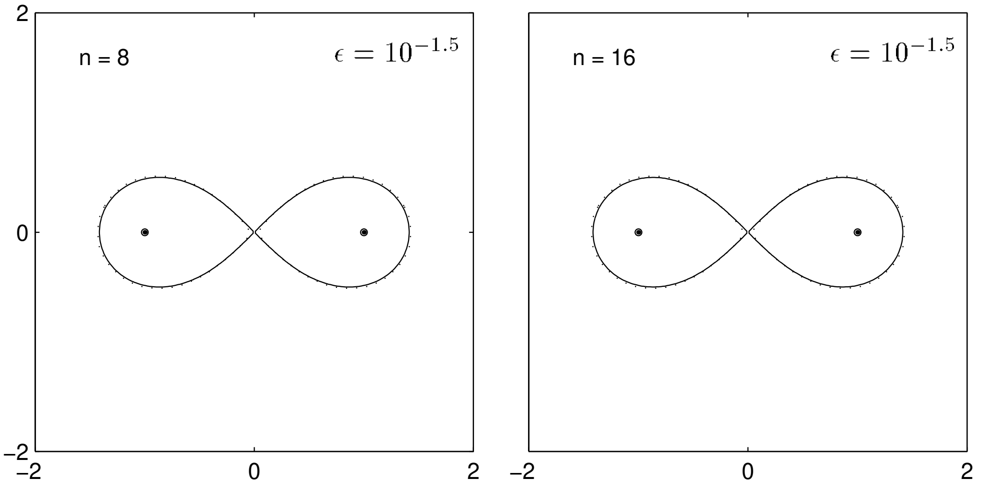

Figure 6: *Lemniscate1. The zeros of* p_n^* *are* $\{-1, 1\}$*. The dotted and solid curves coincide almost exactly.*

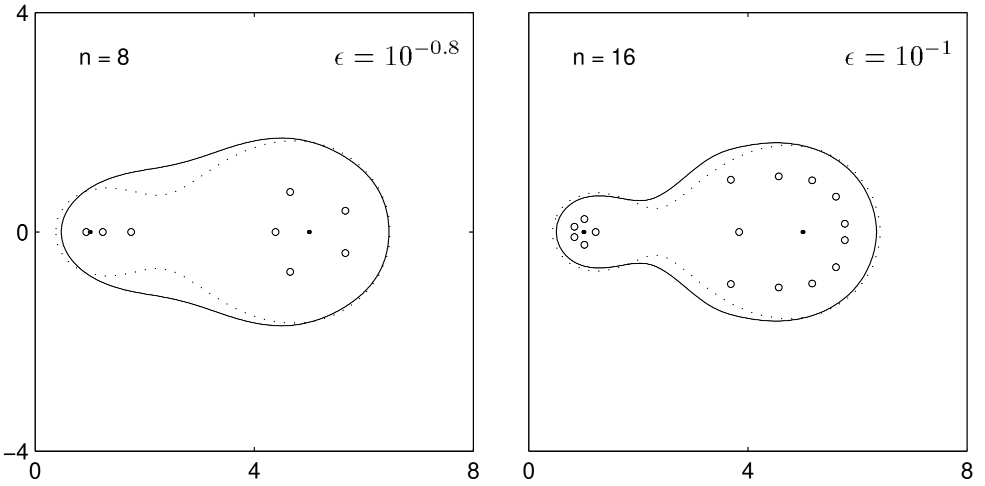

Figure 7: *Lemniscate2. Here one eigenvalue of the matrix has twice the multiplicity of the other.*

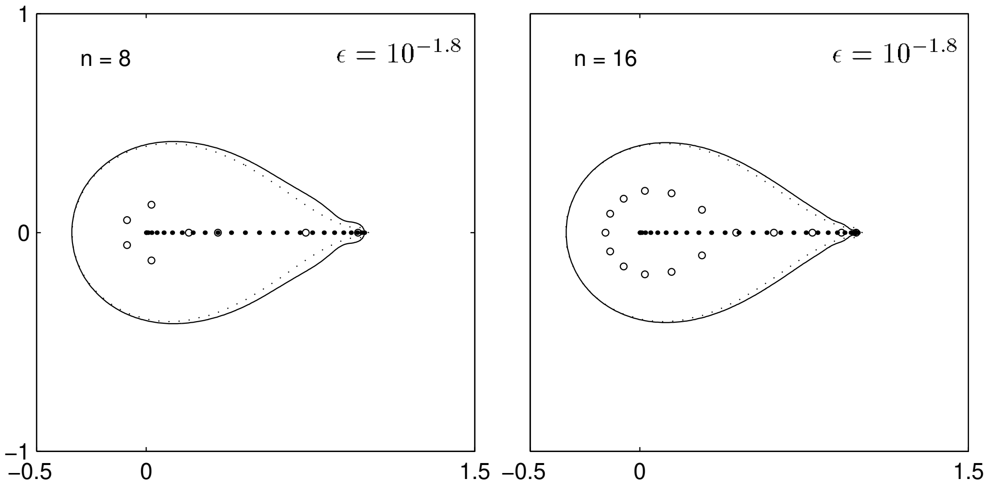

Figure 8: *Gauss-Seidel. Half the eigenvalues are at the origin.*

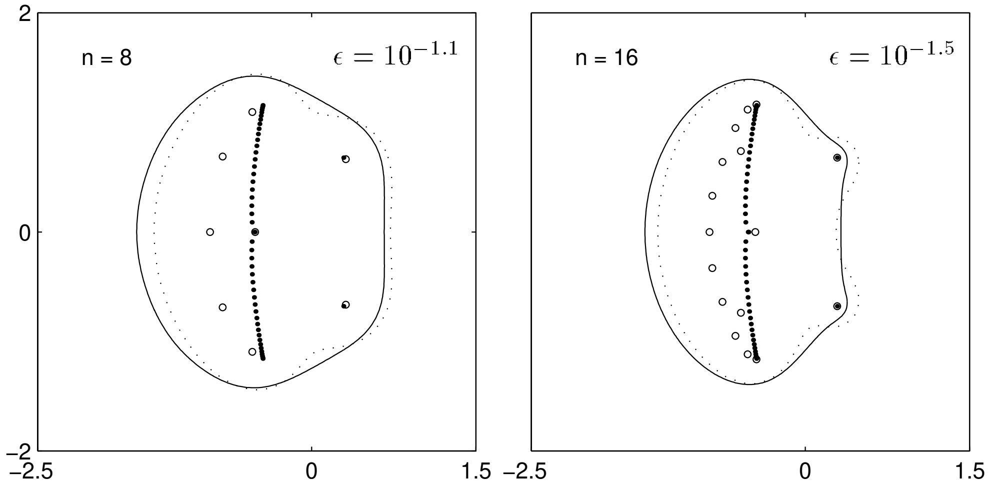

Figure 9: *Beam-Warming — a "quasi–Toeplitz" matrix.*

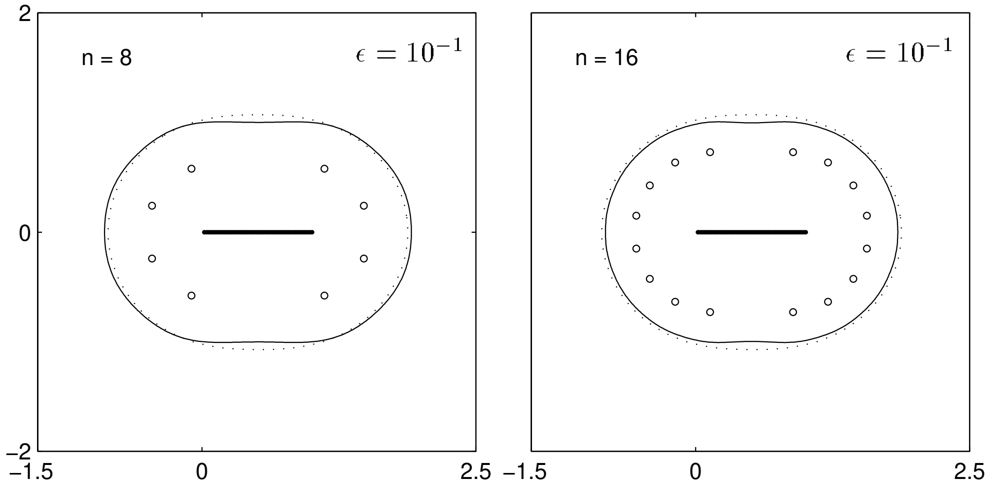

Figure 10: *Wilkinson — evenly spaced eigenvalues on the real axis.*

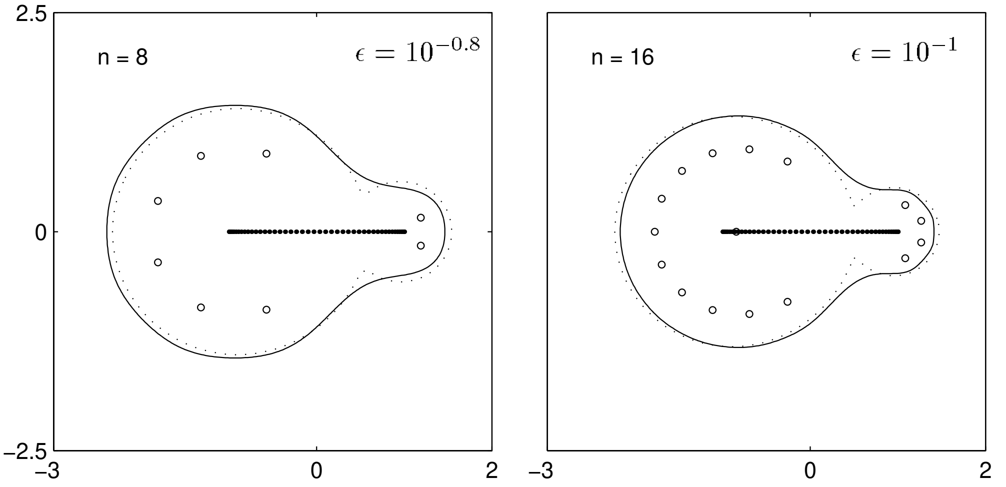

Figure 11: *Chebyshev points — unevenly spaced eigenvalues.*

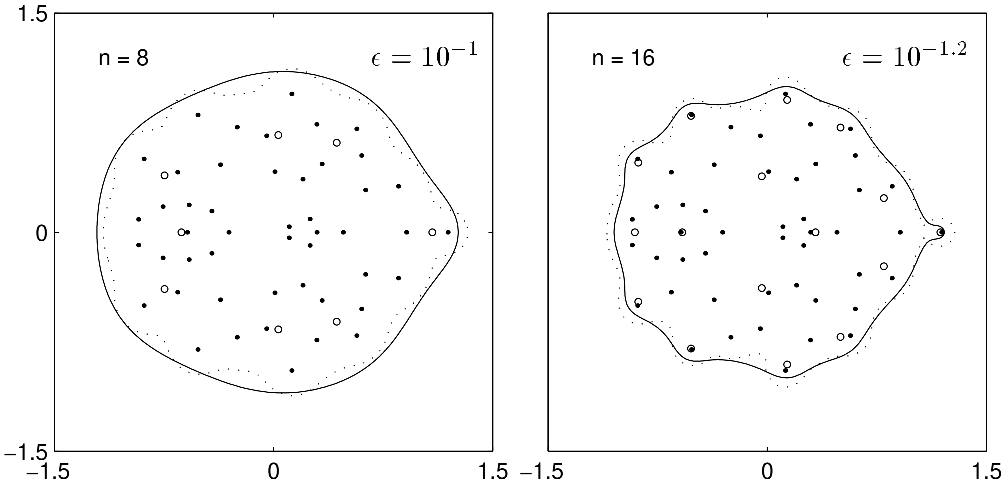

Figure 12: *Random. This matrix is only mildly non-normal.*

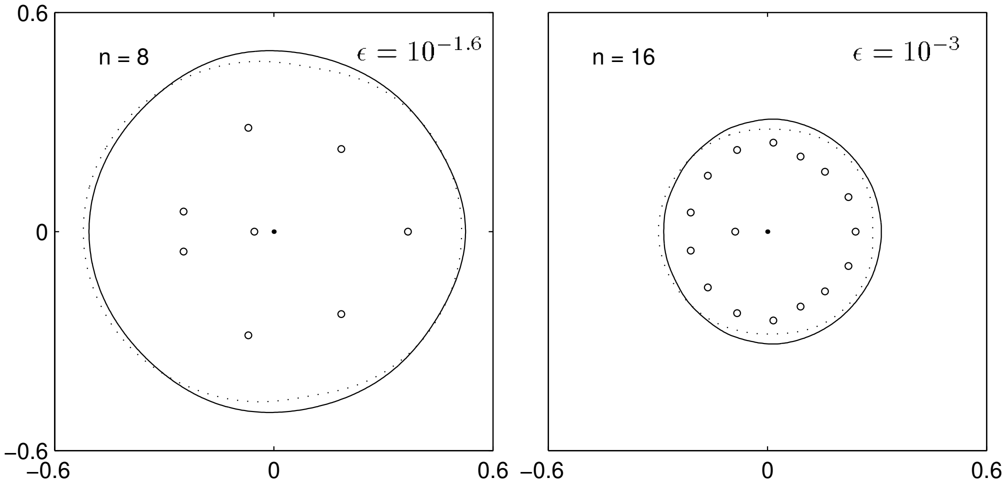

Figure 13: *Random triangular. The non-normality is now pronounced.*

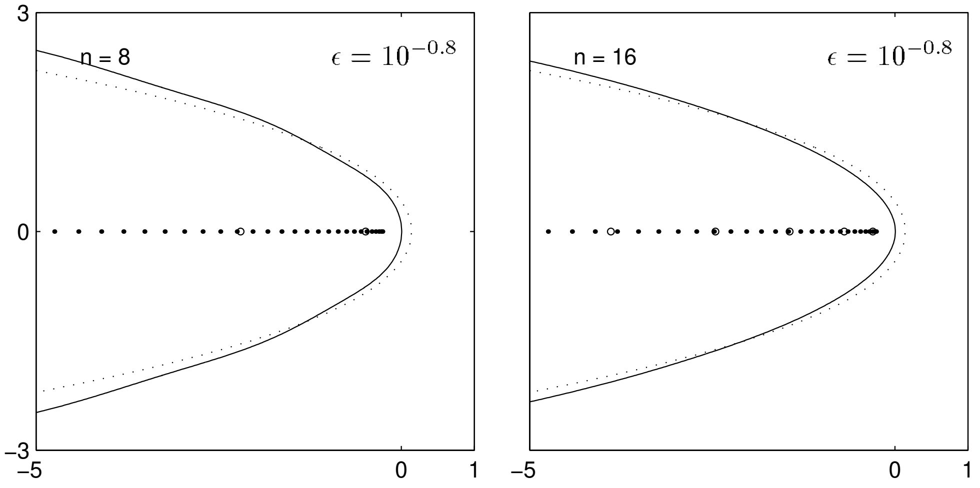

Figure 14: *Convection-diffusion. A matrix approximation to the canonical non-normal differential operator.*