Electronic Transactions on Numerical Analysis.
Volume 18, pp. 1-41, 2004.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

TRANSIENT BEHAVIOR OF POWERS AND EXPONENTIALS OF LARGE TOEPLITZ MATRICES*

ALBRECHT BÖTTCHER†

Abstract. The message of this paper is that powers of large Toeplitz matrices show critical behavior if and only if the L^∞ norm of the symbol is greater than one. Critical behavior means that the norms of the powers grow exponentially to infinity or that they run through a critical transient phase before decaying exponentially to zero. We summarize several known results that are relevant to the problem. Moreover, the paper contains some new results and illuminates the question from a perspective that might be new. The paper may serve as an introduction to a few basic phenomena one encounters in the field.

Key words. Toeplitz matrix, matrix power, matrix exponential, transient behavior.

AMS subject classifications. 47B35, 15A60, 65F35.

1. Introduction. Let A_n be a complex $n \times n$ matrix. The behavior of the norms $\|A_n^k\|$ is of considerable interest in connection with several problems. We specify $\|\cdot\|$ to be the spectral norm (i.e., the operator norm associated with the ℓ^2 vector norm). The norms $\|A_n^k\|$ converge to zero as $k \to \infty$ if and only if $\varrho(A_n) < 1$, where $\varrho(A_n)$ denotes the spectral radius of A_n. However, sole knowledge of the spectral radius or even of all eigenvalues of A_n does not tell us whether the norms $\|A_n^k\|$ run through a critical transient phase, that is, whether there are k for which $\|A_n^k\|$ becomes very large, before eventually decaying exponentially to zero.

We here embark on the case where A_n is a Toeplitz matrix or where A_n can be approximated by a sum of products of Toeplitz matrices. An $n \times n$ Toeplitz matrix is a matrix of the form

$$(a_{j-k})_{j,k=1}^n = \begin{pmatrix} a_0 & a_{-1} & a_{-2} & \dots & a_{-(n-1)} \\ a_1 & a_0 & a_{-1} & \dots & a_{-(n-2)} \\ a_2 & a_1 & a_0 & \dots & a_{-(n-3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-1} & a_{n-2} & a_{n-3} & \dots & a_0 \end{pmatrix}, \tag{1.1}$$

and an infinite Toeplitz matrix is given by

$$(a_{j-k})_{j,k=1}^\infty = \begin{pmatrix} a_0 & a_{-1} & a_{-2} & \dots \\ a_1 & a_0 & a_{-1} & \dots \\ a_2 & a_1 & a_0 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}, \tag{1.2}$$

where $\{a_\ell\}_{\ell\in\mathbb{Z}}$ is a sequence of complex numbers. It is well-known that many properties of a Toeplitz matrix are encoded in its so-called *symbol*. This is an object whose Fourier coefficients are the numbers a_ℓ ($\ell \in \mathbb{Z}$). We here assume that the symbol is an L^∞ function. Thus, let a be a complex-valued function on the complex unit circle $\mathbf{T}$ and suppose $a \in L^\infty := L^\infty(\mathbf{T})$. The Fourier coefficients of a are defined by

$$a_\ell = \frac{1}{2\pi}\int_0^{2\pi} a(e^{i\theta})e^{-i\ell\theta}d\theta \quad (\ell \in \mathbb{Z}).$$

*Received July 23, 2003. Accepted for publication October 13, 2003. Recommended by Richard S. Varga.

†Fakultät für Mathematik, Technische Universität Chemnitz, D-09107 Chemnitz, Germany. E-mail: aboettch@mathematik.tu-chemnitz.de.

ETNA
Kent State University
etna@mcs.kent.edu

We denote the matrices (1.1) and (1.2) by $T_n(a)$ and $T(a)$, respectively. A 1911 result by Otto Toeplitz states that $T(a)$ induces a bounded operator on $\ell^2 := \ell^2(\mathbb{N})$ and that $\|T(a)\| = \|a\|_\infty$, where $\|\cdot\|_\infty$ is the norm in L^∞ (see, e.g., [10], [11]).

What can be said about the norms $\|T_n^k(a)\| := \|(T_n(a))^k\|$? The computer is expected to give a reliable answer in the case where n is small. We therefore assume that n is large. The message of this paper is that $\|T_n^k(a)\|$ has critical behavior as k increases if and only if $\|a\|_\infty > 1$. In other words, to find out whether $\|T_n^k(a)\|$ shows critical transient or limiting behavior, we need only look whether the L^∞ norm of the symbol is greater than one. Notice that it is much more difficult to decide whether $\varrho(T_n(a))$ or $\limsup_{n\to\infty} \varrho(T_n(a))$ is smaller than one.

We think of $\mathbb{C}^n$ as a subspace of ℓ^2 and denote by P_n the orthogonal projection of ℓ^2 onto $\mathbb{C}^n$. The range (i.e., the image) of P_n will be denoted by $\operatorname{Im} P_n$. On identifying an $n \times n$ matrix A_n with $A_n P_n$, we may regard A_n as an operator on ℓ^2.

First of all we remark that $T_n(a)$ is the compression of $T(a)$ to $\operatorname{Im} P_n = \mathbb{C}^n$, that is, $T_n(a) = P_n T(a) P_n | \operatorname{Im} P_n$. This implies at once that

$$\|T_n^k(a)\| \le \|T_n(a)\|^k \le \|T(a)\|^k = \|a\|_\infty^k \tag{1.3}$$

for all n and k. Thus, for the norms of powers of Toeplitz matrices we have the simple universal upper estimate (1.3). In particular, if $\|a\|_\infty \le 1$, then there is no critical behavior.

The following theorem shows that (1.3) reduces to equality as $n \to \infty$. This theorem may be viewed as an argument in support of the statement that $\|T_n^k(a)\|$ shows critical behavior whenever $\|a\|_\infty > 1$.

THEOREM 1.1. *If $a \in L^\infty$, then*

$$\lim_{n\to\infty} \|T_n^k(a)\| = \|a\|_\infty^k \tag{1.4}$$

for each natural number k.

Proof. For fixed k, the operators $T_n^k(a)$ converge strongly (i.e., pointwise) to $T^k(a)$ as $n \to \infty$. Hence, by the Banach-Steinhaus theorem,

$$\liminf_{n\to\infty} \|T_n^k(a)\| \ge \|T^k(a)\|. \tag{1.5}$$

By (1.3) and (1.5), we are left with proving that $\|T^k(a)\| \ge \|a\|_\infty^k$. But $\|T^k(a)\| \ge \varrho(T^k(a)) = (\varrho(T(a)))^k$, and the Hartman-Wintner theorem (see, e.g., [10]) says that the spectrum of $T(a)$ contains the essential range of a, whence $\varrho(T(a)) \ge \|a\|_\infty$. $\square$

The problem studied here amounts to looking for peaks of the "surface" $(k, n) \mapsto \|T_n^k(a)\|$ along the lines $n =$ constant, while Theorem 1.1 concerns the behavior of $\|T_n^k(a)\|$ along the lines $k =$ constant. We will return to this question later, especially in Section 7. For the moment, let us consider an example.

EXAMPLE 1.2. The simplest nontrivial Toeplitz matrix is the Jordan block

$$J_n(\lambda) = \begin{pmatrix} \lambda & 0 & 0 & \dots & 0 \\ 1 & \lambda & 0 & \dots & 0 \\ 0 & 1 & \lambda & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda \end{pmatrix}. \tag{1.6}$$

Clearly, $J_n(\lambda) = T_n(a)$ with $a(t) = \lambda + t$ ($t \in \mathbf{T}$). Figure 1.1 indicates the shape of the "surface" $(k, n) \mapsto \|J_n^k(0.8)\|$. We will omit the quotation marks in the following and will

ETNA
Kent State University
etna@mcs.kent.edu

simply speak of the norm surface. We will also silently identify the surface with its "map" in the k, n plane, that is, we will not distinguish the point $(k, n, f(k, n))$ on the surface from its projection (k, n) in the plane. The spectral radius of $J_n(0.8)$ equals 0.8, while the L^∞ norm of the symbol is $\|a\|_\infty = 1.8$. We see that the surface has a *lowland* (say below the curve $\|J_n^k(0.8)\| = 10^{-6}$, a *steep* (say between the curves $\|J_n^k(0.8)\| = 10^{-6}$ and $\|J_n^k(0.8)\| = 10^2$), and a part where it grows enormously (say above the curve $\|J_n^k(0.8)\| = 10^2$). We call the last part the *sky region*.

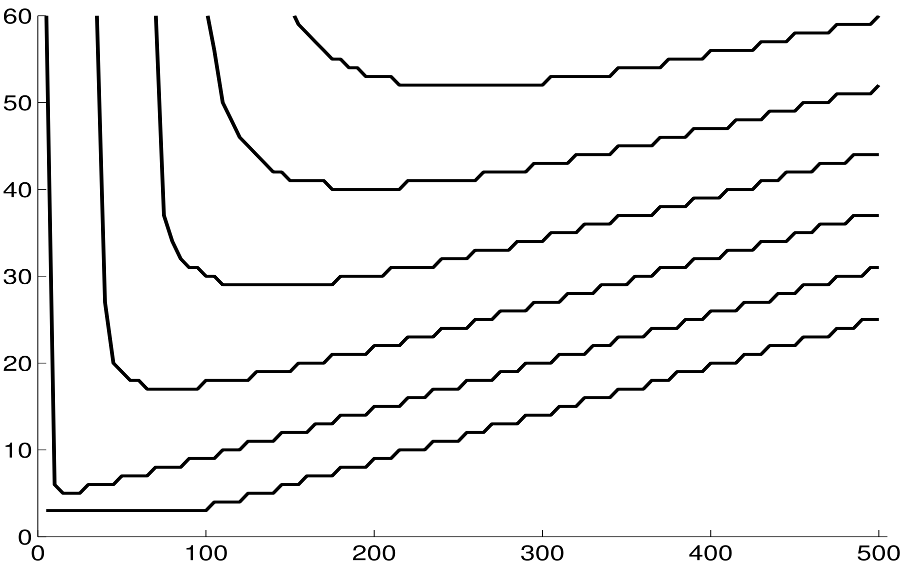

FIG. 1.1. *Level curves* $\|J_n^k(0.8)\| = c$ *for* $c = 10^{-6}$, 10^2, 10^{10}, 10^{18}, 10^{26}, 10^{34} (*the lower curve corresponds to* $c = 10^{-6}$, *the upper to* 10^{34}). *We took* $n = 3, 4, 5, \ldots, 60$ *and* $k = 5, 10, 15, \ldots, 500$.

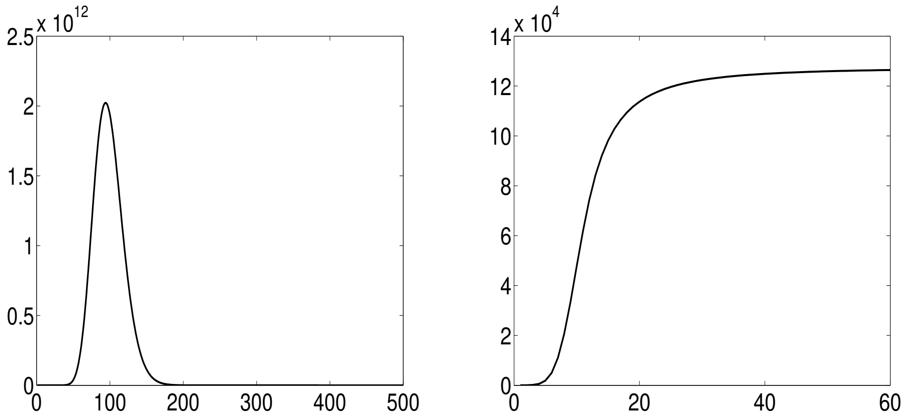

FIG. 1.2. *Movement on the surface* $(k, n) \mapsto \|J_n^k(0.8)\|$ *along the horizontal line* $n = 20$ (*left*) *and along the vertical line* $k = 20$ (*right*).

ETNA
Kent State University
etna@mcs.kent.edu

If n is fixed and k increases, we move horizontally in Figure 1.1. On the surface, we will soon be in the sky region, then step down the steep, and finally be caught in the lowland forever. Thus, we have the critical transient phase shown in the left picture of Figure 1.2. On the other hand, if we fix k and let n increase, then this corresponds to a vertical movement of Figure 1.1. This time we will fairly quickly reach the sky region and move at nearly constant height for the rest of the journey, as shown the right picture of Figure 1.2. Note that in the right picture of Figure 1.2 the norms converge to the limit $1.8^{20} = 12.75 \cdot 10^4$ and that the norms are already very close to this limit beginning with n between 20 and 40. □

The problem of determining the spectral radius of $T_n(a)$ is not trivial. If a is a rational function (without poles on $\mathbf{T}$, because otherwise a would not be in L^∞), then $\varrho(T_n(a))$ converges to a limit as $n \to \infty$, and formulas for this limit are known in a few important special cases; see [14], [15], [27] and Section 7. However, in practice one can compute this limit numerically: the spectral radii of $T_{100}(a)$ or even of $T_{30}(a)$ are often already very close to the limit. Things are more complicated for nonrational symbols. In that case it is known that, under certain assumptions, the inequality $\|a\|_\infty > 1$ implies that $\varrho(T_n(a)) > 1$ for all sufficiently large n. This is, for example, true in the following two situations.

(i) The function a is continuous but not C^∞ on $\mathbf{T}$ and there is a point $t_0 \in \mathbf{T}$ such that a is C^∞ on $\mathbf{T} \setminus \{t_0\}$; see [35], [36] and [11, pp. 168–169].

(ii) There is a point $t_0 \in \mathbf{T}$ such that a is C^2 on $\mathbf{T} \setminus \{t_0\}$, the limits $a(t_0 - 0)$ and $a(t_0 + 0)$ exist but do not coincide, and

$$\left| \arg\Big(a(t_0 - 0) - \lambda\Big) - \arg\Big(a(t_0 + 0) - \lambda\Big) \right| < 2\pi$$

for each λ outside the essential range of a, where $\arg$ denotes any continuous argument of $a(t) - \lambda$ ($t \in \mathbf{T} \setminus \{t_0\}$); see [11, p. 181].

Figure 1.3 shows two examples, one for case (i) and another one for case (ii).

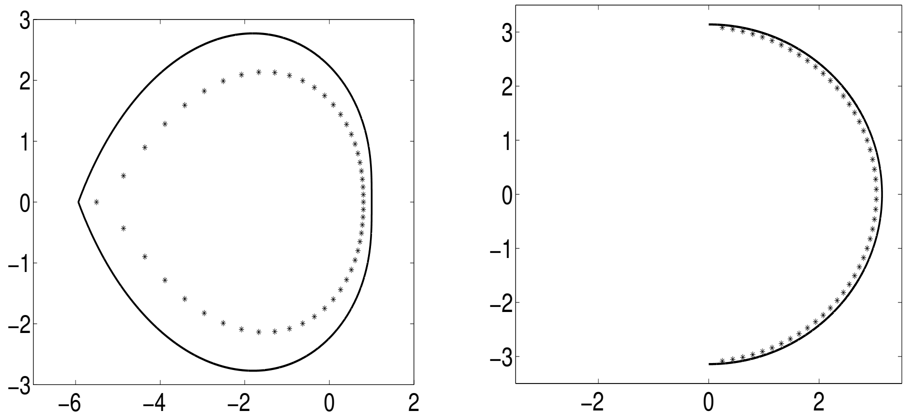

FIG. 1.3. *The symbols are $a(e^{i\theta}) = (1 + \frac{1}{2}\theta^2)e^{i\theta}$ and $a(e^{i\theta}) = i\pi e^{-i\theta/2}$ in the left and right pictures, respectively. We see the image of $\mathbf{T}$ under a and the eigenvalues of $T_{50}(a)$.*

ETNA
Kent State University
etna@mcs.kent.edu

In Figure 1.3, we observe that the eigenvalues cluster around the curve $a(\mathbf{T})$. Consequently, there will be eigenvalues near the point(s) of this curve having maximal distance from the origin. This implies that $\varrho(T_n(a))$ is close to $\|a\|_\infty$. Thus, if a satisfies (i) or (ii) and if $\|a\|_\infty > 1$, then $\|T_n^k(a)\| \to \infty$ as $k \to \infty$ whenever n is sufficiently large. The situation is different for rational symbols. In that case the eigenvalues cluster along certain curves that may stay away from $a(\mathbf{T})$ (see Figure 1.4). Hence, there may remain a gap between $\varrho(T_n(a))$ and $\|a\|_\infty$, which may result in a critical transient phase before the exponential decay to zero.

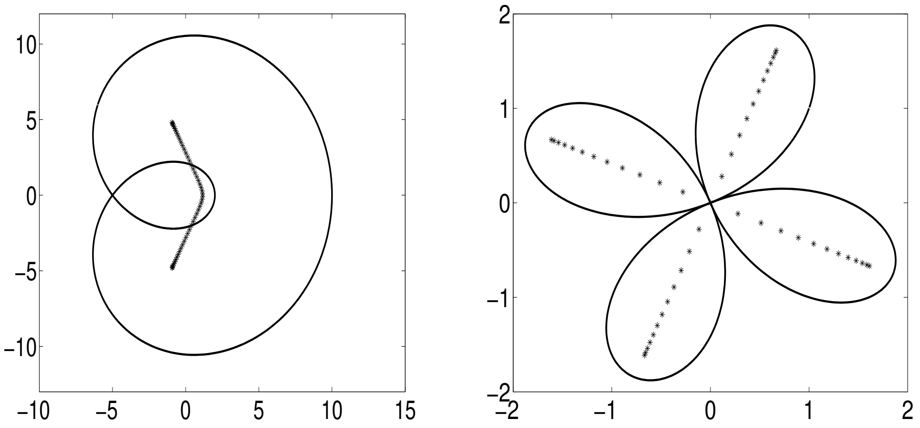

FIG. 1.4. *The symbols are $a(t) = it + 5t^{-1} - 6t^{-2}$ (left) and $a(t) = it + t^{-3}$ (right). The pictures show $a(\mathbf{T})$ and the eigenvalues of $T_{50}(a)$.*

2. Polynomial numerical hulls.

Let A_n be an $n \times n$ matrix. The *polynomial numerical hull $G_k(A_n)$ of degree k* is defined as

$$G_k(A_n) = \{z \in \mathbb{C} : |p(z)| \le \|p(A_n)\| \text{ for all } p \in \mathcal{P}_k^+\},$$

where $\mathcal{P}_k^+$ is the set of all polynomials of the form

$$p(z) = p_0 + p_1 z + \cdots + p_k z^k.$$

The objective of $G_k(A_n)$ is to employ the obvious inequality

$$\|p(A_n)\| \ge \max_{z \in G_k(A_n)} |p(z)|$$

in order to get a lower estimate for $\|p(A_n)\|$. The sets $G_k(A_n)$ were introduced by Olavi Nevanlinna [24], [25] and were independently discovered by Anne Greenbaum [21], [22]. Their works describe various properties of polynomial numerical hulls. In general, polynomial numerical hulls can be computed only numerically (see, e.g., [21]), and the development of algorithms and software for polynomial numerical hulls is still far away from the advanced level of the pseudospectra counterpart [37], [38].

Faber, Greenbaum, and Marshall [17] obtained pretty precise results on the polynomial numerical hulls of Jordan blocks. Let $J_n(\lambda)$ be the Jordan block (1.6). If k is greater than the degree of the minimal polynomial of A_n, then $G_k(A_n)$ collapses to the spectrum of A_n. Thus, $G_k(J_n(\lambda)) = \{\lambda\}$ for $k \ge n$. In [17], it is shown that if $1 \le k \le n-1$, then

ETNA
Kent State University
etna@mcs.kent.edu

$G_k(J_n(\lambda))$ is a closed disk with the center λ whose radius $\varrho_{k,n}$ satisfies

$$\cos\frac{\pi}{n+1} = \varrho_{1,n} \geq \varrho_{k,n} \geq \varrho_{n-1,n} = 1 - \frac{\log(2n)}{n} + \frac{\log(\log(2n))}{n} + o\left(\frac{1}{n}\right).$$

Consequently,

$$\|J_n^k(\lambda)\| \geq \max_{|\zeta-\lambda|\leq\varrho_{k,n}} |\zeta|^k = (|\lambda| + \varrho_{k,n})^k.$$

It is also shown in [17] that $\varrho_{n-1,n}$ is greater than or equal to the positive root of $2r^n + r - 1 = 0$, which implies that

$$\varrho_{n-1,n} > 1 - \frac{\log(2n)}{n} \;\text{ for all }\; n.$$

In fact, in [17] it is proved that $\varrho_{n-1,n}$ is extremely close to the positive root of $2r^n + r - 1 = 0$; if n is even, one actually has equality. This yields, for instance, $\|J_{100}^{50}(0.8)\| \geq 10^{12.11}$. Matlab gives us the exact value: $\|J_{100}^{50}(0.8)\| = 10^{12.76}$.

It should be mentioned that once Faber, Greenbaum, and Marshall had completed the original version of [17], they observed that their asymptotic formula

$$\varrho_{n-1,n} = 1 - \frac{\log(2n)}{n} + \frac{\log(\log(2n))}{n} + o\left(\frac{1}{n}\right)$$

is in fact a very old result. Namely, the problem of determining $\varrho_{n-1,n}$ is equivalent to a classical problem in complex approximation theory (closely related to the Carathéodory-Fejér interpolation problem) which was explicitly solved by Schur and Szegö [28] and then rediscovered with a different proof by Goluzin [19], [20, Theorem 6, pp. 522-523]. Anne Greenbaum wrote [23]: “It indicates that it is a fundamental question that probably has application in a lot of areas, so this gives one more confidence that the idea of the polynomial numerical hull is a fundamental concept that we should be investigating.”

We also remark that a matrix with a norm that is gigantic in comparison with the matrix dimension must have a gigantic entry. For $n > k$, the ℓth entry of the first column of the lower-triangular Toeplitz matrix $J_n^k(\lambda)$ is $\binom{k}{\ell}\lambda^{k-\ell}$. Taking the value $\ell = [k/(1+|\lambda|)]$, where $[\cdot]$ denotes the integer part, we get

$$\|J_n^k(\lambda)\| \geq \binom{k}{[k/(1+|\lambda|)]} |\lambda|^{k-[k/(1+|\lambda|)]}.$$

This simple observation delivers

$$\|J_{100}^{50}(0.8)\| \geq \binom{50}{27} 0.8^{23} = 10^{11.80},$$

which is fairly good.

3. Pseudospectra and resolvent norm. Let A be a bounded operator on some Hilbert space H. Note that we may think of an $n \times n$ matrix as an operator on $\mathbb{C}^n$ with the ℓ^2 norm. For $\varepsilon > 0$, the ε-*pseudospectrum* $\mathrm{sp}_\varepsilon A$ is defined as the set

$$\mathrm{sp}_\varepsilon A = \{\lambda \in \mathbb{C} : \|(A - \lambda I)^{-1}\| \geq 1/\varepsilon\}, \tag{3.1}$$

ETNA
Kent State University
etna@mcs.kent.edu

with the convention that $\|(A-\lambda I)^{-1}\| = +\infty$ if $A-\lambda I$ is not invertible. A great deal of information about pseudospectra is in [11], [12], [16], [26], [29], [30]. We note in particular that $\mathrm{sp}_\varepsilon A$ admits the alternative description

$$\mathrm{sp}_\varepsilon A = \bigcup_{\|E\|\le\varepsilon} \mathrm{sp}\,(A+E), \tag{3.2}$$

where $\mathrm{sp}\,(A+E)$ denotes the spectrum of $A+E$ and the union in (3.2) is over all bounded operators E on H whose norm does not exceed ε.

Tom Wright's package EigTool [37] provides us with fantastic software for computing pseudospectra numerically. In 1999, Trefethen [31] wrote: "In 1990, getting a good plot of pseudospectra on a workstation for a 30×30 matrix took me several minutes. Today I would expect the same of a 300×300 matrix, and pseudospectra of matrices with dimensions in the thousands are around the corner." Concerning the computation of pseudospectra of dense matrices, Wright notes in his 2002 thesis [38]: "What was once a very expensive computation has become one that is practically of a similar order of complexity to that of computing eigenvalues."

The role of pseudospectra in connection with the norms of powers of matrices and operators is as follows: norms of powers can be related to the resolvent norm, and pseudospectra decode information about the resolvent norm in a visual manner. A relation between resolvent and power norms is established by the Kreiss matrix theorem. Let A_n be an $n\times n$ matrix with $\varrho(A_n)\le 1$. For λ outside the closed unit disk $\overline{\mathbf{D}}$, put $R_n(\lambda) = \|(A_n-\lambda I)^{-1}\|$. The Kreiss matrix theorem says that

$$\sup_{|\lambda|>1}(|\lambda|-1)R_n(\lambda) \le \max_{k\ge 0}\|A_n^k\| \le en \sup_{|\lambda|>1}(|\lambda|-1)R_n(\lambda).$$

We refer the reader to [33] for a delightful discussion of the theorem. The "easy half" of the Kreiss matrix theorem is the lower estimate. It implies that if we pick a point $\lambda\in\mathbb{C}$ with $\varrho := |\lambda| > 1$, then the maximum of the norms $\|A_n^k\|$ $(k\ge 0)$ is at least $(\varrho-1)R_n(\lambda)$. Thus, the maximum is seen to be large whenever we can find a λ outside $\overline{\mathbf{D}}$ with large resolvent norm.

The following theorem and its proof are due to Nick Trefethen [32]. This theorem is similar to but nevertheless slightly different from the lower estimate of the Kreiss matrix theorem.

THEOREM 3.1. *Let A be a bounded linear operator and let $\lambda\in\mathbb{C}\setminus\mathrm{sp}\,A$. Put $\varrho = |\lambda|$ and $R(\lambda) = \|(A-\lambda I)^{-1}\|$. If $R(\lambda) > 1/\varrho$, then*

$$\max_{1\le j\le k}\|A^j\| \ge \varrho^k \Bigg/ \left(1 + \frac{\varrho^k-1}{\varrho-1}\,\frac{1}{\varrho R(\lambda)-1}\right). \tag{3.3}$$

Proof. Put $M_k = \max_{1\le j\le k}\|A^j\|$. By assumption, $\varrho R(\lambda)-1$ is positive, so the term in the large brackets of (3.3) is at least 1 and the right-hand side of (3.3) is thus at most ϱ^k. If ϱ does not exceed the spectral radius of A, then (3.3) is accordingly trivial. We may therefore assume that ϱ is larger than the spectral radius of A. In this case we have

$$\lambda(\lambda I - A)^{-1} = (I-\lambda^{-1}A)^{-1} = I + (\lambda^{-1}A) + (\lambda^{-1}A)^2 + (\lambda^{-1}A)^3 + \cdots$$

and the corresponding bound

$$\varrho R(\lambda) \le 1 + \varrho^{-1}\|A\| + \varrho^{-2}\|A^2\| + \varrho^{-3}\|A^3\| + \cdots. \tag{3.4}$$

ETNA
Kent State University
etna@mcs.kent.edu

Let us take the case $k = 2$ for illustration. Clearly,

$$\|A\|, \|A^2\| \le M_2, \quad \|A^3\|, \|A^4\| \le M_2^2, \quad \|A^5\|, \|A^6\| \le M_2^3,$$

and so on. Grouping the terms in (3.4) accordingly into pairs gives

$$\varrho R(\lambda) \le 1 + (\varrho^{-2} M_2)(1+\varrho) + (\varrho^{-2} M_2)^2 (1+\varrho) + \cdots .$$

If $M_2 \ge \varrho^2$, then (3.3) is true, so we may assume that $\varrho^{-2} M_2 < 1$. In this case

$$\varrho R(\lambda) \le 1 + \frac{\varrho^{-2} M_2 (1+\varrho)}{1 - \varrho^{-2} M_2} = 1 + \frac{\varrho + 1}{\varrho^2 / M_2 - 1}.$$

For general k, we obtain similarly that $\varrho R(\lambda)$ is at most

$$\begin{aligned} &1 + (\varrho^{-k} M_k)(1 + \cdots + \varrho^{k-1}) + (\varrho^{-k} M_k)^2 (1 + \cdots + \varrho^{k-1}) + \cdots \\ &= 1 + \frac{1 + \varrho + \cdots + \varrho^{k-1}}{\varrho^k / M_k - 1}, \end{aligned}$$

that is,

$$\varrho R(\lambda) - 1 \le \frac{\varrho^k - 1}{\varrho - 1} \frac{1}{\varrho^k / M_k - 1}.$$

It follows that

$$\frac{1}{\varrho R(\lambda) - 1} \ge \frac{\varrho - 1}{\varrho^k - 1} \left(\frac{\varrho^k}{M_k} - 1 \right),$$

which implies that

$$\frac{\varrho^k}{M_k} - 1 \le \frac{\varrho^k - 1}{\varrho - 1} \frac{1}{\varrho R(\lambda) - 1}.$$

This is (3.3). □

Thus, if $\operatorname{sp} A$ is known to be a subset of the closed unit disk $\overline{\mathbf{D}}$ and if there is an $\varepsilon \le 1$ such that the pseudospectrum $\operatorname{sp}_\varepsilon A$ contains points outside $\overline{\mathbf{D}}$, then, by (3.1), each point $\lambda \in \operatorname{sp}_\varepsilon A \setminus \overline{\mathbf{D}}$ yields an estimate

$$\max_{1 \le j \le k} \|A^j\| \ge \varrho^k \Big/ \left(1 + \frac{\varrho^k - 1}{\varrho - 1} \frac{1}{\varrho/\varepsilon - 1} \right) \tag{3.5}$$

with $\varrho = |\lambda|$. Different choices of λ give different values of the right-hand sides of (3.3) and (3.5), and we want these right-hand sides to be as large as possible. Tom Wright [37], [38] has implemented the search for an optimal λ in EigTool.

Let $\mathcal{R}$ denote the set of all rational functions without poles on the unit circle $\mathbf{T}$. If $a \in \mathcal{R}$, then $a \in L^\infty$. Now fix $a \in \mathcal{R}$ and consider $T_n(a)$. We put

$$M_k(n) = \max_{1 \le j \le k} \|T_n^j(a)\|, \quad M(n) = \lim_{k \to \infty} M_k(n).$$

Clearly, $M(n) = \sup_{k \ge 1} M_k(n)$, that is, $M(n)$ is the height of the highest peak of $\|T_n^k(a)\|$ as k ranges over $\mathbb{N}$. Thus, the powers of $T_n(a)$ go to infinity if and only if $M(n) = \infty$ and they have a critical transient behavior before decaying to zero if and only if $M(n)$ is large but finite. We assume that the plane Lebesgue measure of $\operatorname{sp} T(a)$ is nonzero. Equivalently, we assume that there exist points in the plane that are encircled by the (naturally oriented) curve $a(\mathbf{T})$ with nonzero winding number. Suppose $\|a\|_\infty > 1$. Then there exist points $\lambda \in \mathbb{C} \setminus a(\mathbf{T})$ such that $\varrho = |\lambda| > 1$ and the winding number of a about λ is nonzero. It is well-known that

$$R_n(\lambda) := \|(T_n(a) - \lambda I)^{-1}\| = \|T_n^{-1}(a - \lambda)\|$$

increases at least exponentially (see [26] and [5, Theorem 3.13]), i.e., there exist positive constants $C = C(a, \lambda)$ and $\beta = \beta(a, \lambda)$ such that

$$R_n(\lambda) \geq Ce^{\beta n} \tag{3.6}$$

for all $n \geq 1$. (Note that if a is merely continuous or piecewise continuous, then the growth rate of $R_n(\lambda)$ may drop down to $Cn^{1+\varepsilon}$ or even $Cn(\log n)^{1+\varepsilon}$; see [4] and [7].) Consequently, $R_n(\lambda) > 1$ for all sufficiently large n and we obtain from Theorem 3.1 that

$$M_k(n) \geq \varrho^k \Big/ \left(1 + \frac{\varrho^k - 1}{\varrho - 1}\,\frac{1}{\varrho R_n(\lambda) - 1}\right),$$

and

$$M(n) \geq (\varrho - 1)(\varrho R_n(\lambda) - 1). \tag{3.7}$$

This, in conjunction with (3.6), gives

$$M(n) \geq (\varrho - 1)(\varrho C e^{\beta n} - 1), \tag{3.8}$$

that is, *there must be critical transient or limiting behavior of* $\|T_n^k(a)\|$ *for all* n *larger than a moderately sized* n_0. We will now say something about the constants in (3.6) and (3.8)

We denote by $\mathcal{P}_r$ $(r \geq 1)$ the set of all Laurent polynomials of degree at most r, that is, the set of all functions of the form

$$a(t) = \sum_{j=-r}^{r} a_j t^j. \tag{3.9}$$

Obviously, if $a \in \mathcal{P}_r$, then $T(a)$ is banded with bandwidth $2r + 1$. Let $a \in \mathcal{P}_r$, suppose $\|a\|_\infty > 1$, and choose λ as above. We then have

$$t^r(a(t) - \lambda) = \prod_{i=1}^{s}(t - \mu_i) \prod_{i=s+1}^{2r}(t - \delta_i),$$

with $|\mu_i| > 1$ and $|\delta_i| < 1$. It follows that

$$a(t) - \lambda = t^{-\kappa} \prod_{i=1}^{s}(t - \mu_i) \prod_{i=s+1}^{2r}\left(1 - \frac{\delta_j}{t}\right),$$

where $\kappa = s - r$. Without loss of generality assume that $\kappa \geq 1$ (otherwise replace $a(t)$ by $a(1/t)$, which amounts to replacing $T_n(a)$ by the transpose matrix). Put

$$d(t) = \prod_{i=1}^{s}(t - \mu_i)^{-1} =: \sum_{j=0}^{\infty} d_j t^j \quad (t \in \mathbf{T}), \tag{3.10}$$

$$\|P_n d\|^2 = \sum_{j=0}^{n-1} |d_j|^2, \quad \|Q_n d\|^2 = \sum_{j=n}^{\infty} |d_j|^2. \tag{3.11}$$

The following result is a slight modification of Theorem 3.13 of [5].

THEOREM 3.2. *For all $n \geq 1$,*

$$R_n(\lambda) \geq \frac{\|P_n d\|}{\|a-\lambda\|_\infty \|Q_n d\|}.$$

Proof. Put $b = a - \lambda$. With the notation $\chi_k(t) = t^k$, we have $b = \chi_{-\kappa} b_- b_+$ (a so-called Wiener-Hopf factorization), where

$$b_-(t) = \prod_{i=s+1}^{2r} \left(1 - \frac{\delta_i}{t}\right), \quad b_+(t) = \prod_{i=1}^{s} (t - \mu_i).$$

Define $x^{(n)} \in \mathbb{C}^n$ and $x \in \ell^2$ by

$$x^{(n)} = (d_0, \ldots, d_{n-1}), \quad x = (d_0, d_1, \ldots).$$

Since $T(b) = T(b_-)T(\chi_{-\kappa})T(b_+)$ (see, e.g, [11, Section 1.5]), we obtain

$$T_n(b)x^{(n)} = P_n T(b_-)T(\chi_{-\kappa})T(b_+)x^{(n)}. \tag{3.12}$$

Because $\kappa \geq 1$, we have $P_n T(b_-)T(\chi_{-\kappa})e_0 = 0$, where $e_0 = (1, 0, 0, \ldots)$. As $T(b_+)x = e_0$, it follows that

$$P_n T(b_-)T(\chi_{-\kappa})T(b_+)x = 0. \tag{3.13}$$

From (3.12) and (3.13) we get

$$T_n(b)x^{(n)} = P_n T(b_-)T(\chi_{-\kappa})T(b_+)(x^{(n)} - x) = P_n T(b)(x^{(n)} - x).$$

Taking into account that $\|x^{(n)} - x\| = \|Q_n d\|$, we arrive at the estimate

$$\|T_n(b)x^{(n)}\| \leq \|T(b)\| \, \|Q_b d\| = \|b\|_\infty \|Q_n d\|,$$

and since $\|x^{(n)}\| = \|P_n d\|$, it results that

$$\|T_n^{-1}(b)\| \geq \frac{\|x^{(n)}\|}{\|T_n(b)x^{(n)}\|} \geq \frac{\|P_n d\|}{\|b\|_\infty \|Q_n d\|}. \qquad \square$$

For the sake of simplicity, assume the zeros $\mu_1, \ldots, \mu_s$ are distinct and $|\mu_1| < |\mu_i|$ for all $i \geq 2$. Decomposition into partial fractions gives

$$d(t) = \sum_{i=1}^{s} \frac{A_i}{1 - t/\mu_i} = \sum_{j=0}^{\infty} \sum_{i=1}^{s} \frac{A_i}{\mu_i^j} t^j, \tag{3.14}$$

with explicitly available constants $A_1, \ldots, A_s$. Theorem 3.2 in conjunction with the estimates

$$\|P_n d\| \geq |d_0| = \left| \sum_{i=1}^{s} A_i \right| = \prod_{i=1}^{s} \frac{1}{|\mu_i|}$$

and

$$\|Q_n d\|^2 = \sum_{j=n}^{\infty} \left| \sum_{i=1}^{s} \frac{A_i}{\mu_i^j} \right|^2 \leq s \sum_{i=1}^{s} \sum_{j=n}^{\infty} \frac{|A_i|^2}{|\mu_i|^{2j}} = \sum_{i=1}^{s} \frac{|A_i|^2}{|\mu_i|^2} \frac{s}{1 - 1/|\mu_i|^2}$$

$$=: \sum_{i=1}^{s} \frac{B_i}{|\mu_i|^{2n}} = \frac{B_1}{|\mu_1|^{2n}} \left(1 + \sum_{i=2}^{s} \frac{B_j}{B_1} \left|\frac{\mu_1}{\mu_i}\right|^{2n}\right)$$

yields

$$R_n(\lambda)^2 \geq \frac{|\mu_1|^{2n}}{B_1\|a-\lambda\|_\infty^2 \prod_{i=1}^s |\mu_i|^2} \frac{1}{1+\sum_{i=2}^s (B_j/B_1)|\mu_1/\mu_j|^{2n}}. \tag{3.15}$$

In practice, we could try computing $R_n(\lambda)$ directly via the Matlab command $R_n(\lambda) =$ $\texttt{norm}(\texttt{inv}(T_n(a-\lambda)))$ or $R_n(\lambda) = 1/\texttt{min}(\texttt{svd}(T_n(a-\lambda)))$. However, as (3.6) shows, this is an ill-conditioned problem for large n. Estimate (3.15) is more reliable. It first of all shows that (3.6) and (3.8) are true with

$$\beta = \log|\mu_1|.$$

To get the constants contained in (3.15) we may proceed as follows. We first determine the zeros $\mu_1, \ldots, \mu_s$, we then use Matlab's RESIDUE command to find the numbers A_i in formula (3.14), and finally we define $B_i = s|A_i|^2/(1-1/|\mu_i|^2)$. Note that s is in general much smaller than n, so that possible numerical instabilities are no longer caused by large matrix dimensions but at most by unfortunate location of the zeros $\mu_1, \ldots, \mu_s$. Notice also that (3.10) implies that the numbers $d_0, \ldots, d_{s-1}$ are the entries of the first column of the lower-triangular Toeplitz matrix $T_s(d)$. Thus, alternatively we could solve the $s \times s$ system $T_s(d)x = e_0$ to obtain $d_0, \ldots, d_{s-1}$ and then, taking into account (3.14), find the constants $A_1, \ldots, A_s$ as the solutions of the $s \times s$ Vandermonde system

$$\begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1/\mu_1 & 1/\mu_2 & \ldots & 1/\mu_s \\ \vdots & \vdots & & \vdots \\ 1/\mu_1^{s-1} & 1/\mu_2^{s-1} & \ldots & 1/\mu_s^{s-1} \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_s \end{pmatrix} = \begin{pmatrix} d_0 \\ d_1 \\ \vdots \\ d_{s-1} \end{pmatrix}.$$

Here is an example that can be done by hand.

EXAMPLE 3.3. Let $a(t) = t^{-1} + \alpha^2 t$ with $0 < \alpha < 1/2$. The range $a(\mathbf{T})$ is the ellipse

$$\frac{x^2}{(1+\alpha^2)^2} + \frac{y^2}{(1-\alpha^2)^2} = 1$$

and the eigenvalues of the matrix $T_n(a)$ are densely spread over the interval $(-2\alpha, 2\alpha)$ between the foci of the ellipse. The spectral radius is

$$\varrho(T_n(a)) = 2\alpha \cos \frac{\pi}{n+1}.$$

This is smaller than 1 but may be close to 1. The norm of the symbol is given by $\|a\|_\infty =$ $1+\alpha^2 > 1$. Fix $\lambda = \varrho \in (1, 1+\alpha^2)$. The zeros of $t(a(t)-\varrho)$ are

$$\mu_1 = \frac{\varrho - \sqrt{\varrho^2-\alpha^2}}{2\alpha^2}, \quad \mu_2 = \frac{\varrho + \sqrt{\varrho^2-\alpha^2}}{2\alpha^2}.$$

The numbers μ_1 and μ_2 are greater than 1, and we have

$$A_1 = \frac{1}{\mu_1-\mu_2}, \quad A_2 = \frac{1}{\mu_2-\mu_1},$$
$$B_1 = \frac{2A_1^2}{1-1/\mu_1^2}, \quad B_2 = \frac{2A_2^2}{1-1/\mu_2^2}.$$

ETNA
Kent State University
etna@mcs.kent.edu

Thus, (3.15) becomes

$$R_n(\lambda)^2 \geq \frac{\alpha^4 \mu_1^{2n}}{B_1(\varrho+1+\alpha^2)^2} \Bigg/ \left(1+\frac{B_2}{B_1}\left(\frac{\mu_1}{\mu_2}\right)^{2n}\right).$$

Here are a few concrete samples. In each case we picked $\lambda = \varrho = 1.01$. Note that (3.7) with $\varrho = 1.01$ gives

$$M(n) \geq 0.1(1.01\, R_n(1.01) - 1),$$

which is almost the same as the Kreiss estimate $M(n) \geq 0.1 R_n(1.01)$.
$\boldsymbol{\alpha = 0.2.}$ We have $\mu_1 = 1.0323$, $\mu_2 = 24.2177$,

$$R_n(1.01) \geq 10^{0.0138\, n - 1.0863} =: E_{0.2}(n).$$

In particular, $E_{0.2}(1000) = 10^{12.70}$. Matlab result: $R_{1000}(1.01) = 10^{15.03}$.
$\boldsymbol{\alpha = 0.4.}$ Now $\mu_1 = 1.2296$, $\mu_2 = 5.0829$,

$$R_n(1.01) \geq 10^{0.0897\, n - 0.9316} =: E_{0.4}(n).$$

We get $E_{0.4}(100) = 10^{8.04}$ and $E_{0.4}(1000) = 10^{88.77}$, while Matlab gives $R_{100}(1.01) = 10^{9.48}$ and $R_{1000}(1.01) = 10^{90.28}$.
$\boldsymbol{\alpha = 0.49.}$ This time $\mu_1 = 1.5945$, $\mu_2 = 2.6121$,

$$R_n(1.01) \geq 10^{0.2026\, n - 1.2233} =: E_{0.49}(n).$$

We obtain

$$E_{0.49}(50) = 10^{8.91}, \quad E_{0.49}(100) = 10^{19.04},$$
$$E_{0.49}(1000) = 10^{201.38}, \quad E_{0.49}(2000) = 10^{403.98},$$

and Matlab delivers

$$R_{50}(1.01) = 10^{10.22}, \quad R_{100}(1.01) = 10^{20.35}, \quad R_{1000}(1.01) = 10^{202.71},$$

the last two values with a warning. Matlab returns a warning without a value for $n = 2000$. Thus, we have eventually beaten the unbeatable Matlab. □

4. Simply norms. In this section we collect some known results on the asymptotic behavior of the norms of Toeplitz-like matrices.

Frequently one has not to deal with pure Toeplitz matrices but with matrices that are composed of Toeplitz matrices. Let $\mathcal{B}(\ell^2)$ be the C^*-algebra of all bounded linear operators on ℓ^2 and let $\mathbf{A}(L^\infty)$ denote the smallest closed subalgebra of $\mathcal{B}(\ell^2)$ that contains all Toeplitz operators $T(a)$ with $a \in L^\infty$. For example, if $\{a_{jk}\}$ is a finite collection of functions in L^∞, then

$$\sum_i \prod_j T(a_{jk}) \in \mathbf{A}(L^\infty). \tag{4.1}$$

Notice also that $e^{\tau T(a)} \in \mathbf{A}(L^\infty)$ for each $\tau \in \mathbb{C}$ and each $a \in L^\infty$, because

$$\left\| e^{\tau T(a)} - \left(I + \frac{\tau}{1!}T(a) + \frac{\tau^2}{2!}T^2(a) + \ldots + \frac{\tau^N}{N!}T^N(a)\right)\right\| \to 0$$

as $N \to \infty$. We define $V^{(-n)}$ and V^n on ℓ^2 by

$$V^{(-n)} : (x_1, x_2, x_3, \ldots) \mapsto (x_{n+1}, x_{n+2}, \ldots),$$
$$V^n : (x_1, x_2, x_3, \ldots) \mapsto (0, \ldots, 0, x_1, x_2, \ldots) \quad (n \text{ zeros}).$$

THEOREM 4.1. *If $A \in \mathbf{A}(L^\infty)$, then the strong limit*

$$s\text{-}\lim_{n\to\infty} V^{(-n)} A V^n$$

exists and is a Toeplitz operator $T(a)$ with $a \in L^\infty$.

For the history of this theorem, see [1] and [11, p. 152]. A full proof is in [10]. The function a of Theorem 4.1 is referred to as the *symbol* of the operator A and is denoted by $\text{sym}\, A$. Since, obviously, $\|A\| \ge \|V^{(-n)} A V^n\|$ for all n, Theorem 4.1 in conjunction with the Banach-Steinhaus theorem implies the useful estimate

$$\|A\| \ge \|T(\text{sym}\, A)\| = \|\text{sym}\, A\|_\infty. \tag{4.2}$$

The symbol of the operator (4.1) and of $e^{\tau T(a)}$ is $\sum_j \prod_k a_{jk}$ and $e^{\tau a}$, respectively.

Let $\mathcal{F}$ denote the set of all sequences $\{A_n\}_{n=1}^\infty$, where A_n is an $n \times n$ matrix and

$$\|\{A_n\}_{n=1}^\infty\| := \sup_{n\ge 1} \|A_n\| < \infty. \tag{4.3}$$

The set $\mathcal{F}$ becomes a C^*-algebra with the norm (4.3) and the operations

$$\{A_n\} + \{B_n\} = \{A_n + B_n\}, \quad \alpha\{A_n\} = \{\alpha A_n\},$$
$$\{A_n\}\{B_n\} = \{A_n B_n\}, \quad \{A_n\}^* = \{A_n^*\},$$

A_n^* denoting the Hermitian adjoint of A_n. Given a subset B of L^∞, we define $\mathbf{S}(B)$ as the smallest closed subalgebra of $\mathcal{F}$ that contains all elements of the form $\{T_n(a)\}_{n=1}^\infty$ with $a \in B$. Important cases are $B = L^\infty$, $B = PC$, and $B = C$. Here C and PC denote the C^*-algebras of all continuous and piecewise continuous functions on $\mathbf{T}$, respectively. A function a on $\mathbf{T}$ is said to be piecewise continuous if $a \in L^\infty$ and the one-sided limits $a(t-0)$ and $a(t+0)$ exist for each t on the counter-clockwise oriented unit circle. For instance, if a_{jk} are finitely many functions in B, then

$$\left\{ \sum_j \prod_k T_n(a_{jk}) \right\}_{n=1}^\infty \in \mathbf{S}(B), \tag{4.4}$$

and if $a \in B$, then $\left\{e^{\tau T_n(a)}\right\}_{n=1}^\infty \in \mathbf{S}(B)$ for each $\tau \in \mathbb{C}$.

The operator W_n is defined on ℓ^2 by

$$W_n : (x_1, x_2, x_3, \ldots) \mapsto (x_n, x_{n-1}, \ldots, x_1, 0, 0, \ldots).$$

Thus, W_n is P_n followed by reversal of the coordinates.

THEOREM 4.2. *Let $\{A_n\}_{n=1}^\infty \in \mathbf{S}(L^\infty)$. Then the sequences $\{A_n\}_{n=1}^\infty$ and $\{W_n A_n W_n\}_{n=1}^\infty$ converge strongly to the operators A and $\widetilde{A}$ in $\mathbf{A}(L^\infty)$, respectively, and*

$$\liminf_{n\to\infty} \|A_n\| \ge \max(\|A\|, \|\widetilde{A}\|). \tag{4.5}$$

If $\{A_n\}_{n=1}^\infty \in \mathbf{S}(PC)$, then

$$\lim_{n\to\infty} \|A_n\| = \max(\|A\|, \|\widetilde{A}\|). \tag{4.6}$$

ETNA
Kent State University
etna@mcs.kent.edu

A proof of the existence of the strong limits A and $\widetilde{A}$ is in [10], for example. Estimate (4.5) is a simple consequence of the equality $\|W_n A_n W_n\| = \|A_n\|$ and the Banach-Steinhaus theorem. Equality (4.6) was established in [3], and a full proof is also contained in [11].

An easy computation shows that $W_n T_n(a) W_n = T_n(\widetilde{a})$, where $\widetilde{a}(t) := a(1/t)$ $(t \in \mathbf{T})$. In other terms, $T_n(\widetilde{a})$ is simply the transpose of $T_n(a)$. If $\{A_n\}$ is of the form (4.4), then

$$A = \sum_j \prod_k T(a_{jk}), \quad \widetilde{A} = \sum_j \prod_k T(\widetilde{a}_{jk}).$$

In particular, for $A_n = T_n^k(a)$ we obtain $A = T^k(a)$ and $\widetilde{A} = T^k(\widetilde{a})$, so that $\widetilde{A}$ is just the transpose of A. Note also that (4.5) can be sharpened to (1.4) in this case. If $A_n = (T_n(b)T_n(c))^k$, then $A = (T(b)T(c))^k$ and $\widetilde{A} = (T(\widetilde{b})T(\widetilde{c}))^k$. Now $\widetilde{A}$ is not necessarily the transpose of A. Combining Theorem 4.1 and (4.2) we obtain

$$\begin{aligned} \liminf_{n\to\infty} \|(T_n(b)T_n(c))^k\| &\geq \max(\|(T(b)T(c))^k\|, \|(T(\widetilde{b})T(\widetilde{c}))^k\|) \\ &\geq \max(\|(bc)^k\|_\infty, \|(bc)^k\|_\infty) = \|bc\|_\infty^k. \end{aligned} \tag{4.7}$$

We also remark that it may happen that $\|\widetilde{A}\| > \|A\|$. For example, Proposition 4.12 of [5] implies that if

$$b(t) = \frac{t-\delta}{t-\gamma} \quad \text{and} \quad c(t) = \frac{1-\gamma t}{1-\delta t} \quad (t \in \mathbf{T})$$

with $|\gamma| < 1$ and $|\delta| < 1$, then $\|T(b)T(c)\| < \|T(\widetilde{b})T(\widetilde{c})\|$.

The next question after Theorems 1.1 and 4.2 is the question about the convergence speed. Recall that $\mathcal{R}$ stands for the rational functions without poles on $\mathbf{T}$. The following result is from [6] (also see [5, Theorem 4.1]).

THEOREM 4.3. *Let $a \in \mathcal{R}$ and assume that the modulus $|a|$ is not constant on $\mathbf{T}$. Denote by $2\gamma \in \{2, 4, 6, \ldots\}$ the maximal order of the zeros of $\|a\|_\infty - |a(t)|$ for $t \in \mathbf{T}$. Then there exist constants c, d such that $0 < c < d < \infty$ and*

$$\|a\|_\infty - \frac{d}{n^{2\gamma}} \leq \|T_n(a)\| \leq \|a\|_\infty - \frac{c}{n^{2\gamma}} \tag{4.8}$$

for all $n \geq 1$. On the other hand, if $|a|$ is constant on $\mathbf{T}$, then there are constants d and δ in $(0, \infty)$ such that for all $n \geq 1$,

$$\|a\|_\infty - de^{-\delta n} \leq \|T_n(a)\| \leq \|a\|_\infty.$$

EXAMPLE 4.4. Let $a \in \mathcal{R}$ be of the form $a(t) = \sum_{j=0}^\infty a_j t^j$, that is, suppose $T(a)$ is lower-triangular. Assume that $|a|$ is not constant on $\mathbf{T}$. We have $T_n^k(a) = T_n(a^k)$, and hence (4.8) is applicable. It follows that there are constants c_k and d_k such that

$$\|a\|_\infty^k - \frac{d_k}{n^{2\gamma}} \leq \|T_n^k(a)\| \leq \|a\|_\infty^k - \frac{c_k}{n^{2\gamma}}$$

for all $n \geq 1$. However, the constants c_k and d_k may be large if k is large. We remark that the constants C of inequalities like

$$\|a\|_\infty - |a(t)| \leq C|t - t_0|^{2\gamma} \tag{4.9}$$

enter into the d of (4.8). For the powers of a, we obtain from (4.9) something like

$$\|a\|_\infty^k - |a^k(t)| \lessapprox Ck\|a\|_\infty^k |t-t_0|^{2\gamma},$$

and since $Ck\|a\|_\infty^k$ is large whenever $\|a\|_\infty > 1$ and k is large, we can expect that d_k is also large.

Let us consider a concrete example. Take

$$a(t) = 10\left(-\frac{1}{16} - \frac{1}{4}t + t^2 - \frac{1}{4}t^3 - \frac{1}{16}t^4\right) = 10t^2\left(\frac{11}{8} - \frac{1}{16}|1-t|^4\right).$$

In this case $2\gamma = 4$. The spectral radius of $T_n(a)$ is about 0.63 and $\|a\|_\infty$ equals $110/8 = 13.75$. The level curves of the norm surface are shown in Figure 4.1 and we clearly see the expected critical transient behavior. An interesting feature of Figure 4.1 is the indents of the level curves. These indents correspond to vertical valleys in the surface. For instance, along the horizontal line $n = 18$ we have Figure 4.2. The interesting piece of Figure 4.2 is between $k = 15$ and $k = 40$. If we move vertically along one of the lines $k = 15$, $k = 20, \ldots$, $k = 40$, we obtain Figure 4.3. Figure 4.4 reveals some mild turbulency in the steep, but Figure 4.3 convincingly shows that in the sky region everything goes smoothly. In particular, the speed of the convergence of $\|T_n^k(a)\|$ to $\|T^k(a)\| = \|a\|_\infty^k$ is not affected by small fluctuations of k. $\square$

We now consider matrices of the form

$$A_n = T_n(b) + P_n X P_n + W_n Y W_n, \tag{4.10}$$

where $b \in \mathcal{R}$ and X and Y are "small" in comparison with the "Toeplitz part" $T_n(b)$. To be more precise, we suppose that the moduli $|X_{jk}|$ and $|Y_{jk}|$ of the entries are bounded from above by $C\sigma^{j+k}$ with constants $C < \infty$ and $\sigma \in (0,1)$. These conditions guarantee that $\{A_n\}$ belongs to $\mathbf{S}(C)$. The operators A and $\widetilde{A}$ of Theorem 4.2 can be shown to be

$$A = T(b) + X, \quad \widetilde{A} = T(\widetilde{b}) + Y.$$

From Theorem 4.2 we infer that $\|A_n\| \to N$ as $n \to \infty$, where

$$N := \max(\|T(b) + X\|, \|T(\widetilde{b}) + Y\|).$$

Put

$$N_0 := \|T(b)\| = \|b\|_\infty.$$

Since $\operatorname{sym}(T(b) + X) = b$ and $\operatorname{sym}(T(\widetilde{b}) + Y) = \widetilde{b}$, we deduce from (4.2) that always $N_0 \leq N$.

ETNA
Kent State University
etna@mcs.kent.edu

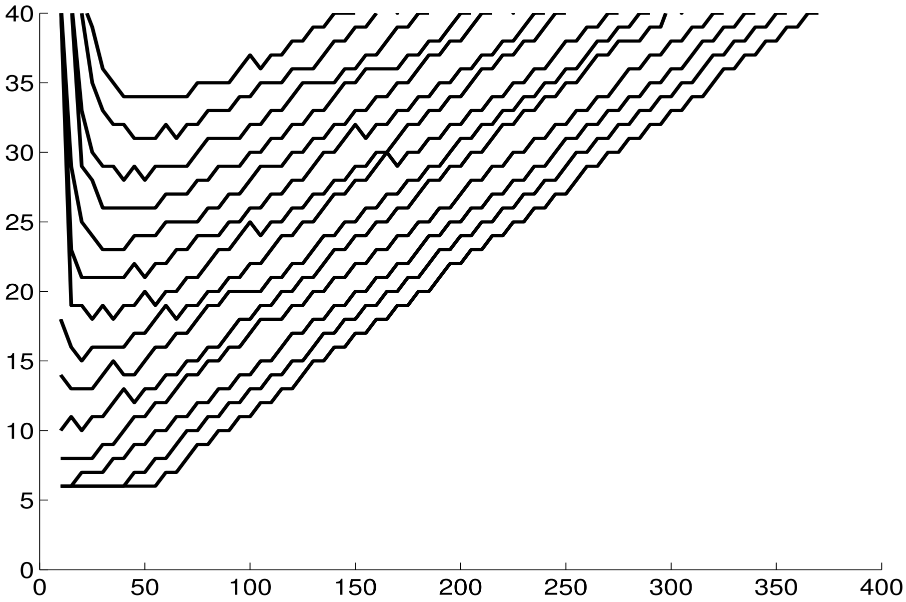

FIG. 4.1. *The symbol is* $a(t) = 10\left(-\frac{1}{16} - \frac{1}{4}t + t^2 - \frac{1}{4}t^3 - \frac{1}{16}t^4\right)$. *The plot shows the level curves* $\|T_n^k(a)\| = c$ *for* $c = 10^{-2}, 1, 10^2, 10^4, \ldots, 10^{24}$ *(the lower curve corresponds to* $c = 10^{-2}$*, the upper to* 10^{24}*). We took* $6 \le n \le 40$ *and* $k = 10, 15, 20, \ldots, 400$.

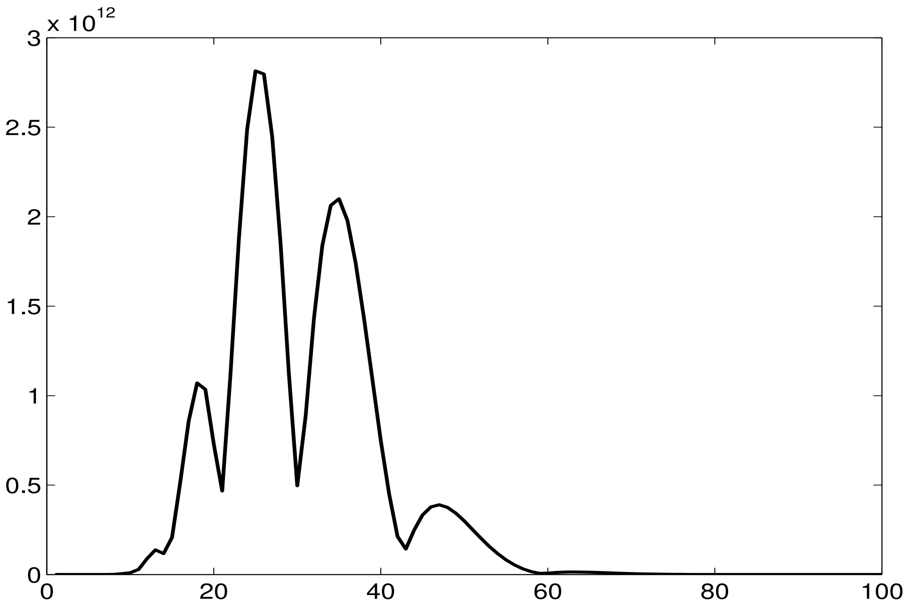

FIG. 4.2. *The symbol is as in Figure 4.1. The picture shows the norms* $\|T_{18}^k(a)\|$ *for* $1 \le k \le 100$.

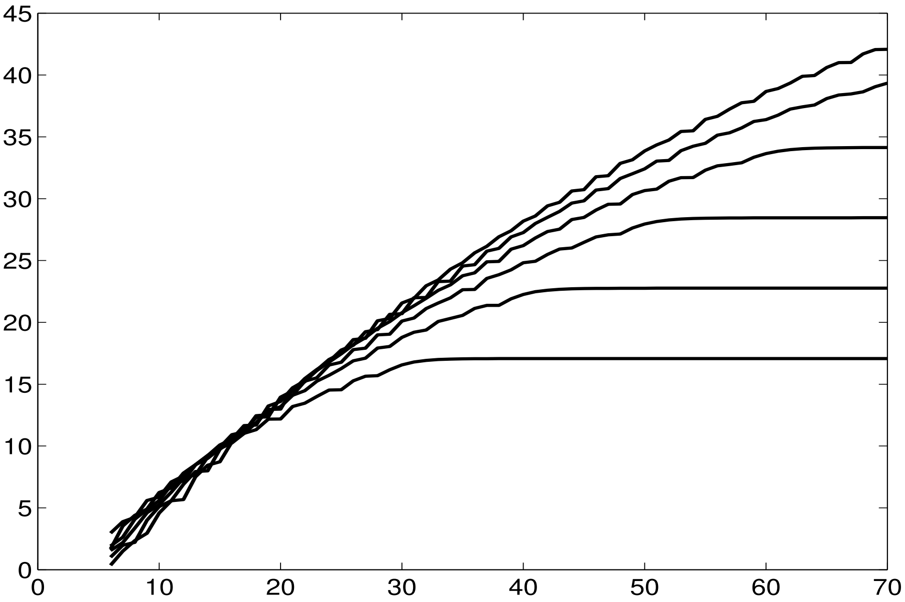

FIG. 4.3. *The symbol is again as in Figure 4.1. We see the values of of* $\log_{10} \|T_n^k(a)\|$ *for* $6 \le n \le 70$ *and* $k = 15, 20, 25, \ldots, 40$*. Eventually higher curves correspond to higher values of* k*.*

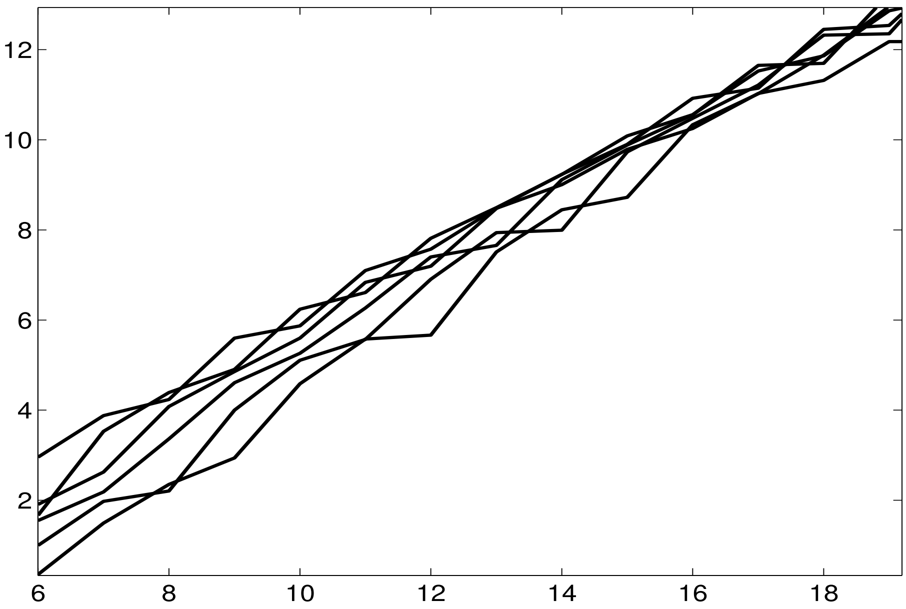

FIG. 4.4. *A close-up of Figure 4.3.*

ETNA
Kent State University
etna@mcs.kent.edu

The following theorem was established in [6] (it is also Theorem 4.16 of [5]).

THEOREM 4.5. *If* $N > N_0$, *then*

$$\Big| \|A_n\| - N \Big| = O\left(e^{-\gamma\sqrt{n}}\right)$$

for some $\gamma > 0$, *and if* $N = N_0$, *then*

$$\Big| \|A_n\| - N \Big| = O\left(\frac{\log n}{n}\right).$$

Now pick $a \in \mathcal{R}$. One can show that then $A_n = T_n^k(a)$ is of the form (4.10) up to an exponentially small term. The b equals a^k. Thus,

$$T_n^k(a) = T_n(a^k) + P_n X_k P_n + W_n Y_k W_n + E_n, \tag{4.11}$$

where $\|E_n\| = O(e^{-\delta n})$ for some $\delta > 0$. Consequently, we can apply Theorem 4.5 to this situation. Since $A = T^k(a)$ and $\widetilde{A} = T^k(\widetilde{a})$, it follows that $N = N_0$, and hence, we cannot extract more than $O(\log n/n)$ convergence speed, which is nevertheless more than nothing.

Finally, suppose A_n is of the form (4.10) with some Laurent polynomial $b \in \mathcal{P}_s$ and with $s \times s$ matrices X and Y. In this case the results of [9] (see also Theorems 4.3 and 4.4 of [5]) give the following.

THEOREM 4.6. *If* $n \geq 41s$, *then*

$$N_0\left(1 - \frac{41s}{n}\right) \leq \|A_n\| \leq N\left(1 + \frac{2s}{n}\right).$$

In particular, for $N = N_0$ *and* $n \geq 41s$ *we have*

$$N\left(1 - \frac{41s}{n}\right) \leq \|A_n\| \leq N\left(1 + \frac{2s}{n}\right). \tag{4.12}$$

If $N > N_0$ *and* $n \geq 8s + 2$, *then*

$$N(1 - c\varrho^n) \leq \|A_n\| \leq N(1 + c\varrho^n),$$

where

$$c = 2s\left(\frac{N}{N_0}\right)^{2+1/(2s)}, \quad \varrho = \left(\frac{N_0}{N}\right)^{1/(4s)}.$$

If $a \in \mathcal{P}_r$, then $T_n^k(a)$ differs from $T_n(a^k)$ only by $r(k-1) \times r(k-1)$ matrices in the upper-left and lower-right corners, and, moreover, these matrices are independent of n (see Lemma 6.1). In other words, (4.11) is valid with $s = r(k-1)$. Thus, Theorem 4.6 with $s = r(k-1)$ and $N = N_0 = \|a\|_\infty^k$ yields the inequality

$$\|a\|_\infty^k\left(1 - \frac{41r(k-1)}{n}\right) \leq \|T_n^k(a)\| \leq \|a\|_\infty^k \tag{4.13}$$

for all n and k. Inequality (4.13) will have its grand scene in Section 7.

5. Gauss-Seidel for large Toeplitz matrices. To solve the $n \times n$ system $C_n x = y$, one decomposes C_n into a sum $C_n = L_n + U_n$ of a lower-triangular matrix L_n and an upper-triangular matrix U_n with zeros on the main diagonal. If L_n is invertible, then the system $C_n x = y$ is equivalent to the system

$$x = -L_n^{-1} U_n x + L_n^{-1} y. \tag{5.1}$$

Gauss-Seidel iteration consists in choosing an initial x_0 and in computing the iterations by

$$x_{k+1} = -L_n^{-1} U_n x_k + L_n^{-1} y. \tag{5.2}$$

Sometimes (5.1) and (5.2) are written in the form

$$x = x + L_n^{-1}(y - C_n x), \quad x_{k+1} = x_k + L_n^{-1}(y - C_n x_k).$$

The iteration matrix is $A_n = -L_n^{-1} U_n = I - L_n^{-1} C_n$, and the iteration converges whenever $\varrho(A_n) < 1$. The problem is whether a critical transient behavior of the norms $\|A_n^k\|$ may garble the solution.

Now suppose $C_n = T_n(c)$ is a Toeplitz matrix and let us, for the sake of simplicity, assume that

$$c(t) = \sum_{j=-\infty}^{\infty} c_j t^j, \quad \text{where} \quad \sum_{j=-\infty}^{\infty} |c_j| < \infty.$$

We write $C_n = L_n + U_n = T_n(c_+) + T_n(c_-)$ with

$$c_+(t) = \sum_{j=0}^{\infty} c_j t^j, \quad c_-(t) = \sum_{j=-\infty}^{-1} c_j t^j.$$

If $c_+(z) \neq 0$ for $|z| \leq 1$, then the inverse of $T_n(c_+)$ is $T_n(c_+^{-1})$ and the iteration matrix becomes

$$A_n = -T_n(c_+^{-1}) T_n(c_-).$$

The spectrum of A_n is the set of all complex numbers λ for which

$$-T_n(c_+^{-1}) T_n(c_-) - \lambda I = -T_n(c_+^{-1}) T_n(c_- + \lambda c_+)$$

is not invertible. Since $T_n(c_+^{-1})$ is invertible by assumption, we have to look for the λ's for which $T_n(c_- + \lambda c_+)$ is not invertible. From (4.7) we deduce that

$$\liminf_{n\to\infty} \|A_n^k\| \geq \|c_+^{-1} c_-\|_\infty^k, \tag{5.3}$$

and, since $\{A_n\} \in \mathbf{S}(C)$, the limit inferior is actually a limit. Thus, if n is large enough, a critical transient phase will certainly occur in case $\|c_+^{-1} c_-\|_\infty > 1$.

If $a(t) = a_{-1} t^{-1} + a_0 + a_1 t$ (which means that $T(a)$ is tridiagonal), then the eigenvalues of $T_n(a)$ are known to be

$$a_0 + 2\sqrt{a_1 a_{-1}} \cos \frac{\pi j}{n+1} \quad (j = 1, \ldots, n).$$

Hence, in case c is a trinomial, $c(t) = c_{-1} t^{-1} + c_0 + c_1 t$, the matrix $T_n(c_- + \lambda c_+)$ has the eigenvalues

$$\lambda c_0 + 2\sqrt{c_1 c_{-1}} \sqrt{\lambda} \cos \frac{\pi j}{n+1} \quad (j = 1, \ldots, n).$$

It follows that

$$\operatorname{sp} A_n = \left\{ -\frac{4c_1 c_{-1}}{c_0^2} \cos^2 \frac{\pi j}{n+1} : j = 1, \ldots, n \right\},$$

and

$$\varrho(A_n) = \frac{4|c_1|\,|c_{-1}|}{|c_0|^2} \cos^2 \frac{\pi}{n+1}.$$

I learned from [29] that these two formulas are a 1950 result by Frankel [18].

Anne Greenbaum [21] discussed the example

$$c(t) = -1.16t^{-1} + 1 + 0.16t.$$

The spectral radius of A_n is about 0.73. Using (numerically computed) polynomial numerical hulls, she obtained that $\|A_{30}^{29}\| \geq 1.256^{29} \approx 700$ and trying the computer directly, she arrived at the much better result $\|A_{30}^{29}\| \approx 10^4$. Our estimate (5.3) with $c_+(t) = 1 + 0.16t$ and $c_-(t) = -1.16t^{-1}$ gives

$$\liminf_{n\to\infty} \|A_n^{29}\| \geq \|c_+^{-1} c_-\|_\infty^{29} = \max_{t\in\mathbf{T}} \left| \frac{-1.16t^{-1}}{1+0.16t} \right|^{29} \tag{5.4}$$
$$= \left(\frac{1.16}{1-0.16} \right)^{29} = 1.381^{29} = 10^{4.07}.$$

Thus, although (5.4) is an "$n \to \infty$" result, it is already strikingly good for n about 30. Figure 5.1 shows the norm surface.

Nick Trefethen [29] considered the symbol

$$c(t) = t^{-1} - 2 + t.$$

In this case $c_+(t) = -2 + t$ and $c_-(t) = t^{-1}$, so

$$\|c_+^{-1} c_-\|_\infty = \max_{t\in\mathbf{T}} \left| \frac{t^{-1}}{-2+t} \right| = 1,$$

and hence (5.3) does not provide any useful piece of information. However, the brute estimate

$$\|A_n^k\| \leq \|T_n(c_+^{-1})\|^k \|T_n(c_-)\|^k \leq \|c_+^{-1}\|_\infty^k \|c_-\|_\infty^k = \max_{t\in\mathbf{T}} \left| \frac{1}{-2+t} \right|^k \cdot 1^k = 1$$

shows that there is no critical behavior. The spectral radius of A_n is $\cos^2 \frac{\pi}{n+1}$, which is approximately 0.9990325 for $n = 100$. Matlab gives $\varrho(A_{100}) = 0.9990$ (fantastic!) and tells us that the norms $\|A_{100}^k\|$ are $0.9080, 0.6169$, and 0.3804 for $k = 100, 500, 1000$, respectively.

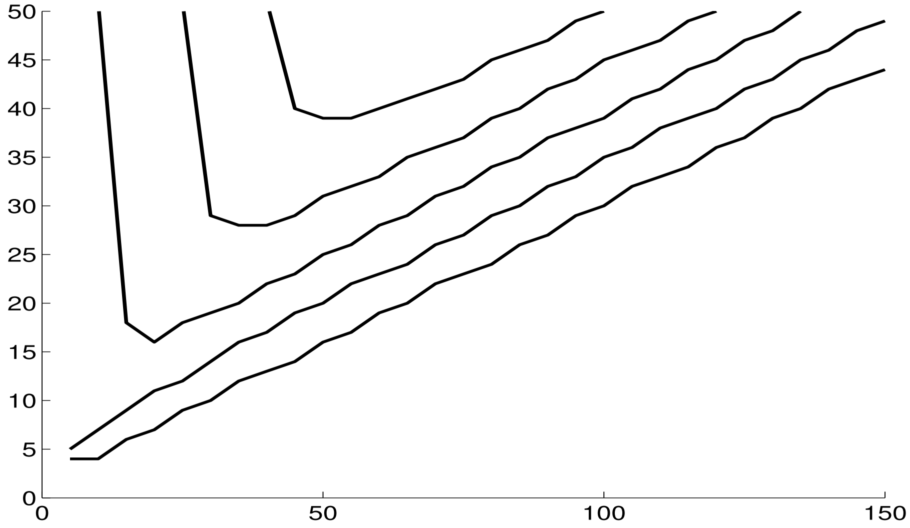

FIG. 5.1. *The norm surface for $A_n = T_n(c_+^{-1})T_n(c_-)$ with $c_+(t) = 1 + 0.16t$ and $c_-(t) = -1.16t^{-1}$. The picture shows the level curves $\|A_n^k\| = c$ for $c = 10^{-2}$, 1, 10^2, 10^4, 10^6 (the lower curve corresponds to $c = 10^{-2}$, the upper to 10^5). We took $3 \le n \le 50$ and $k = 5, 10, 15, \dots, 150$.*

6. Genuinely finite results. Theorem 4.5 is an asymptotic result. In contrast to this, Theorem 4.6 and estimate (4.13) are genuinely finite. However, the lower bound of (4.13) is positive for $n > 41r(k-1)$ only, which is not yet of much use for n's below 1000. The following well-known lemma repeats what was said before (4.13). This lemma is already applicable to $n > 2r(k-1)$.

LEMMA 6.1. *Let $a \in \mathcal{P}_r$ and $n > 2r(k-1)$. Then*

$$T_n^k(a) = T_n(a^k) + P_n X_k P_n + W_n Y_k W_n$$

with $r(k-1) \times r(k-1)$ matrices X_k and Y_k independent of n.

Proof. For $k = 1$, the assertion is true with $X_1 = Y_1 = 0$. Assume the asserted equality is valid for k. We prove it for $k+1$. Obviously,

$$T_n^{k+1}(a) = T_n(a^k)T_n(a) + P_n X_k P_n T(a) P_n + W_n Y_k W_n P_n T(a) P_n.$$

By a famous formula of Harold Widom [34] (see also [11, Prop. 2.12]),

$$T_n(a^k)T_n(a) = T_n(a^{k+1}) - P_n H(a^k) H(\widetilde{a}) P_n - W_n H(\widetilde{a}^k) H(a) W_n,$$

where $H(c)$ is the infinite Hankel matrix $(c_{j+k-1})_{j,k=1}^{\infty}$. We have

$$H(a^k) = \begin{pmatrix} Z_{11} & \dots & Z_{1k} & 0 & \dots \\ \vdots & & \vdots & \vdots & \\ Z_{k1} & \dots & Z_{kk} & 0 & \dots \\ 0 & \dots & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}, \quad H(\widetilde{a}) = \begin{pmatrix} U & 0 & \dots \\ 0 & 0 & \dots \\ \dots & \dots & \dots \end{pmatrix},$$

with certain $r \times r$ blocks Z_{ij} and U, which implies that

$$H(a^k)H(\widetilde{a}) = \begin{pmatrix} Z_{11}U & 0 & \dots \\ \vdots & \vdots & \\ Z_{k1}U & 0 & \dots \\ 0 & 0 & \dots \\ \dots & \dots & \dots \end{pmatrix}.$$

Thus, $P_nH(a^k)H(\widetilde{a})P_n = P_nX'_{k+1}P_n$ with a $kr \times r$ matrix X'_{k+1} independent of n. Analogously, $W_nH(\widetilde{a}^k)H(a)W_n = W_nY'_{k+1}W_n$ with some $kr \times r$ matrix Y'_{k+1}. Further,

$$X_k = \begin{pmatrix} X_{11} & \dots & X_{1,k-1} & 0 & \dots \\ \vdots & & \vdots & \vdots & \\ X_{k-1,1} & \dots & X_{k-1,k-1} & 0 & \dots \\ 0 & \dots & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix},$$

$$T(a) = \begin{pmatrix} A_0 & A_{-1} & & & \\ A_1 & A_0 & A_{-1} & & \\ & A_1 & A_0 & A_{-1} & \\ & & \dots & \dots & \dots \end{pmatrix}$$

with $r \times r$ blocks X_{ij} and A_ℓ independent of n. This shows that at most the first $(k-1)r$ rows of $X_kP_nT(a)$ are nonzero. For $1 \le j \le k-1$ and $\ell \ge k+1$, the j, ℓ block of $X_kP_nT(a)$ is

$$\sum_{m=1}^{k-1} X_{jm}A_{m-\ell} = X_{j,\ell-1}A_{-1} + X_{j,\ell}A_0 + X_{j,\ell+1}A_1,$$

and as $X_{j,\ell-1} = X_{j,\ell} = X_{j,\ell+1} = 0$ for $\ell \ge k+1$, it follows that at most the first kr columns of $X_kP_nT(a)$ are nonzero. Thus, $P_nX_kP_nT(a)P_n = P_nX''_{k+1}P_n$ with some $(k-1)r \times kr$ matrix X''_{k+1} independent of n. Finally,

$$W_nY_kW_nP_nT(a)P_n = W_nY_kP_nT(\widetilde{a})W_n,$$

and the same argument as above shows that this is $W_nY''_{k+1}W_n$ with some $(k-1)r \times kr$ matrix Y''_{k+1} that does not depend on n. $\square$

The $m-1$st Fejér mean $\sigma_m b$ of a function $b \in L^\infty$ is defined by

$$(\sigma_m b)(t) = \sum_{|j| \le m-1} \left(1 - \frac{|j|}{m}\right) b_j t^j \quad (t \in \mathbf{T}).$$

The following result was established in [8]. We cite the proof because it is a pleasure.

THEOREM 6.2. *If $b \in L^\infty$, then $\|T_m(b)\| \ge \|\sigma_m b\|_\infty$ for every $m \ge 1$.*

Proof. Fix $t = e^{i\theta} \in \mathbf{T}$ and let $x_t \in \mathbb{C}^m$ be the vector

$$x_t = \frac{1}{\sqrt{m}}\,(1, \overline{t}, \dots, \overline{t}^{m-1}).$$

ETNA
Kent State University
etna@mcs.kent.edu

We have $\|x_t\| = 1$ and thus $\|T_m(b)\| \ge |(T_m(b)x_t, x_t)|$. Since

$$(T_m(b)x_t, x_t) = \frac{1}{m} \sum_{j,k=0}^{m-1} b_{j-k} e^{-ik\theta} e^{ij\theta} = \frac{1}{m} \sum_{j,k=0}^{m-1} b_{j-k} e^{i(j-k)\theta}$$
$$= \frac{1}{m} \sum_{|j| \le m-1} (m - |j|) b_j e^{ij\theta} = (\sigma_m b)(e^{i\theta}),$$

it follows that $\|T_m(b)\| \ge |(\sigma_m b)(e^{i\theta})| = |(\sigma_m b)(t)|$. As $t \in \mathbf{T}$ can be chosen arbitrarily, we arrive at the assertion. ☐

COROLLARY 6.3. *If $a \in \mathcal{P}_r$ and $n > 2r(k-1)$, then*

$$\|T_n^k(a)\| \ge \|\sigma_{n-2r(k-1)}(a^k)\|_\infty .$$

Proof. By Lemma 6.1,

$$T_n^k(a) = T_n(a^k) + \begin{pmatrix} X_k & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & W_k Y_k W_k \end{pmatrix} = \begin{pmatrix} * & * & * \\ * & A & * \\ * & * & * \end{pmatrix},$$

where X_k and $W_k Y_k W_k$ are $(k-1)r \times (k-1)r$ matrices and $A = T_{n-2r(k-1)}(a^k)$. This implies that

$$\|T_n^k(a)\| \ge \|A\| = \|T_{n-2r(k-1)}(a^k)\|.$$

It remains to combine the last inequality and Theorem 6.2. ☐

We denote by $\|\cdot\|_W$ the Wiener norm of a Laurent polynomial,

$$\|b\|_W = \sum_j |b_j|.$$

THEOREM 6.4. *If $a \in \mathcal{P}_r$ and $n > 2r(k-1)$, then*

$$\|T_n^k(a)\| \ge \|a\|_\infty^k - \frac{kr}{n - 2r(k-1)} \|a\|_W^k .$$

Proof. For $b \in \mathcal{P}_{kr}$,

$$(\sigma_m b)(t) = \sum_{|j| \le kr} \left(1 - \frac{|j|}{m}\right) b_j t^j = \sum_{|j| \le kr} b_j t^j - \frac{1}{m} \sum_{|j| \le kr} |j| b_j t^j,$$

whence

$$|(\sigma_m b)(t)| \ge \left| \sum_{|j| \le kr} b_j t^j \right| - \frac{kr}{m} \sum_{|j| \le kr} |b_j|,$$

and thus

$$\|\sigma_m b\|_\infty \ge \|b\|_\infty - \frac{kr}{m} \|b\|_W .$$

This inequality in conjunction with Corollary 6.3 gives the assertion of the theorem. □

COROLLARY 6.5. *If $a \in \mathcal{P}_r$ and $a_j \geq 0$ for all j, then*

$$\|T_n^k(a)\| \geq \frac{n-3kr}{n-2kr}\,\Big(\sum_j a_j\Big)^k \quad \textit{for} \quad n \geq 3kr.$$

Proof. In this case $\|a\|_\infty = \|a\|_W = \sum_j a_j$ and hence, by Theorem 6.4,

$$\|T_n^k(a)\| \geq \left(1 - \frac{kr}{n-2r(k-1)}\right)\Big(\sum a_j\Big)^k \geq \frac{n-3kr}{n-2kr}\,\Big(\sum a_j\Big)^k. \quad \square$$

EXAMPLE 6.6. Let $a(t) = t + \alpha^2 t^{-1}$ with $\alpha \in \mathbb{R}$. From Corollary 6.5 we infer that

$$\frac{1}{2}\,(1+\alpha^2)^k \leq \|T_{4k}^k(a)\| \leq (1+\alpha^2)^k$$

for all $k \geq 1$. □

EXAMPLE 6.7. Suppose $a \in L^\infty$ and $T(a)$ is lower-triangular. Then $T_n^k(a) = T_n(a^k)$ for all n and k, and hence we can have immediate recourse to Theorem 6.2 to obtain that

$$\|\sigma_n(a^k)\|_\infty \leq \|T_n^k(a)\| \leq \|a\|_\infty^k \tag{6.1}$$

for all n and k.

Let $a(t) = \lambda + t$. Thus, $T_n(a)$ is the Jordan block $J_n(\lambda)$. For $k \leq n$,

$$(\sigma_n(a^k))(t) = \sum_{j=0}^{k}\left(1 - \frac{j}{n}\right)\binom{n}{k}\lambda^{n-j}t^j = (\lambda+t)^k\left(1 - \frac{k}{n}\,\frac{t}{t+\lambda}\right)$$

and hence (6.1) yields

$$\left(1 - \frac{k}{n}\,\frac{1}{1+|\lambda|}\right)(1+|\lambda|)^k \leq \|J_n^k(\lambda)\| \leq (1+|\lambda|)^k \quad (k \leq n).$$

In particular,

$$10^{12.62} \leq \|J_{100}^{50}(0.8)\| \leq 10^{12.77}, \quad 10^{25.17} \leq \|J_{100}^{100}(0.8)\| \leq 10^{25.53},$$

which is better than the results of Section 2.

Now let $a(t) = t + t^2$. Then $T_n(a)$ is a “super Jordan block” [29]. For every k,

$$\|\sigma_{2k}(a^k)\|_\infty = \sum_{j=k}^{2k-1}\left(1 - \frac{j}{2k}\right)\binom{k}{k-j} = \frac{3}{8}\,2^k,$$

and thus (6.1) implies that $(3/8)\,2^k \leq \|T_{2k}^k(a)\| \leq 2^k$. We remark that $T_n(a)$ is triangular, so that genuinely finite results can also be derived from [17]. □

7. The sky region contains an angle. Let $a \in \mathcal{P}_r$ and suppose $\|a\|_\infty > 1$. In Figures 1.1 and 4.1, the sky region looks approximately like an angle: it is bounded by a nearly vertical line on the left and by a curve close to the graph of a linear function $n = ck + d$ from the right and below. The question is whether this remains true beyond the cut-outs we see in the pictures and whether this is valid in general. To be more precise, we fix a (large) number B and we call the set

$$S_B = \{(k,n) : \|T_n^k(a)\| > B\}$$

the *sky region* (for our choice of B). The following theorem proves that the lower-right boundary of the sky region is always linear or sublinear.

THEOREM 7.1. *If $a \in \mathcal{P}_r$ and $\|a\|_\infty > 1$, then there exist positive constants c and k_0, depending on a and B, such that S_B contains the angle $\{(k,n) : k > k_0, n > ck\}$.*

Proof. This is a simple consequence of inequality (4.13), which shows that

$$\|T_n^k(a)\| \ge \|a\|_\infty^k \left(1 - \frac{41rk}{n}\right) \tag{7.1}$$

for all k and n: if $k > k_0$, where $\|a\|_\infty^{k_0} > 2B$ and $n > 82rk$, then the right-hand side of (7.1) is greater than B. □

Thus, if we walk on the norm surface $(k,n) \mapsto \|T_n^k(a)\|$ along a curve whose projection in the k,n plane is given by $n = \varphi(k)$, then we will eventually reach any height B and stay above this height forever provided $\varphi(k)/k \to \infty$. Note that this is satisfied for $\varphi(k) = k \log\log k$. Or in yet other terms, if $\varphi(k)/k \to \infty$, then $\|T_{\varphi(k)}^k(a)\| \to \infty$.

Theorem 7.1 is the deciding argument in support of the statement that $\|T_n^k(a)\|$ runs through a critical transient phase if $\lim_{n\to\infty} \varrho(T_n(a)) < 1$ but $\|a\|_\infty > 1$. Suppose, for example, $k = 100$. If the sky region were roughly of the form $S_B = \{(k,n) : k > k_0, n > k^2\}$, then $\|T_n^{100}(a)\|$ were larger than B for $n > 10000$ only. It is the linearity or sublinearity of the lower-right border of the sky region that allows us to conclude that $\|T_n^{100}(a)\|$ is already larger than B for n in the hundreds.

We call the number

$$L_B(n) = \#\{k : \|T_n^k(a)\| > B\}$$

the *length of the critical transient phase for the matrix dimension n*. Theorem 7.1 implies that $L_B(n) > n/c - k_0$. In other words, $L_B(n)$ increases at least linearly with the matrix dimension n. Equivalently, and a little more elegantly,

$$\liminf_{n\to\infty} \frac{L_B(n)}{n} > 0.$$

May the lower-right boundary of the sky region be strictly sublinear? Or equivalently, may the lowland contain an angle? Let $a \in \mathcal{P}_r$ be given by $a(t) = \sum_{j=-r}^{r} a_j t^j$. For $\mu \in (0,\infty)$, we define $a_\mu \in \mathcal{P}_r$ by

$$a_\mu(t) = \sum_{j=-r}^{r} a_j \mu^j t^j.$$

THEOREM 7.2. *Let $a \in \mathcal{P}_r$, $\|a\|_\infty > 1$, and $B > 1$. If there exists a number $\mu \in (0,\infty)$ such that $\|a_\mu\|_\infty < 1$, then S_B is contained in an angle $\{(k,n) : n > bk + d\}$ with $b > 0$ and $d > 0$.*

Proof. The key observation, due to Schmidt and Spitzer [27], is that

$$T_n(a) = D_\mu^{-1}\, T_n(a_\mu)\, D_\mu,$$

where $D_\mu = \mathrm{diag}\,(1, \mu, \dots, \mu^{n-1})$. Letting $M = \max(\mu, 1/\mu)$, we get

$$\|T_n^k(a)\| \le \|D_\mu^{-1} T_n^k(a_\mu) D_\mu\| \le \|D_\mu^{-1}\|\,\|T_n^k(a)\|\,\|D_\mu\| \le M^{n-1}\|a_\mu\|_\infty^k$$

for all n and k. Thus, if $(k, n) \in S_B$, then $B < M^{n-1}\|a_\mu\|_\infty^k$. Since $M > 1$, it follows that

$$n > k\,\frac{\log(1/\|a_\mu\|_\infty)}{\log M} + \frac{\log B}{\log M} + 1 =: bk + d,$$

and as $\|a_\mu\|_\infty < 1$ and $B > 1$, we see that $b > 0$ and $d > 0$. □

Schmidt and Spitzer [27] showed that always

$$\lim_{n\to\infty} \varrho(T_n(a)) \le \inf_{\mu\in(0,\infty)} \|a_\mu\|_\infty$$

(see also [36]) and that in certain special cases the equality

$$\lim_{n\to\infty} \varrho(T_n(a)) = \inf_{\mu\in(0,\infty)} \|a_\mu\|_\infty \tag{7.2}$$

holds. Equality (7.2) is in particular true if $T(a)$ is Hermitian or tridiagonal or triangular or nonnegative, where nonnegativity means that $a_j \ge 0$ for all j (see also [15] for the nonnegative case). Hermitian matrices are uninteresting in our context, because for them the value given by (7.2) coincides with $\|a\|_\infty$. However, if $T(a)$ is tridiagonal or triangular or nonnegative and if $\lim \varrho(T_n(a)) < 1$, then (7.2) implies that we can find a $\mu \in (0, \infty)$ such that $\|a_\mu\|_\infty < 1$ and hence, by Theorem 7.2, the sky region is contained in an angle. Equivalently, the lowland contains an angle. It also follows that in these cases

$$\limsup_{n\to\infty} \frac{L_B(n)}{n} < \infty.$$

Suppose now that $a \in \mathcal{P}_r$, $\lim \varrho(T_n(a)) < 1$, and $\|a\|_\infty > 1$. If the lower border of the sky region is strictly sublinear, then, by Theorem 7.2, $\inf \|a_\mu\|_\infty$ must be at least 1. I looked for such symbols in $\mathcal{P}_3$ and observed that they are difficult to identify. The determination of $\inf \|a_\mu\|_\infty$ for a given $a \in \mathcal{P}_3$ is simple. The problem comes with checking whether $\lim \varrho(T_n(a))$ is smaller than 1 for a given candidate $a \in \mathcal{P}_3$. It turns out that $\inf \|a_\mu\|_\infty$ and $\lim \varrho(T_n(a))$ are usually extremely close. I took 3000 random symbols $a \in \mathcal{P}_3$ whose real and imaginary parts of the 7 coefficients $a_{-3}, \dots, a_3$ were drawn from the uniform distribution on $(-1, 1)$. Each time Matlab computed

$$q = \inf_{\mu\in(0,\infty)} \|a_\mu\|_\infty / \varrho(T_{64}(a)).$$

The result was as follows:

$$\begin{array}{llll}
1.00 \le q < 1.02 & \text{in} & 2793 \ \text{samples} & (= 93.1\,\%)\\
1.02 \le q < 1.04 & \text{in} & 117 \ \text{samples} & (= 3.9\,\%)\\
1.04 \le q < 1.06 & \text{in} & 48 \ \text{samples} & (= 1.6\,\%)\\
1.06 \le q < 1.08 & \text{in} & 23 \ \text{samples} & (= 0.77\,\%)\\
1.08 \le q < 1.10 & \text{in} & 10 \ \text{samples} & (= 0.33\,\%)\\
1.10 \le q < 1.20 & \text{in} & 9 \ \text{samples} & (= 0.3\,\%)
\end{array}$$

and there was no sample with $q \ge 1.20$. Thus, the dice show that if $\varrho(T_{64}(a)) \le 0.98$, then $\inf \|a_\mu\|_\infty$ is at most $1.02 \cdot 0.98 = 0.9996 < 1$ with probability about 93 % and if $\varrho(T_{64}(a)) \le 0.83$, then $\inf \|a_\mu\|_\infty$ does not exceed $1.20 \cdot 0.83 = 0.996 < 1$ almost surely. This result reveals that if we had a symbol $a \in \mathcal{P}_3$ for which $\lim \varrho(T_n(a)) < 1$ and $\inf \|a_\mu\|_\infty \ge 1$, then $\lim \varrho(T_n(a))$ were dramatically close to 1. Ensuring that $\lim \varrho(T_n(a))$ is really strictly smaller than 1 and guaranteeing at the same time that $\inf \|a_\mu\|_\infty \geq 1$ requires subtle tiny adjustments in the higher decimals after the comma of the coefficients. As I was not sure whether these subtleties survive the numerics needed to plot the norm surface, that is, to compute $\|T_n^k(a)\|$ for n in the 30's and k in the hundreds, I gave up. Thus, I do not know a single symbol with strictly sublinear lower-right border of the sky region or, equivalently, with a lowland that does not contain an angle.

EXAMPLE 7.3. Figure 7.1 shows the norm surface for the symbol $a(t) = t^{-1} + 0.49^2 t$. The spectral radius of $T_n(a)$ is exactly $0.98 \cos \frac{\pi}{n+1}$ and $\|a\|_\infty = 1 + 0.49^2 = 1.2401$. We clearly see the critical transient behavior of $\|T_n^k(a)\|$ for n exceeding 20 or 30. We also nicely see the indents in the level curves. The matrix $T(a)$ is nonnegative, and hence, by Theorem 7.2, the sky region must be contained in an angle. The strange thing with Figure 7.1 is that the lower-right pieces of the level curves nevertheless look slightly sublinear. Let $n = \varphi_B(k)$ be the equation of the lower-right piece of the level curve $\|T_n^k(a)\| = B$. If φ_B were sublinear, then $\varphi_B(k)/k$ would approach zero as $k \to \infty$. The left picture of Figure 7.2, showing $10\,\varphi_B(k)/k$ and $10\,\varphi_B(k) \log k/k$ for $100 \leq k \leq 1000$ does not yet convincingly indicate that $\varphi_B(k)/k$ tends to a positive finite limit. However, the right picture of Figure 7.2, where we extended the range of the k's up to 3000, reveals that there must be a positive finite limit for the lower curve $10\,\varphi_B(k)/k$. $\square$

REMARK 7.4. Theorem 7.1 tells us that the lower-right boundary of the sky region is always sublinear and Theorem 7.2 shows that it is superlinear in many cases. Things are completely different for the $n \times n$ truncations A_n of arbitrary bounded linear operators on ℓ^2.

Let $\psi : \mathbb{N} \to (e, \infty)$ be any function such that $\psi(n) \to \infty$ as $n \to \infty$. We put

$$\xi(n) = \frac{\psi(n)}{\log \psi(n)}, \quad \lambda_n = e^{-1/\xi(n)}.$$

Clearly, $1/e < \lambda_n < 1$. We define the operator A by

$$\begin{aligned} A &= \operatorname{diag}\,(J_2(\lambda_1), J_2(\lambda_2), \ldots) \\ &:= \operatorname{diag}\left(\begin{pmatrix} \lambda_1 & 0 \\ 1 & \lambda_1 \end{pmatrix}, \begin{pmatrix} \lambda_2 & 0 \\ 1 & \lambda_2 \end{pmatrix}, \ldots \right). \end{aligned}$$

Since $\lambda_n < 1$ for all n, the operator A is bounded on ℓ^2. We have

$$A_n^k = \begin{cases} \operatorname{diag}\,(J_2^k(\lambda_1), \ldots, J_2^k(\lambda_m)) & \text{for} \quad n = 2m, \\ \operatorname{diag}\,(J_2^k(\lambda_1), \ldots, J_2^k(\lambda_m), \lambda_{m+1}^k) & \text{for} \quad n = 2m+1. \end{cases}$$

The equality

$$J_2^k(\lambda) = \begin{pmatrix} \lambda^k & 0 \\ k\lambda^{k-1} & \lambda^k \end{pmatrix}$$

ETNA
Kent State University
etna@mcs.kent.edu

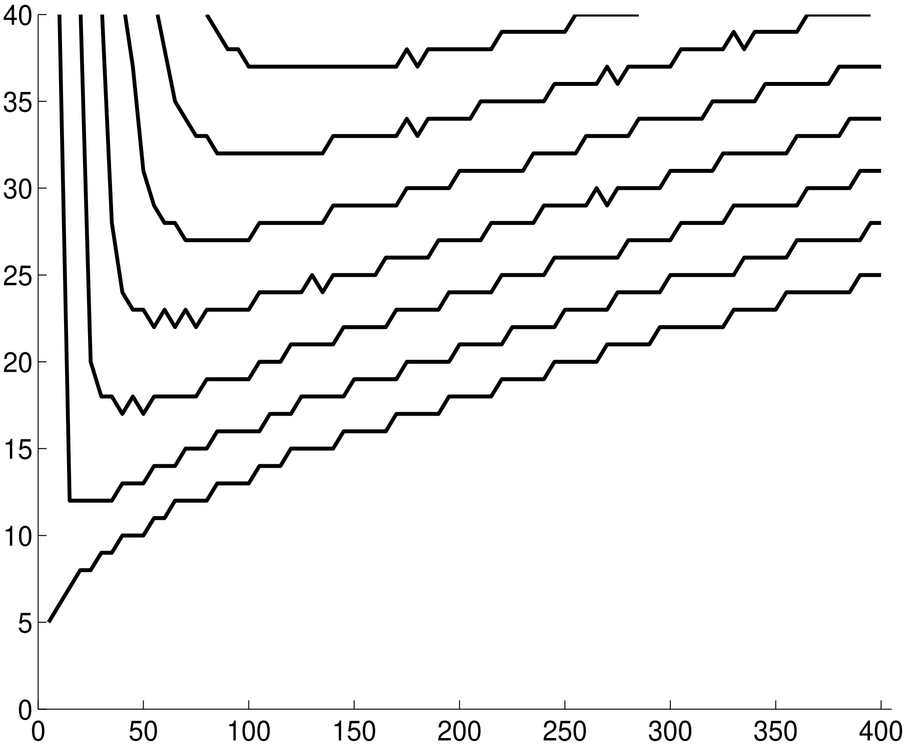

FIG. 7.1. *The symbol is $a(t) = t^{-1} + 0.49^2 t$. The picture shows the level curves $\|T_n^k(a)\| = c$ for $c = 1$, 10, 10^2, 10^3, …, 10^6 (the lower curve corresponds to $c = 1$, the upper to 10^6). We took $3 \le n \le 40$ and* $k = 5, 10, 15, \ldots, 400$.

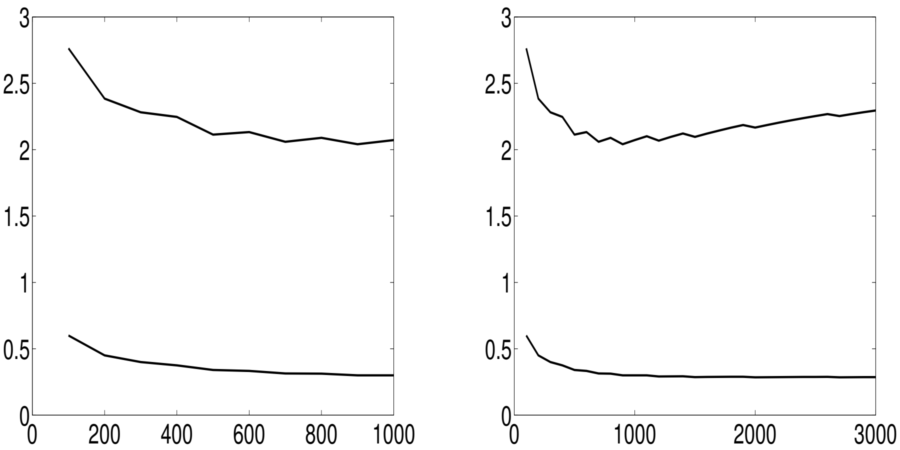

FIG. 7.2. *The symbol is $a(t) = t^{-1} + 0.49^2 t$. The pictures show $10\,\varphi_B(k)/k$ and $10\,\varphi_B(k)\log k/k$ for $B = 10^{-5}$ over two different ranges of k.*

and the condition $1/e < \lambda < 1$ imply that

$$k\lambda^k \le k\lambda^{k-1} \le \|J_2^k(\lambda)\| \le \lambda^k + k\lambda^{k-1} \le (1+ek)\lambda^k.$$

Consequently, up to constants independent of n and k, we may replace $\|A_n^k\|$ by $k\lambda_m^k$ with $m = [n/2]$. The function $f_m(x) := x\lambda_m^x$ has its maximum at $\xi(m)$ and $f_m(\xi(m)) = (1/e)\xi(m)$. If $k \ge \psi(m)$, then $f_m(k) \le f_m(\psi(m))$ because f_m is monotonically decreasing on the right of $\psi(m)$ (note that $\psi(m) \ge \xi(m)$). As $f_m(\psi(m)) = 1$, it follows that

$$S_B \subset \{(k,n) : k < \psi([n/2])\} \tag{7.3}$$

once B is large enough, say $B > 10$. Thus, on choosing very slowly increasing functions ψ, we obtain very narrow sky regions. In particular, the sky regions need not contain any angles.

This is the right point to come back to what was said after Theorem 7.1. For the operator A just constructed, we have

$$\|A_n^k\| \simeq k\lambda_{[n/2]}^k \to k \quad \text{as} \quad n \to \infty.$$

Thus, when moving along the line $k = 10^4$ on the norm surface, we will eventually be at a height of about 10^4 and may conclude that $\|A_n^k\|$ is close to 10^4 for all sufficiently large n. But if, for example, $\psi(n) = \log n$ for large n, then (7.3) shows that we will be in $S_{10^4/2}$ only for the n's satisfying $\log[n/2] > k = 10^4$, that is, for $n > 2\exp(10^4) \approx 2 \cdot 10^{4343}$. We beautifully see that in this case movement along the lines $k =$ constant does practically not provide us with information about the evolution of the norms along the lines $n =$ constant. Moral: Theorem 1.1 is a good reason for expecting critical behavior whenever $\|a\|_\infty > 1$, but *it is Theorem 1.1 in conjunction with Theorem 7.1 that justifies this expectation within reasonable dimensions.*

In the language of critical transient phase lengths, (7.3) says that

$$\limsup_{n\to\infty} \frac{L_B(n)}{\psi([n/2])} < \infty$$

for the operator constructed above. An estimate in the reverse direction is also easy. Namely, it is readily seen that

$$f_m\left(\frac{1}{2}\,\xi(m)\right) = \frac{1}{2\sqrt{e}}\,\xi(m), \qquad f_m(2\,\xi(m)) = \frac{2}{2e^2}\,\xi(m),$$

and since $2/e^2 < 1/(2\sqrt{e})$, this shows that

$$f_m(k) > \frac{2}{e^2}\,\xi(m) \quad \text{for} \quad \frac{1}{2}\,\xi(m) < k < 2\,\xi(m).$$

Consequently, $L_B(n) \ge (3/2)\xi([n/2])$ if only $(2/e^2)\xi([n/2]) > B$. As the last inequality is satisfied for all sufficiently large n, we arrive at the conclusion that

$$\liminf_{n\to\infty} \frac{L_B(n)}{\xi([n/2])} \ge \frac{3}{2}.$$

Finally, since $\xi(m) = \psi(m)/\log\psi(m)$, it follows that

$$\liminf_{n\to\infty} \frac{L_B(n)}{\psi([n/2])}\,\log\psi([n/2]) > 0.$$

Rapidly growing functions ψ, such as $\psi(n) = e^n$, therefore yield gigantic critical phase lengths. $\square$

ETNA
Kent State University
etna@mcs.kent.edu

8. Oscillations. Let $a(t) = t^{-1} + 0.49^2 t$ be as in Example 7.3. While plotting $\|T_{30}^k(a)\|$ for k between 1 and 300, I saw Figure 8.1 on the screen and my first thought was that I had done something wrong. However, after a while I began to understand the reason for the oscillating behavior of the norms and now I am sure that Figure 8.1 is correct.

Let A_n be an $n \times n$ matrix and suppose $\varrho(A_n) < 1$. For the sake of simplicity, assume that all eigenvalues $\lambda_1, \ldots, \lambda_n$ of A_n are simple. We then have $A_n = C\Lambda C^{-1}$, where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and C is an invertible $n \times n$ matrix. It follows that $A_n^k = C\Lambda^k C^{-1}$, and hence

$$A_n^k = C_1\lambda_1^k + \ldots + C_n\lambda_n^k \tag{8.1}$$

for certain $n \times n$ matrices $C_1, \ldots, C_n$ that do not depend on k (note that C_j is simply the product of the jth column of C by the jth row of C^{-1}).

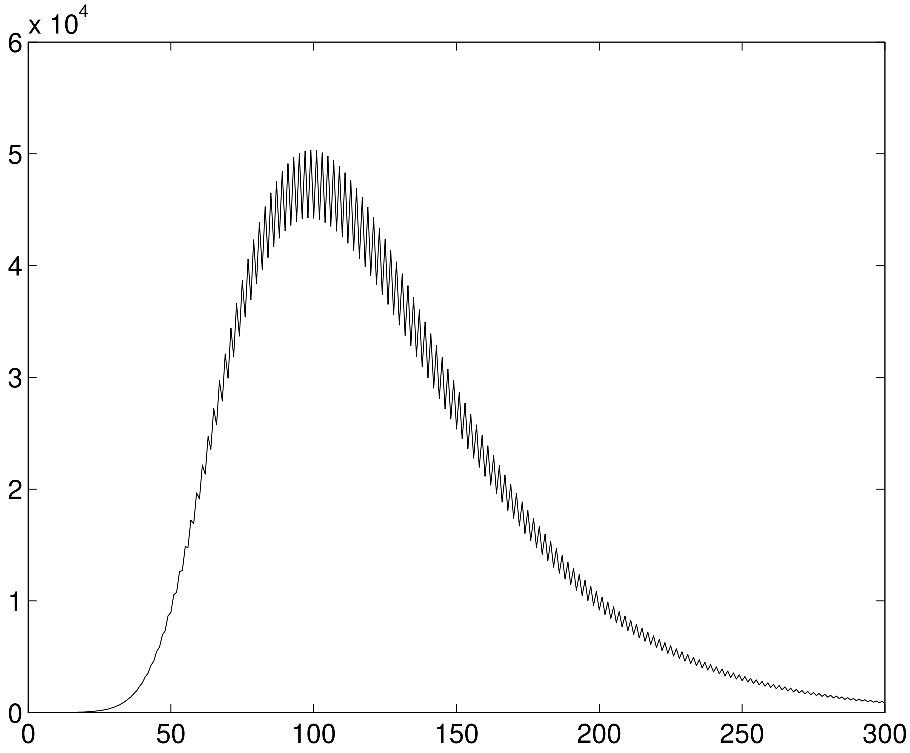

FIG. 8.1. *The symbol is $a(t) = t^{-1} + 0.49^2 t$. We see the norms $\|T_{30}^k(a)\|$ for $k = 1, 2, 3, \ldots, 300$. A close-up is in the left picture of Figure 8.2.*

Assume that

$$|\lambda_1| = \ldots = |\lambda_s| > \max_{j \geq s+1} |\lambda_j|.$$

Then (8.1) gives

$$A_n^k = C_1\lambda_1^k + \ldots + C_s\lambda_s^k + O(\sigma^k)$$

with $\sigma = \max_{j\geq s+1} |\lambda_j|/|\lambda_1| < 1$. Thus, if k is large, then

$$\|A_n^k\| \approx \|C_1\lambda_1^k + \ldots + C_s\lambda_s^k\|, \tag{8.2}$$

which has good chances for oscillatory behavior due to the fact that λ_1 to λ_s have equal moduli.

In the case where $A_n = T_n(a)$ with $a(t) = t^{-1} + \alpha^2 t$, the eigenvalues are given by

$$2\alpha \cos \frac{\pi j}{n+1} \quad (j = 1, \dots, n).$$

The two dominating eigenvalues are

$$\lambda_1 = 2\alpha \cos \frac{\pi}{n+1} \quad \text{and} \quad \lambda_2 = 2\alpha \cos \frac{n\pi}{n+1} = -2\alpha \cos \frac{\pi}{n+1}.$$

Consequently,

$$\|T_n^k(a)\| \approx \left(|2\alpha| \cos \frac{\pi}{n+1}\right)^k \|C_1 + (-1)^k C_2\| \tag{8.3}$$

for large k. Thus, the damping factor $|2\alpha|^k \cos^k \frac{\pi}{n+1}$ has an amplitude that equals $\|C_1 + C_2\|$ for even k and $\|C_1 - C_2\|$ for odd k. If the damping factor is not too small, then (8.3) (which holds for large k only) should be already valid in the critical transient phase, so that we can see it with our eyes. This would be an explanation for the oscillating behavior in Figure 8.1. Let us check our example. Thus, take $\alpha = 0.49$ and $n = 30$. Then

$$\|T_{30}^{100}(a)\| = 10^{4.6457}, \quad \left(0.98 \cos \frac{\pi}{31}\right)^{100} \|C_1 + C_2\| = 10^{5.1438},$$
$$\|T_{30}^{101}(a)\| = 10^{4.7014}, \quad \left(0.98 \cos \frac{\pi}{31}\right)^{101} \|C_1 - C_2\| = 10^{5.1959},$$

that is, (8.3) cannot be said to be satisfied. The point is that the modulus of the quotient of the dominant eigenvalues and the next eigenvalue is very close to 1, which implies that approximation (8.3) is not yet good enough in the high transient phase. However, consideration of a few more terms does set things right:

$$\left(0.98 \cos \frac{\pi}{31}\right)^k \left\| C_1 + (-1)^k C_2 + (C_3 + (-1)^k C_4) \left(\frac{\cos(2\pi/31)}{\cos(\pi/31)}\right)^k \right\|$$

equals $10^{4.4597}$ and $10^{4.5201}$ for $k = 100$ and $k = 101$, respectively, and

$$\left(0.98 \cos \frac{\pi}{31}\right)^k \left\| C_1 + (-1)^k C_2 + (C_3 + (-1)^k C_4) \left(\frac{\cos(2\pi/31)}{\cos(\pi/31)}\right)^k \right.$$
$$\left. + (C_5 + (-1)^k C_6) \left(\frac{\cos(3\pi/31)}{\cos(\pi/31)}\right)^k \right\|$$

is $10^{4.6512}$ and $10^{4.7067}$ for $k = 100$ and $k = 101$, respectively.

From the paper [13] by Brian Davies, I learned that such oscillation phenomena for semigroup norms are well-known. I also learned from [13] that the kind of oscillation may depend on the norm chosen. Figures 8.2 and 8.3 show the different oscillatory behavior of the spectral norms $\|T_n^k(a)\|$ and the Frobenius norms $\|T_n^k(a)\|_F$ fairly convincingly. Finally, Figure 8.4 illustrates what happens when walking on the norm surface along the lines $n =$

ETNA
Kent State University
etna@mcs.kent.edu

constant.

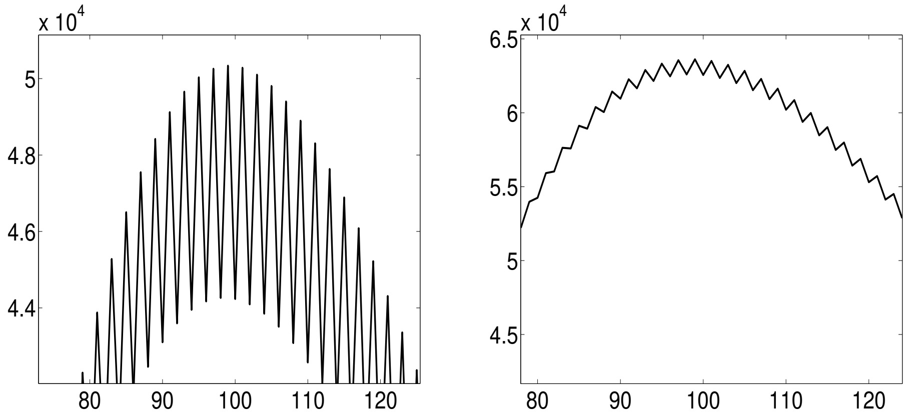

FIG. 8.2. *The symbol is $a(t) = t^{-1} + 0.49^2 t$. The left picture is a close-up of Figure 8.1. The right picture shows the Frobenius norms $\|T_{30}^k(a)\|_{\mathrm{F}}$.*

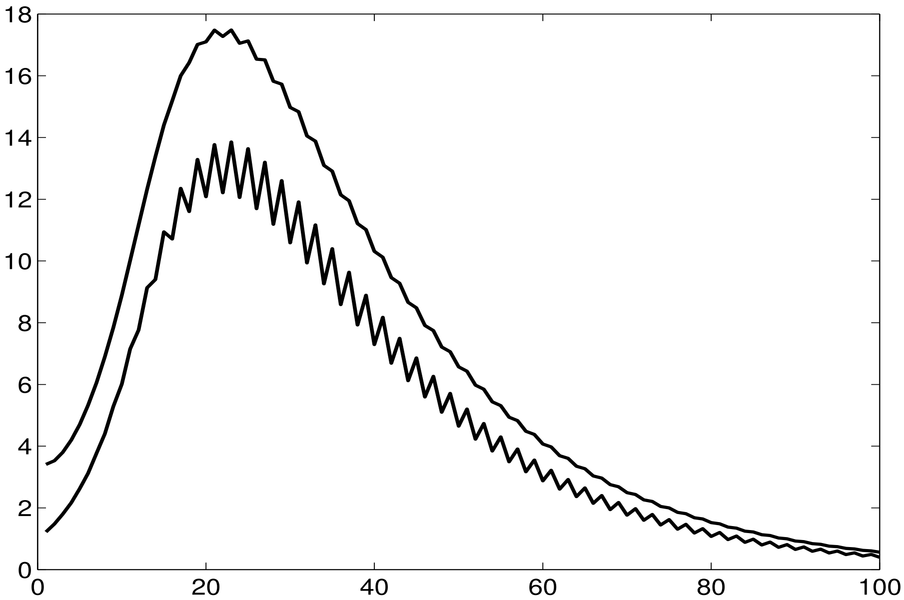

FIG. 8.3. *The symbol is $a(t) = t^{-1} + 0.49^2 t$. We see the norms $\|T_{12}^k(a)\|$ (lower curve) and the Frobenius norms $\|T_{12}^k(a)\|_{\mathrm{F}}$ (upper curve) for $k = 1, 2, 3, \ldots, 100$. The picture reminds me of Steven Spielberg's white shark.*

ETNA
Kent State University
etna@mcs.kent.edu

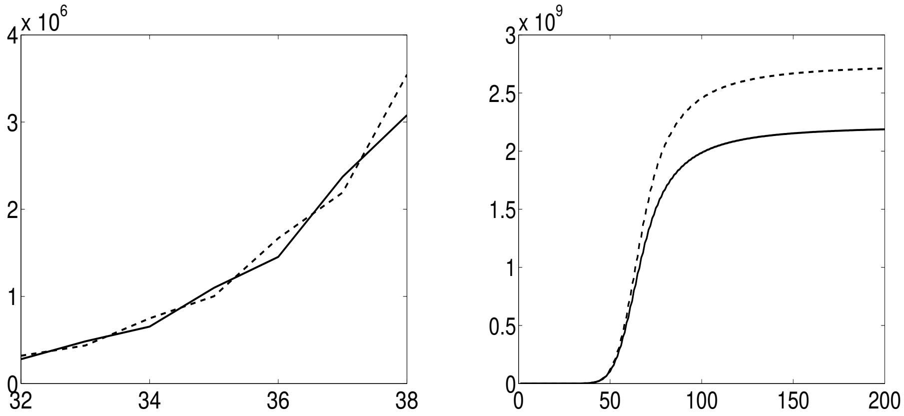

FIG. 8.4. *The symbol is again as in Figure 8.1. The pictures show the norms* $\|T_n^{100}(a)\|$ (*solid*) *and* $\|T_n^{101}(a)\|$ (*dashed*) *for two different ranges of* n.

Another example is considered in Figures 8.5 and 8.6. Now the symbol is $a(t) = \frac{10}{19}(t + t^{-2})$. We have $\varrho(T_{30}(a)) = 0.9847$ and $\|a\|_\infty = 1.04$. The dominating eigenvalues of $T_{30}(a)$ are

$$\lambda_1 = \mu, \;\; \lambda_2 = \mu\omega, \;\; \lambda_3 = \mu\omega^2 \quad (\mu = 0.9847, \;\; \omega = e^{2\pi i/3})$$

and hence

$$\|T_{30}^k(a)\| \approx \begin{cases} \|C_1 + C_2 + C_3\| \, |\mu|^k & \text{for} \quad k \equiv 0 \,(\mathrm{mod}\, 3) \\ \|C_1 + C_2\omega + C_3\omega^2\| \, |\mu|^k & \text{for} \quad k \equiv 1 \,(\mathrm{mod}\, 3) \\ \|C_1 + C_2\omega^2 + C_3\omega\| \, |\mu|^k & \text{for} \quad k \equiv 2 \,(\mathrm{mod}\, 3) \end{cases}$$

The period 3 is nicely seen in the right picture of Figure 8.6.

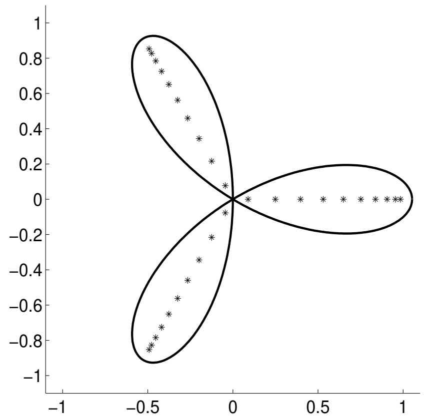

FIG. 8.5. *The range* $a(\mathbf{T})$ *for* $a(t) = \frac{10}{19}(t + t^{-2})$ *and the eigenvalues of* $T_{30}(a)$. *The three dominating eigenvalues are* 0.9847, 0.9847ω, $0.9847\omega^2$ *with* $\omega = e^{2\pi i/3}$. *The maximum modulus of the remaining* 27 *eigenvalues is* 0.9549.

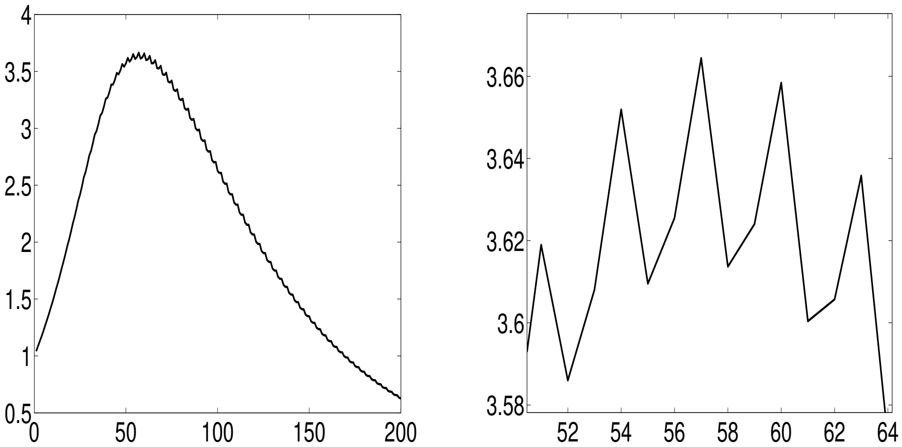

FIG. 8.6. *The pictures show the norms* $\|T_{30}^k(a)\|$ *for the symbol* $a(t) = \frac{10}{19}(t + t^{-2})$.

9. Exponentials. Let A_n be an $n \times n$ matrix and let $\tau > 0$. Then $\|e^{\tau A_n}\| \to 0$ as $\tau \to \infty$ if and only if the spectrum of A_n is contained in the open left half-plane. Now let $A_n = T_n(a)$ with $a \in L^\infty$. The decomposition $a = \operatorname{Re} a + i \operatorname{Im} a$ of a into the real and imaginary parts yields

$$\operatorname{Re} A_n := \frac{1}{2}(A_n + A_n^*) = T_n(\operatorname{Re} a),$$
$$\operatorname{Im} A_n := \frac{1}{2i}(A_n - A_n^*) = T_n(\operatorname{Im} a).$$

We put

$$\sup \operatorname{Re} a := \operatorname*{ess\,sup}_{t \in \mathbf{T}} \operatorname{Re} a(t).$$

Notice that $\sup \operatorname{Re} a = \max \operatorname{Re} a$ if a is continuous. Here is the analogue of Theorem 1.1.

THEOREM 9.1. *Let* $a \in L^\infty$. *Then*

$$\|e^{\tau T_n(a)}\| \le e^{\tau \sup \operatorname{Re} a} \tag{9.1}$$

for all $\tau > 0$ *and all* $n \in \mathbb{N}$. *Moreover,*

$$\lim_{n \to \infty} \|e^{\tau T_n(a)}\| = e^{\tau \sup \operatorname{Re} a} \tag{9.2}$$

for each $\tau > 0$.

Proof. We have $\|e^{A_n}\| \le \|e^{\operatorname{Re} A_n}\|$ for every matrix A_n (see, e.g., [2, p. 258]). Furthermore, $\|e^{\operatorname{Re} A_n}\| = e^{\lambda_{\max}(\operatorname{Re} A_n)}$, where $\lambda_{\max}(\operatorname{Re} A_n)$ is the maximal eigenvalue of the Hermitian matrix $\operatorname{Re} A_n$. Thus,

$$\|e^{\tau T_n(a)}\| \le e^{\tau \lambda_{\max}(T_n(\operatorname{Re} a))}.$$

It is well-known that $\lambda_{\max}(T_n(\operatorname{Re} a))$ does not exceed $\sup \operatorname{Re} a$. The proof is as follows. Pick $\mu > \sup \operatorname{Re} a$ and assume $T_n(\operatorname{Re} a)x = \mu x$ for some nonzero $x = (x_j)_{j=0}^{n-1} \in \mathbb{C}^n$.

Put

$$x(e^{i\theta}) = x_0 + x_1 e^{i\theta} + \ldots + x_{n-1} e^{i(n-1)\theta}.$$

A straightforward computation shows that

$$(T_n(\mu - \operatorname{Re} a)x, x) = \frac{1}{2\pi} \int_0^{2\pi} (\mu - (\operatorname{Re} a)(e^{i\theta}))|x(e^{i\theta})|^2 d\theta, \tag{9.3}$$

and since $T_n(\mu - \operatorname{Re} a)x = 0$, the right-hand side of (9.3) must be zero. As $x(e^{i\theta})$ vanishes only at finitely many $e^{i\theta} \in \mathbf{T}$, it follows that $\mu - \operatorname{Re} a = 0$ almost everywhere, which is impossible for $\mu > \sup \operatorname{Re} a$. Thus, the proof of (9.1) is complete

Since $T_n(a) \to T(a)$ strongly and hence $e^{\tau T_n(a)} \to e^{\tau T(a)}$ strongly, the Banach-Steinhaus theorem gives

$$\liminf_{n\to\infty} \|e^{\tau T_n(a)}\| \geq \|e^{\tau T(a)}\|.$$

The symbol $\operatorname{sym} e^{\tau T(a)}$ is nothing but the function $e^{\tau a}$, and hence we deduce from (4.2) that

$$\|e^{\tau T(a)}\| \geq \|T(e^{\tau a})\| = \|e^{\tau a}\|_\infty.$$

Since $\|e^{\tau a}\|_\infty = e^{\tau \sup \operatorname{Re} a}$, we get

$$\liminf_{n\to\infty} \|e^{\tau T_n(a)}\| \geq e^{\tau \sup \operatorname{Re} a}.$$

This and (9.1) imply (9.2). □

Thus, Theorem 9.1 tells us that if n is large, then $e^{\tau T_n(a)}$ has critical behavior if and only if $\sup \operatorname{Re} a > 0$. To get realistic estimates, one can employ the analogue of (3.5). Trefethen's note [32] contains an analogue of Theorem 3.1 for exponentials. This result implies that if the pseudospectrum $\operatorname{sp}_\varepsilon A$ contains points in the open right half-plane, then each point $\lambda \in \operatorname{sp}_\varepsilon A$ with $\beta := \operatorname{Re} \lambda > 0$ gives us an estimate of the form

$$\sup_{0<\tau\leq\tau_0} \|e^{\tau A}\| \geq e^{\beta\tau_0} \Big/ \left(1 + \varepsilon \frac{e^{\beta\tau_0} - 1}{\beta}\right)$$

(see also [13] and [38]).

The goal of the rest of this section is to establish an analogue of Theorem 7.1. Things are a little more complicated, because I do not know an analogue of (4.13). We begin with an example.

EXAMPLE 9.2. Let $T(a)$ be the tridiagonal matrix generated by $a(t) = t^{-1} + \alpha^2 t - \lambda$ with $\alpha \in [0, \infty)$ and $\lambda \in \mathbb{C}$. The eigenvalues of $T_n(a)$ are densely spread over the interval $(-2\alpha - \lambda, 2\alpha - \lambda)$. Hence $\|e^{\tau T_n(a)}\| \to 0$ as $\tau \to \infty$ for each n if and only if $2\alpha < \operatorname{Re} \lambda$. We have $\max \operatorname{Re} a = -\operatorname{Re} \lambda + 1 + \alpha^2$. Consequently, $\|e^{\tau T_n(a)}\|$ has a critical transient phase before decaying to zero for large n if and only if $2\alpha < \operatorname{Re} \lambda < 1 + \alpha^2$. Figure 9.1 reveals that the norm surfaces of exponentials look much like their counterparts for powers. □

ETNA
Kent State University
etna@mcs.kent.edu

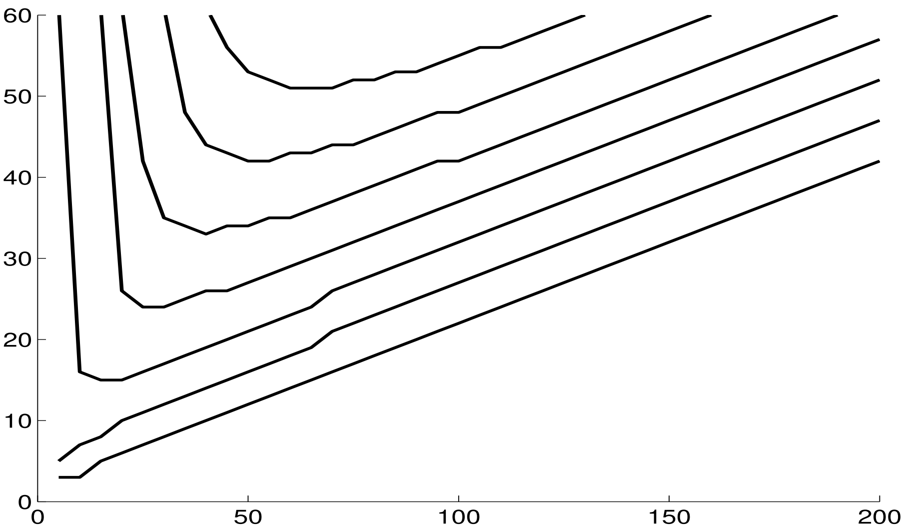

FIG. 9.1. *The symbol is* $a(t) = t^{-1} + 3^2 t - 7$. *The picture shows the level curves* $\|e^{\tau T_n(a)}\| = c$ *for* $c = 10^{-2}, 1, 10^2, 10^4, \ldots, 10^{10}$ *(the lower curve corresponds to* $c = 10^{-2}$*, the upper to* 10^{10}*). We took* $3 \le n \le 60$ *and* $\tau = 5, 10, 15, \ldots, 200$.

LEMMA 9.3. *If* $a \in \mathcal{R}$*, then there are constants* $D, \mu, \gamma \in (0, \infty)$ *depending only on* a *such that*

$$\|e^{\tau T_n(a)}\| \ge \|\sigma_{[n/3]}(e^{\tau a})\|_\infty - De^{\mu\tau - \gamma n/2}$$

for all sufficiently large n.

Proof. Let $a \in \mathcal{R}$ and $\lambda \in \mathbb{C} \setminus \operatorname{sp} T(a)$. In [6], we proved that

$$T_n^{-1}(a - \lambda) = T_n[(a - \lambda)^{-1}] + P_n X_\lambda P_n + W_n Y_\lambda W_n + E_{\lambda,n},$$

where X_λ and Y_λ are compact operators on ℓ^2 whose entries satisfy

$$|(X_\lambda)_{j,\ell}| \le De^{-\gamma(j+\ell)}, \quad |(Y_\lambda)_{j,\ell}| \le De^{-\gamma(j+\ell)},$$

and $E_{\lambda,n}$ is an $n \times n$ matrix such that $\|E_{\lambda,n}\| \le De^{-\gamma n}$. Here D and γ are positive finite constants depending only on a and λ. A check of the computations and estimates of [6] reveals that the constants D and $1/\gamma$ are uniformly bounded on compact subsets of $\mathbb{C} \setminus \operatorname{sp} T(a)$. Consequently, from the representation

$$e^{\tau T_n(a)} - T_n(e^{\tau a}) = \frac{i}{2\pi} \int_\Gamma e^{\tau\lambda} \Big(T_n^{-1}(a - \lambda) - T_n[(a - \lambda)^{-1}] \Big) d\lambda,$$

where Γ is a counter-clockwise oriented smooth curve in $\mathbb{C} \setminus \operatorname{sp} T(a)$ that encircles $\operatorname{sp} T_n(a)$ exactly once, we obtain that

$$\Big[e^{\tau T_n(a)} - T_n(e^{\tau a}) \Big]_{j,\ell} \le D\, e^{\mu\tau} \left(e^{-\gamma(j+\ell)} + e^{-\gamma(2n-j-\ell)} + e^{-\gamma n} \right)$$

with some constant $\mu \in (0, \infty)$ depending only on a and Γ. For $n/3 \le j, \ell \le 2n/3$, this gives

$$\Big[e^{\tau T_n(a)} - T_n(e^{\tau a}) \Big]_{j,\ell} \le D\, e^{\mu\tau} \left(e^{-2n\gamma/3} + e^{-2n\gamma/3} + e^{-\gamma n} \right) \le 3De^{\mu\tau - 2n\gamma/3}.$$

It follows that the Frobenius norm of the block of $e^{\tau T_n(a)} - T_n(e^{\tau a})$ that is formed by the entries with $n/3 \le j, \ell \le 2n/3$ is at most $3nDe^{\mu\tau - 2n\gamma/3}$, which does not exceed $De^{\mu\tau - n\gamma/2}$ if n is large enough. Theorem 6.2 and the argument of the proof of Corollary 6.3 therefore yield

$$\|e^{\tau T_n(a)}\| \ge \|\sigma_{[n-2n/3]}(e^{\tau a})\|_\infty - De^{\mu\tau - \gamma n/2} = \|\sigma_{[n/3]}(e^{\tau a})\|_\infty - De^{\mu\tau - \gamma n/2},$$

which completes the proof. $\square$

LEMMA 9.4. *Let $a \in \mathcal{R}$ and $\tau \ge 1$. Then for each $\varepsilon > 0$ there exists a constant E depending only on a and ε such that*

$$\|\sigma_m(e^{\tau a})\|_\infty \ge \|e^{\tau a}\|_\infty \left(1 - \frac{E\tau(\log m)^{1+\varepsilon}}{m}\right)$$

for all $m \ge 1$.

Proof. We may, without loss of generality, assume that $\|e^{\tau a}\|_\infty = |e^{\tau a(1)}|$. We put $b(\theta) := a(e^{i\theta})$ for $\theta \in \mathbb{R}$. The function b is smooth and we have $b(0) - a(1) = 0$. Hence, there is a constant F depending only on a such that $|b(\theta) - a(1)| \le F|\theta|$ for $|\theta| \le \pi$. Let s be the maximum of $|b(\theta) - a(1)|$ for $|\theta| \le \pi$. If $|z| \le s$, then

$$\begin{aligned} |1 - e^z| &= |z| \left|1 + \frac{z}{2!} + \frac{z^2}{3!} + \dots\right| \\ &\le |z| \left(1 + \frac{s}{2!} + \frac{s^2}{3!} + \dots\right) = |z| \, \frac{e^s - 1}{s} =: M_s|z|. \end{aligned}$$

Thus,

$$\left|1 - e^{\tau(b(\theta) - a(1))}\right| \le \tau M_s |b(\theta) - a(1)| \le \tau F M_s |\theta| \tag{9.4}$$

for $|\theta| \le \pi$. With the Fejér kernel

$$K_m(\theta) = \frac{1}{2\pi m} \, \frac{\sin^2(m\theta/2)}{\sin^2(\theta/2)},$$

we have

$$\left(\sigma_m\left(\frac{e^{\tau a}}{e^{\tau a(1)}}\right)\right)(1) = 1 - \int_{|\theta|<\pi} K_m(\theta)\left(1 - \frac{e^{\tau b(\theta)}}{e^{\tau a(1)}}\right) d\theta. \tag{9.5}$$

Put $\delta_m = 1/(\log m)^{1+\varepsilon}$. We split the integral $\int_{|\theta|<\pi}$ on the right of (9.5) into two integrals $I_1 = \int_{|\theta|<\delta_m}$ and $I_2 = \int_{|\theta|>\delta_m}$. Using (9.4), we can estimate the first integral as follows:

$$\begin{aligned} |I_1| &\le \frac{1}{2\pi m} \int_{-\delta_m}^{\delta_m} \frac{\sin^2(m\theta/2)}{\sin^2(\theta/2)} \left|1 - e^{\tau(b(\theta) - a(1))}\right| d\theta \\ &\le \frac{1}{2\pi m} \int_{-\delta_m}^{\delta_m} \frac{\sin^2(m\theta/2)}{\sin^2(\theta/2)} \, \tau M_s F |\theta| \, d\theta \\ &= \frac{\tau M_s F}{\pi m} \int_0^{\delta_m} \frac{\sin^2(m\theta/2)}{\sin^2(\theta/2)} \, \theta \, d\theta \\ &\le \frac{\tau M_s F}{\pi m} \int_0^{\delta_m} \frac{\sin^2(m\theta/2)}{(\theta/\pi)^2} \, \theta \, d\theta \end{aligned}$$

ETNA
Kent State University
etna@mcs.kent.edu

$$
\begin{aligned}
&= \frac{\tau\pi M_s F}{m} \int_0^{\delta_m} \frac{\sin^2(m\theta/2)}{\theta}\, d\theta \\
&= \frac{\tau\pi M_s F}{m} \int_0^{m\delta_m/2} \frac{\sin^2 x}{x}\, dx \\
&\le \frac{\tau\pi M_s F}{m} \left(\int_0^{1} \frac{\sin^2 x}{x}\, dx + \int_1^{m\delta_m/2} \frac{dx}{x} \right) \\
&\le \frac{\tau\pi M_s F}{m} \left(\frac{1}{2} + \log \frac{m\delta_m}{2} \right) \\
&= \frac{\tau\pi M_s F}{m} \left(\frac{1}{2} + \log m - (1+\varepsilon)\log\log m - \log 2 \right) \\
&\le \frac{G\tau \log m}{m}
\end{aligned}
$$

for some constant G depending only on a and ε. Since $|e^{\tau b(\theta)}/e^{\tau a(1)}| \le 1$, the modulus of the second integral is

$$
\begin{aligned}
|I_2| &\le \frac{1}{2\pi m} \int_{|\theta|>\delta_m} \frac{\sin^2(m\theta/2)}{\sin^2(\theta/2)} \left| 1 - e^{\tau(b(\theta)-a(1))} \right| d\theta \\
&\le \frac{2}{2\pi m} \int_{|\theta|>\delta_m} \frac{\sin^2(m\theta/2)}{\sin^2(\theta/2)}\, d\theta \\
&= \frac{2}{\pi m} \int_{\delta_m}^{\pi} \frac{\sin^2(m\theta/2)}{\sin^2(\theta/2)}\, d\theta \\
&\le \frac{2}{\pi m} \int_{\delta_m}^{\pi} \frac{\sin^2(m\theta/2)}{(\theta/\pi)^2}\, d\theta \\
&= \frac{2\pi}{m} \int_{\delta_m}^{\pi} \frac{\sin^2(m\theta/2)}{\theta^2}\, d\theta \\
&= \pi \int_{m\delta_m/2}^{m\pi/2} \frac{\sin^2 x}{x^2}\, dx \\
&\le \pi \int_{m\delta_m/2}^{m\pi/2} \frac{dx}{x^2} = \pi \left(\frac{2}{m\delta_m} - \frac{2}{m\pi} \right) < \frac{2\pi}{m\delta_m} \\
&= \frac{2\pi}{m} (\log m)^{1+\varepsilon}.
\end{aligned}
$$

In summary,

$$
\begin{aligned}
&\frac{\|\sigma_m(e^{\tau a})\|_\infty}{\|e^{\tau a}\|_\infty} \ge \frac{|(\sigma_m(e^{\tau a}))(1)|}{|e^{\tau a(1)}|} = \left| \left(\sigma_m \left(\frac{e^{\tau a}}{e^{\tau a(1)}} \right) \right) (1) \right| \\
&= |1 - I_1 - I_2| \ge 1 - |I_1| - |I_2| \\
&\ge 1 - \frac{G\tau \log m}{m} - \frac{2\pi (\log m)^{1+\varepsilon}}{m} \ge 1 - \frac{E\tau (\log m)^{1+\varepsilon}}{m}
\end{aligned}
$$

with $E = G + 2\pi$ (recall that $\tau \ge 1$). □

The following result is an analogue of Theorem 7.1. Clearly, this result is weaker than Theorem 7.1, because of the presence of the factor $(\log \tau)^{1+\varepsilon}$. From a practical point of view we can say that, for large τ, this factor may be ignored in comparison with τ. As for the theoretical side of the problem, we remark that this factor emerges from our techniques and that it can probably be removed by more powerful machinery.

THEOREM 9.5. *Let* $a \in \mathcal{R}$ *and suppose* $\max \operatorname{Re} a > 0$. *For* $B > 0$, *put*

$$S_B = \{(\tau, n) \in (0, \infty) \times \mathbb{N} : \|e^{\tau T_n(a)}\| > B\}.$$

Then, for each $\varepsilon > 0$, *there exist positive and finite constants* τ_0 *and* c *depending only on* B, a, ε *such that*

$$S_B \supset \left\{(\tau, n) : \quad \tau > \tau_0, \quad n > c\tau(\log \tau)^{1+\varepsilon}\right\}.$$

Proof. We abbreviate $\max \operatorname{Re} a$ to β. Combining Lemmas 9.3 and 9.4 we get

$$\|e^{\tau T_n(a)}\| \geq \left(1 - \frac{E\tau(\log[n/3])^{1+\varepsilon}}{[n/3]}\right) e^{\tau\beta} - De^{\mu\tau - \gamma n/2}$$
$$= \left(1 - \frac{E\tau(\log[n/3])^{1+\varepsilon}}{[n/3]} - De^{(\mu-\beta)\tau - \gamma n/2}\right) e^{\tau\beta}$$

for all sufficiently large n and all $\tau > 1$. Throughout what follows we assume that $\tau > e$, so that $\log \tau$ and $\log \log \tau$ are positive. Put

$$c = \max\left(1 + \frac{2}{\gamma}(\mu - \beta), 36E\right) \tag{9.6}$$

and let $n > c\tau(\log \tau)^{1+\varepsilon}$. There is a τ_1 such that $e^{\tau_1\beta} > 3B$. If $\mu \leq \beta$, then $De^{(\mu-\beta)\tau-\gamma n/2} < 1/3$ for all $\tau > \tau_2$. So assume that $\mu > \beta$. We then have

$$De^{(\mu-\beta)\tau-\gamma n/2} < De^{(\mu-\beta)\tau-\gamma c\tau(\log \tau)^{1+\varepsilon}/2} < e^{(\mu-\beta-\gamma c/2)\tau(\log \tau)^{1+\varepsilon}}, \tag{9.7}$$

and because $\mu - \beta - \gamma c/2 < 0$ by virtue of (9.6), the last term of (9.7) is certainly smaller than $1/3$ for all $\tau > \tau_2$. Clearly,

$$\frac{E\tau(\log[n/3])^{1+\varepsilon}}{[n/3]} \leq \frac{2E\tau(\log(n/3))^{1+\varepsilon}}{n/3} \tag{9.8}$$

for all n large enough. As the function on the right of (9.8) is monotonically decreasing for all sufficiently large n, it follows that

$$\frac{E\tau(\log[n/3])^{1+\varepsilon}}{[n/3]} \leq \frac{2E\tau(\log((c/3)\tau(\log \tau)^{1+\varepsilon}))^{1+\varepsilon}}{(c/3)\tau(\log \tau)^{1+\varepsilon}} \tag{9.9}$$
$$= \frac{6E}{c}\left(\frac{\log(c/3) + \log \tau + (1+\varepsilon)\log\log \tau}{\log \tau}\right)^{1+\varepsilon}.$$

Since the right-hand side of (9.9) converges to $6E/c$ as $\tau \to \infty$ and since $6E/c \leq 1/6$ due to (9.6), we arrive at the conclusion that (9.9) is smaller than $1/3$ for all $\tau > \tau_3$. Thus, if $\tau_0 = \max(\tau_1, \tau_2, \tau_3)$, then

$$\|e^{\tau T_n(a)}\| > \left(1 - \frac{1}{3} - \frac{1}{3}\right) 3B = B$$

for $\tau > \tau_0$ and $n > c\tau(\log \tau)^{1+\varepsilon}$. ∎

Theorem 9.5 implies that if $a \in \mathcal{R}$ and $\max \operatorname{Re} a > 0$, then $\|e^{\tau T_{\varphi(\tau)}(a)}\| \to \infty$ as soon as

$$\frac{\varphi(\tau)}{\tau(\log \tau)^{1+\varepsilon}} \to \infty \quad \text{for some} \quad \varepsilon > 0.$$

It also follows that the length of the critical phase satisfies

$$\liminf_{n\to\infty} \frac{L_B(n)}{n} (\log n)^{1+\varepsilon} > 0 \quad \text{for each} \quad \varepsilon > 0.$$

In analogy to Theorem 7.2 we have the following.

THEOREM 9.6. *Let $a \in \mathcal{P}_r$, $\max \operatorname{Re} a > 0$, and $B > 1$. If there exists a number $\nu \in (0, \infty)$ such that $\max \operatorname{Re} a_\nu < 0$, then S_B is a subset of $\{(\tau, n) : n > b\tau + d\}$ for certain $b > 0$ and $d > 0$.*

Proof. We have $T_n(a) = D_\nu^{-1} T_n(a_\nu) D_\nu$ and hence

$$\begin{aligned} \|e^{\tau T_n(a)}\| &= \|D_\nu^{-1} e^{\tau T_n(a_\nu)} D_\nu\| \\ &\leq \|D_\nu^{-1}\| \, \|e^{\tau T_n(a_\nu)}\| \, \|D_\nu\| \\ &\leq \|D_\nu^{-1}\| \, \|D_\nu\| \, e^{\tau \max \operatorname{Re} a_\nu} \end{aligned}$$

(recall (9.1)). Thus, with $M := \max(\nu, 1/\nu) > 1$,

$$\|e^{\tau T_n(a)}\| \leq M^{n-1} e^{\tau \max \operatorname{Re} a_\nu}.$$

For $(\tau, n) \in S_B$, this yields

$$n > \tau \, |\max \operatorname{Re} a_\nu| + \frac{\log B}{\log M} + 1. \quad \square$$

Let $a(t) = t^{-1} + \alpha^2 t - \lambda$ be as in Example 9.2 and suppose that $2\alpha < \operatorname{Re} \lambda < 1 + \alpha^2$. Choosing $\nu = 1/\alpha$ we get $\max \operatorname{Re} a_\nu = \alpha + \alpha - \operatorname{Re} \lambda = 2\alpha - \operatorname{Re} \lambda < 0$, and hence the sky region is contained in an angle. This is nicely seen in Figure 9.1.

Acknowledgement. I thank Anne Greenbaum and Nick Trefethen for valuable remarks on an earlier version of the paper.

REFERENCES

[1] S. AXLER, *Paul Halmos and Toeplitz operators*, in Paul Halmos Celebrating 50 Years of Mathematics, J. H. Ewing and F. W. Gehring, eds., Springer-Verlag, New York, 1991, pp. 257–263.

[2] R. BHATIA, *Matrix Analysis*, Springer-Verlag, New York, 1997.

[3] A. BÖTTCHER, *Pseudospectra and singular values of large convolution operators*, J. Integral Equations Appl., 6 (1994), pp. 267–301.

[4] A. BÖTTCHER, M. EMBREE, AND L. N. TREFETHEN, *Piecewise continuous Toeplitz matrices and operators: slow approach to infinity*, SIAM J. Matrix Analysis Appl., 24 (2002), pp. 484–489.

[5] A. BÖTTCHER AND S. GRUDSKY, *Toeplitz Matrices, Asymptotic Linear Algebra, and Functional Analysis*, Hindustan Book Agency, New Delhi, 2000 and Birkhäuser Verlag, Basel, 2000.

[6] A. BÖTTCHER AND S. GRUDSKY, *Condition numbers of large Toeplitz-like matrices*, Contemp. Math., 280 (2001), pp. 273–299.

[7] A. BÖTTCHER AND S. GRUDSKY, *Toeplitz matrices with slowly growing pseudospectra*, in Factorization, Singular Operators and Related Problems, S. Samko, A. Lebre, A. F. dos Santos, eds., Kluwer, Dordrecht, 2003, pp. 43–54.

[8] A. BÖTTCHER AND S. GRUDSKY, *Fejér means and norms of large Toeplitz matrices*, Acta Sci. Math. (Szeged), 69 (2003), pp. 889–990.

[9] A. Böttcher, S. Grudsky, A. Kozak, and B. Silbermann, *Norms of large Toeplitz band matrices*, SIAM J. Matrix Anal. Appl., 21 (1999), pp. 547–561.

[10] A. Böttcher and B. Silbermann, *Analysis of Toeplitz Operators*, Akademie-Verlag, Berlin, 1989 and Springer-Verlag, Berlin, 1990.

[11] A. Böttcher and B. Silbermann, *Introduction to Large Truncated Toeplitz Matrices*, Universitext, Springer-Verlag, New York, 1999.

[12] E. B. Davies, *Pseudospectra of differential operators*, J. Operator Theory, 43 (2000), pp. 243–262.

[13] E. B. Davies, *Semigroup growth bounds*, Los Alamos archive, arXiv:math.SP/0302144 v1, 12 February 2003.

[14] K. M. Day, *Measures associated with Toeplitz matrices generated by the Laurent expansion of rational functions*, Trans. Amer. Math. Soc., 209 (1975), pp. 175–183.

[15] L. Elsner and S. Friedland, *The limit of the spectral radius of block Toeplitz matrices with nonnegative entries*, Integral Equations Operator Theory, 36 (2000), pp. 193–200.

[16] M. Embree and L. N. Trefethen, *Pseudospectra Gateway*. Available on line from http://www.comlab.ox.ac.uk/pseudospectra.

[17] V. Faber, A. Greenbaum, and D. E. Marshall, *The polynomial numerical hulls of Jordan blocks and related matrices*, Linear Algebra Appl., 374 (2003), pp. 231–246.

[18] S. Frankel, *Convergence rates of iterative treatments of partial differential equations*, Math. Comp., 4 (1950), pp. 65–75.

[19] G. M. Goluzin, *Some estimates for bounded functions*, Matem. Sbornik, N. S. 26 (68) (1950), pp. 7–18 [Russian].

[20] G. M. Goluzin, *Geometric Theory of Functions of a Complex Variable*, Amer. Math. Soc., Providence, RI, 1969.

[21] A. Greenbaum, *Generalizations of the field of values useful in the study of polynomial functions of a matrix*, Linear Algebra Appl., 347 (2002), pp. 233–249.

[22] A. Greenbaum, *Card shuffling and the polynomial numerical hull of degree k*, SIAM J. Sci. Comput., to appear.

[23] A. Greenbaum, personal communication, June 2003.

[24] O. Nevanlinna, *Convergence of Iterations for Linear Equations*, Birkhäuser, Basel, 1993.

[25] O. Nevanlinna, *Hessenberg matrices in Krylov subspaces and the computation of the spectrum*, Numer. Funct. Anal. Optim., 16 (1995), pp. 443–473.

[26] L. Reichel and L. N. Trefethen, *Eigenvalues and pseudo-eigenvalues of Toeplitz matrices*, Linear Algebra Appl., 162/164 (1992), pp. 153–185.

[27] P. Schmidt and F. Spitzer, *The Toeplitz matrices of an arbitrary Laurent polynomial*, Math. Scand., 8 (1960), pp. 15–38.

[28] I. Schur and G. Szegö, *Über die Abschnitte einer im Einheitskreise beschränkten Potenzreihe*, Sitzungsberichte Preuss. Akad. Wiss. Berlin (1925), pp. 545–560.

[29] L. N. Trefethen, *Pseudospectra of matrices*, in Numerical Analysis 1991, D. F. Griffiths and G. A. Watson, eds., Longman Sci. Tech, Harlow, Essex, UK, 1992, pp. 234–266,

[30] L. N. Trefethen, *Pseudospectra of linear operators*, SIAM Review, 39 (1997), pp. 383–406.

[31] L. N. Trefethen, *Computation of pseudospectra*, Acta Numerica, 8 (1999), pp. 247–295.

[32] L. N. Trefethen, personal communication, February 2003.

[33] E. Wegert and L. N. Trefethen, *From the Buffon needle problem to the Kreiss matrix theorem*, Amer. Math. Monthly, 101 (1994), pp. 132–139.

[34] H. Widom, *Asymptotic behavior of block Toeplitz matrices and determinants, II*, Adv. in Math., 21 (1976), pp. 1–19.

[35] H. Widom, *Eigenvalue distribution of nonselfadjoint Toeplitz matrices and the asymptotics of Toeplitz determinants in the case of nonvanishing index*, Operator Theory: Adv. and Appl., 48 (1990), pp. 387–421.

[36] H. Widom, *Eigenvalue distribution for nonselfadjoint Toeplitz matrices*, Operator Theory: Adv. and Appl., 71 (1994), pp. 1–8.

[37] T. Wright, *EigTool software package*. Available on line from http://www.comlab.ox.ac.uk/pseudospectra/eigtool.

[38] T. Wright, *Algorithms and Software for Pseudospectra*, Thesis, University of Oxford, UK, 2002.