Taylor's Theorem for Matrix Functions with Applications to Condition Number Estimation

Deadman, Edvin and Relton, Samuel

2015

ISSN 1749-9097

Taylor's Theorem for Matrix Functions with Applications to Condition Number Estimation

Edvin Deadman[a,1,2], Samuel D. Relton[a,1,2]

[a] *School of Mathematics, The University of Manchester, Manchester, M13 9PL, UK*

Abstract

We derive an explicit formula for the remainder term of a Taylor polynomial of a matrix function. This formula generalizes a known result for the remainder of the Taylor series for an analytic function of a complex scalar. We investigate some consequences of this result, which culminate in new upper bounds for the level-1 and level-2 condition numbers of a matrix function in terms of the pseudospectrum of the matrix. Numerical experiments show that, although the bounds can be pessimistic, they can be computed almost three orders of magnitude faster than the standard methods for the 1-norm condition number of $f(A) = A^t$. This makes the upper bounds ideal for a quick estimation of the condition number whilst a more accurate (and expensive) method can be used if further accuracy is required.

Keywords: matrix function, Taylor series, remainder, condition number, pseudospectrum, Fréchet derivative, Kronecker form
2000 MSC: 15A12, 15A16

1. Introduction

Taylor's theorem is a standard result in elementary calculus (see e.g. [16]). If $f : \mathbb{R} \to \mathbb{R}$ is k times continuously differentiable at $a \in \mathbb{R}$, then the theorem states that there exists $R_k : \mathbb{R} \to \mathbb{R}$ such that

$$f(x) = \sum_{j=0}^{k} \frac{f^{(j)}(a)}{j!}(x-a)^j + R_k(x)$$

and $R_k(x) = o(|x-a|^k)$ as $x \to a$. Depending on any additional assumptions on f, various precise formulae for the remainder term $R_k(x)$ are available. For

[1] This work was supported by European Research Council Advanced Grant MATFUN (267526).

[2] *Email addresses:* `edvin.deadman@manchester.ac.uk` (E. Deadman), `samuel.relton@manchester.ac.uk` (S. D. Relton)

example, if f is $k+1$ times continuously differentiable on the closed interval between a and x, then

$$R_k(x) = \frac{f^{(k+1)}(c)}{(k+1)!}(x-a)^{k+1} \tag{1}$$

for some c between a and x. This is known as the Lagrange form of the remainder. Alternative expressions, such as the Cauchy form or the integral form for the remainder are well known [16].

Taylor's theorem generalizes to analytic functions in the complex plane: instead of (1) the remainder is now expressed in terms of a contour integral. If $f(z)$ is complex analytic in an open subset $\mathcal{D} \subset \mathbb{C}$ of the complex plane, the kth-degree Taylor polynomial of f at $a \in \mathcal{D}$ satisfies

$$f(z) = \sum_{j=0}^{k} \frac{f^{(k)}(a)}{k!}(z-a)^j + R_k(z),$$

where

$$R_k(z) = \frac{(z-a)^{k+1}}{2\pi i} \int_\gamma \frac{f(w)dw}{(w-a)^{k+1}(w-z)}, \tag{2}$$

and γ is a closed curve defining a region $\mathcal{W} \subset \mathcal{D}$ containing a. See [1, §3.1] for a proof of this result.

The first goal of this paper is to generalize (2) to matrices, thereby providing an explicit expression for the remainder term for the kth-degree Taylor polynomial of a matrix function. Note that it will not be possible to obtain an expression similar to (1) because its derivation relies on the mean value theorem, which does not have an exact analogue for matrix-valued functions. Our second goal is to investigate applications of this result to pseudospectra and condition numbers. In particular we show how upper bounds on the condition number of the matrix function A^t, for $t \in (0,1)$, can be estimated very efficiently using a pseudospectral bound derived from the remainder term of a Taylor expansion of the function $f(z) = z^t$ about the matrix A. The bound offers substantial speedups over existing methods to estimate the condition number, though the bound can be much looser.

Convergence results for Taylor series of matrix functions have been known since the work of Hensel [8] and Weyr [18] (see [9, Thm. 4.7] for a more recent exposition). Mathias [15] also obtains a normwise truncation error bound for matrix function Taylor polynomials, which form part of the Schur–Parlett algorithm [5]. However, to our knowledge, this paper represents the first time an explicit remainder term (as opposed to a bound) has been obtained for the Taylor polynomial of a matrix function.

The remaining sections of this paper are organized as follows. In section 2 we state and prove the remainder term for the kth-degree Taylor polynomial of a matrix function. In section 3 we investigate some applications of this result by bounding the first order remainder term using pseudospectral techniques and relating it to the condition number of $f(A)$. In section 4 we extend these results to the level-2 condition number of a matrix function, introduced in [11]. In section 5 we examine the behaviour of the pseudospectral bounds on some test problems. Finally in section 6 we present our conclusions and discuss some potential extensions of this work.

2. Remainder term for Taylor polynomials

The Taylor series theorems found in Higham's monograph [9] primarily involve expanding $f(A)$ about a multiple of the identity matrix, I:

$$f(A) = \sum_{j=0}^{\infty} \frac{f^{(j)}(\alpha)}{j!}(A - \alpha I)^j.$$

Our starting point is the more general Taylor series expansion in terms of Fréchet derivatives, obtained by Al-Mohy and Higham [2, Thm. 1]. Suppose that f is analytic in an open subset $\mathcal{D} \subset \mathbb{C}$ of the complex plane. Then, given $A, E \in \mathbb{C}^{n\times n}$ with $\Lambda(A+E) \subset \mathcal{D}$ (where $\Lambda(X)$ denotes the spectrum of the matrix X), Al-Mohy and Higham proved that

$$f(A+E) = \sum_{j=0}^{\infty} \frac{1}{j!} D_f^{[j]}(A,E), \tag{3}$$

where

$$D_f^{[j]}(A,E) = \frac{d^j}{dt^j}\bigg|_{t=0} f(A+tE). \tag{4}$$

They called the $D_j^{[j]}(A,E)$ terms Fréchet derivatives. More precisely, $D_f^{[j]}(A,E)$ is a special case of the jth order Fréchet derivative described by Higham and Relton [11], in which the perturbations in the j directions are all E. The first of these terms, $D_f^{[1]}(A,E)$, coincides with the "standard" Fréchet derivative $L_f(A,E)$. Additionally, if A and E commute then we have $D_f^{[j]}(A,E) = E^j f^{(j)}(A)$, where $f^{(j)}$ denotes the jth derivative of the scalar function $f(x)$.

Before writing down the remainder term obtained by truncating the Taylor series in (3), we first recall the standard result that, for any invertible A and B,

$$A^{-1} - B^{-1} = A^{-1}(B-A)B^{-1}. \tag{5}$$

We will also need the following lemma.

Lemma 2.1. *Let $X(t) = A - tB$, where t is a scalar. Then*

$$\frac{d^j}{dt^j}\bigg|_{t=0} X(t)^{-1} = j!\, A^{-1}(BA^{-1})^j.$$

Proof. Note that

$$\frac{d}{dt}X^{-1} = -X^{-1}X'X^{-1},$$

where X' denotes the derivative of X, and that, since higher derivatives of X vanish,

$$\frac{d^j}{dt^j} X^{-1} = (-1)^j j! \, X^{-1}(X'X^{-1})^j.$$

The result then follows by substituting $X = A - tB$ and setting $t = 0$.

□

We now state and prove the main result of this paper, which gives an explicit form of the remainder term when truncating (3).

Theorem 2.2. *Let $f : \mathbb{C} \to \mathbb{C}$ be analytic in an open subset $\mathcal{D} \subset \mathbb{C}$. Let $A,\ E \in \mathbb{C}^{n\times n}$ be such that $\Lambda(A),\ \Lambda(A+E) \subset \mathcal{D}$. Then for any $k \in \mathbb{N}$*

$$f(A+E) = T_k(A,E) + R_k(A,E),$$

where

$$T_k(A,E) = \sum_{j=0}^{k} \frac{1}{j!} D_f^{[j]}(A,E), \tag{6}$$

$$R_k(A,E) = \frac{1}{2\pi i} \int_\Gamma f(z)(zI - A - E)^{-1}[E(zI-A)^{-1}]^{k+1} dz, \tag{7}$$

and Γ is a closed contour in $\mathcal{D}$ enclosing $\Lambda(A)$ and $\Lambda(A+E)$.

Proof. The result is proved by induction on k. For the case $k = 0$ we have $f(A+E) = f(A) + R_0(A,E)$. Then

$$\begin{aligned} R_0(A,E) &= f(A+E) - f(A) \\ &= \frac{1}{2\pi i}\int_\Gamma f(z)[(zI-A-E)^{-1} - (zI-A)^{-1}]dz, \end{aligned}$$

using the Cauchy integral definition of a matrix function. It follows from (5) that

$$R_0(A,E) = \frac{1}{2\pi i}\int_\Gamma f(z)(zI-A-E)^{-1}E(zI-A)^{-1}dz.$$

For the inductive step, we assume that $f(A+E) = T_k(A,E) + R_k(A,E)$. The remainder for the degree-$(k+1)$ Taylor polynomial is given by

$$\begin{aligned} R_{k+1}(A,E) &= f(A+E) - T_{k+1}(A,E) \\ &= f(A+E) - T_k(A,E) - \frac{1}{(k+1)!} D_f^{[k+1]}(A,E) \\ &= R_k(A,E) - \frac{1}{(k+1)!}\, \frac{d^{k+1}}{dt^{k+1}}\bigg|_{t=0} f(A+tE). \end{aligned}$$

Substituting the inductive hypothesis for $R_k(A,E)$ and the Cauchy integral form for $f(A+tE)$ gives

$$R_{k+1}(A,E) = \frac{1}{2\pi i}\int_\Gamma f(z)(zI-A-E)^{-1}[E(zI-A)^{-1}]^{k+1}dz \\ - \frac{1}{2\pi i(k+1)!}\frac{d^{k+1}}{dt^{k+1}}\int_\Gamma f(z)(zI-A-tE)^{-1}dz.$$

By continuity, we can differentiate the integrand in the second term, and simplify it using Lemma 2.1. We obtain

$$\begin{aligned} R_{k+1}(A,E) &= \frac{1}{2\pi i}\int_\Gamma f(z)\left[(zI-A-E)^{-1}[E(zI-A)^{-1}]^{k+1}\right. \\ &\qquad \left.-(zI-A)^{-1}[E(zI-A)^{-1}]^{k+1}\right] dz \\ &= \frac{1}{2\pi i}\int_\Gamma f(z)(zI-A-E)^{-1}[E(zI-A)^{-1}]^{k+2}dz, \end{aligned}$$

where (5) has been used once more. This completes the proof. □

We end this section by briefly describing how Theorem 2.2 also allows us to obtain a remainder term for Padé approximants (this was first done in the scalar case by Elliot [6]).

Suppose that we approximate $f(z)$ using a rational function $p_m(z)/q_n(z)$, where $p_m(z)$ and $q_n(z)$ are polynomials of degree m and n respectively. The Padé approximant is the unique choice (up to scalar multiples) of $p_m(z)$ and $q_n(z)$ such that $f(z) - p_m(z)/q_n(z) = O(z^{m+n+1})$. Therefore, using the same rational function to approximate the corresponding matrix function we have $q_n(A)f(A) - p_m(A) = O(\|A\|^{m+n+1})$. We introduce the truncation error term $S_{m,n}(A)$ to the Padé approximant such that

$$f(A) = \frac{p_m(A)}{q_n(A)} - S_{m,n}(A).$$

Then

$$q_n(A)S_{m,n}(A) = q_n(A)f(A) - p_m(A) = O(\|A\|^{m+n+1}).$$

The term $q_n(A)S_{m,n}(A)$ is then the remainder term for the Taylor series for $q_n(A)f(A)$ about 0 with degree $m+n$. From (7) we obtain

$$S_{m,n}(A) = \frac{q_n(A)^{-1}A^{m+n+1}}{2\pi i}\int_\Gamma \frac{q_n(z)f(z)(zI-A)^{-1}}{z^{m+n+1}}\,dz,$$

where the closed contour Γ encloses $\Lambda(A)$ and the origin.

3. Application to condition numbers and pseudospectra

In this section we use Theorem 2.2 to study the behaviour of the condition number of a matrix function, which measures the sensitivity of $f(A)$ to small perturbations in A. We approach this using techniques borrowed from the analysis of pseudospectra. Recall that the ϵ-pseudospectrum of a matrix X is the set

$$\Lambda_\epsilon(X) = \left\{ z \in \mathbb{C} : \|(zI - X)^{-1}\| \geq \epsilon^{-1} \right\}. \tag{8}$$

To begin, the following lemma provides some pseudospectral bounds on the size of the remainder terms.

Lemma 3.1. *Let $\epsilon > 0$ be such that f in Theorem 2.2 is analytic on the ϵ-pseudospectra of A and $A + E$, that is $\Lambda_\epsilon(A) \subset \mathcal{D}$ and $\Lambda_\epsilon(A + E) \subset \mathcal{D}$. Let $\tilde{\Gamma}_\epsilon \subset \mathcal{D}$ be a closed contour that encloses $\Lambda_\epsilon(A)$ and $\Lambda_\epsilon(A + E)$. Then the remainder term $R_k(A, E)$ is bounded by*

$$\|R_k(A, E)\| \leq \frac{\|E\|^{k+1} \tilde{L}_\epsilon}{2\pi\epsilon^{k+2}} \max_{z \in \tilde{\Gamma}_\epsilon} |f(z)|, \tag{9}$$

where $\tilde{L}_\epsilon$ is the length of $\tilde{\Gamma}_\epsilon$. In particular, when a circular contour can be used,

$$\|R_k(A, E)\| \leq \frac{\|E\|^{k+1}}{\epsilon^{k+2}} \max_{\theta \in [0, 2\pi]} |f(\tilde{\rho}_\epsilon e^{i\theta})|, \tag{10}$$

where $\tilde{\rho}_\epsilon = \max\{|z| : z \in \Lambda_\epsilon(A + E) \cap \Lambda_\epsilon(A)\}$.

(Note that tildes on $\tilde{L}_\epsilon$, $\tilde{\Gamma}_\epsilon$, and $\tilde{\rho}_\epsilon$ are used because, for this result only, the contour needs to enclose $\Lambda_\epsilon(A + E)$ in addition to $\Lambda_\epsilon(A)$. For subsequent results, the contour need only enclose $\Lambda_\epsilon(A)$ and the tildes are dropped.)

Proof. The proof is analogous to that of the bound

$$\|f(A)\| \leq \frac{\tilde{L}_\epsilon}{2\pi\epsilon} \max_{z \in \tilde{\Gamma}_\epsilon} |f(z)|,$$

obtained by Trefethen and Embree [17, Ch. 14]. We bound the norm of $R_k(A, E)$ by noting that

$$\|R_k(A, E)\| \leq \frac{\|E\|^{k+1}}{2\pi} \int_{\tilde{\Gamma}_\epsilon} |f(z)| \|(zI - A - E)^{-1}\| \|(zI - A)^{-1}\|^{k+1}.$$

On $\tilde{\Gamma}_\epsilon$ we have $\|(zI - A - E)^{-1}\| \leq \epsilon^{-1}$ and $\|(zI - A)^{-1}\| \leq \epsilon^{-1}$. The first part of the lemma follows immediately. For the second part, take $\tilde{\Gamma}_\epsilon$ to be a circle with centre 0 and radius $\tilde{\rho}_\epsilon = \max\{|z| : z \in \Lambda_\epsilon(A + E) \cap \Lambda_\epsilon(A)\}$. Note that a circular contour is not applicable for all functions, for example those with a branch cut. □

We can also use this result to bound the absolute condition number of a matrix function. Recall that the absolute condition number measures the first order sensitivity of $f(A)$ to small perturbations in A and is given by [9, Chap. 3]

$$\begin{aligned} \mathrm{cond}_{\mathrm{abs}}(f, A) &:= \lim_{\tau \to 0} \sup_{\|E\| \leq \tau} \frac{\|f(A + E) - f(A)\|}{\tau} \\ &= \max_{\|E\| \leq 1} \|L_f(A, E)\|. \end{aligned} \tag{11}$$

Lemma 3.1 provides us with the following bound on the absolute condition number.

Corollary 3.2. *Let $\epsilon > 0$ be such that f in Theorem 2.2 is analytic on the ϵ-pseudospectrum of A and let $\mathit{\Gamma}_\epsilon \subset \mathcal{D}$ be a closed contour of length L_ϵ that encloses the pseudospectrum. Then*

$$\mathrm{cond}_{\mathrm{abs}}(f, A) \leq \frac{L_\epsilon}{2\pi\epsilon^2} \max_{z \in \mathit{\Gamma}_\epsilon} |f(z)|. \tag{12}$$

In particular, when a circular contour can be used,

$$\mathrm{cond}_{\mathrm{abs}}(f, A) \leq \frac{\rho_\epsilon}{\epsilon^2} \max_{\theta \in [0, 2\pi]} |f(\rho_\epsilon e^{i\theta})|, \tag{13}$$

where $\rho_\epsilon = \max\{|z| : z \in \mathit{\Lambda}_\epsilon(A)\}$ is the pseudospectral radius of A.

Proof. Set $k = 0$ in (9). Suppose that $\|E\| = \alpha$. Then, since $R_0(A, E) = L_f(A, E) + o(\|E\|)$, we have

$$\|L_f(A, E) + o(\alpha)\| \leq \frac{\alpha \tilde{L}_\epsilon}{2\pi\epsilon^2} \max_{z \in \mathit{\Gamma}_\epsilon} |f(z)|.$$

We divide by α and take the supremum over all E such that $\|E\| \leq \alpha$ to obtain

$$\sup_{\|E\| \leq \alpha} \|L_f(A, E/\alpha) + o(\alpha)/\alpha\| \leq \frac{\tilde{L}_\epsilon}{2\pi\epsilon^2} \max_{z \in \tilde{\mathit{\Gamma}}_\epsilon} |f(z)|.$$

Note that the curve $\tilde{\mathit{\Gamma}}_\epsilon$ must enclose $\mathit{\Lambda}_\epsilon(A + E)$ for each $\|E\| \leq \alpha$. The proof of (12) is completed by taking the limit $\alpha \to 0$ and recalling that the absolute condition number of a matrix function is given by operator norm of the Fréchet derivative (11). In the limit $\alpha \to 0$, the curve $\tilde{\mathit{\Gamma}}_\epsilon \subset \mathcal{D}$ can become any closed contour $\mathit{\Gamma}_\epsilon \subset \mathcal{D}$ enclosing $\mathit{\Lambda}_\epsilon(A)$.

The proof of (13) is essentially the same, except that (10) is taken as the starting point rather than (9).

Note that an alternative proof of the corollary can be obtained by starting with the integral representation of the Fréchet derivative

$$L_f(A, E) = \frac{1}{2\pi i} \int_{\mathit{\Gamma}_\epsilon} f(z)(zI - A)^{-1} E (zI - A)^{-1} dz,$$

and bounding it above using the techniques from the proof of Lemma 3.1.

□

Assuming that these bounds can be computed efficiently they are of considerable interest since most existing results regarding the estimation of the condition number provide only lower bounds [9, Chap. 3]. In section 5 we show how to calculate this bound efficiently for matrix powers A^t, where $t \in (0, 1)$. We obtain large speedups over existing methods whilst sacrificing some accuracy.

We end this section by briefly mentioning a related theorem due to Lui [14, Thm. 3.1], concerning the relationship between the pseudospectra of A and $f(A)$. The theorem is restated here in our notation. Recall that $R_k(A,E)$ was defined in Theorem 2.2 and that $R_0(A,E) = L_f(A,E) + o(\|E\|)$.

Lemma 3.3 (Lui). *Let ϵ, f, and Γ_ϵ satisfy the conditions of Corollary 3.2. Furthermore let $f(\Lambda_\epsilon(A)) = \{f(z) : z \in \Lambda_\epsilon(A)\}$ and $M = \max_{\|E\|\le\epsilon} \|R_0(A,E)\|$. Then $f(\Lambda_\epsilon(A)) \subset \Lambda_M(f(A))$.*

Proof. If z is an eigenvalue of $A+E$ with $\|E\| \le \epsilon$ (so that $z \in \Lambda_\epsilon(A)$), then $f(z)$ is an eigenvalue of $f(A+E) = f(A) + R_0(A,E)$ and $\|R_0(A,E)\| \le M$. □

This result shows that, to first order in ϵ, the ϵ-pseudospectrum of A is related to the δ-pseudospectrum of $f(A)$ via $f(\Lambda_\epsilon(A)) \subset \Lambda_\delta(f(A))$, where $\delta = \mathrm{cond}_{\mathrm{abs}}(f,A)\epsilon$.

4. Application to higher order condition numbers

Higham and Relton [11] introduce the level-q condition number for matrix functions, which is defined recursively by

$$\mathrm{cond}_{\mathrm{abs}}^{(q)}(f,A) := \lim_{\alpha\to 0} \sup_{\|Z\|\le\alpha} \frac{|\,\mathrm{cond}_{\mathrm{abs}}^{(q-1)}(f,A+Z) - \mathrm{cond}_{\mathrm{abs}}^{(q-1)}(f,A)|}{\alpha}, \qquad (14)$$

where $\mathrm{cond}_{\mathrm{abs}}^{(1)}(f,A) := \mathrm{cond}_{\mathrm{abs}}(f,A)$. In section 3 we focused on the first order remainder term, $R_0(A,E)$, and results concerning the condition number $\mathrm{cond}_{\mathrm{abs}}(f,A)$ but—by choosing $k > 0$ in Lemma 3.1—we can attempt to extend results such as Corollary 3.2 to these higher order condition numbers.

Before proceeding, we must first investigate the relationship between the $D_f^{[j]}(A,E)$ defined in (4) and higher order Fréchet derivatives. Recall that $D_f^{[j]}(A,E)$ is a special case of the jth order Fréchet derivative in which the perturbation in each direction is E. In [11] a definition of the jth order Fréchet derivative is given in terms of the mixed partial derivative:

$$L_f^{(j)}(A,E_1,\dots,E_j) = \frac{\partial}{\partial s_1}\cdots\frac{\partial}{\partial s_j}\bigg|_{(s_1,\dots,s_j)=0} f(A + s_1E_1 + \cdots + s_jE_j). \qquad (15)$$

The following theorem expresses this jth order Fréchet derivative in terms of a contour integral.

Theorem 4.1. *The jth order Fréchet derivative of a matrix function $f(A)$ in the directions $E_1,\dots,E_j$ is given by*

$$L_f^{(j)}(A,E_1,\dots,E_j) = \frac{1}{2\pi i}\int_\Gamma f(z)(zI-A)^{-1} \sum_{\sigma\in\mathcal{S}_j} \prod_{i=1}^{k} E_{\sigma(i)}(zI-A)^{-1}dz, \qquad (16)$$

where $\mathit{\Gamma}$ is a closed curve enclosing $\mathit{\Lambda}(A)$, within which f is analytic, and $\mathcal{S}_j$ is the set of permutations of $\{1, 2, \ldots, k\}$. In particular the derivative $D_f^{[j]}(A, E)$ is given by

$$D_f^{[j]}(A, E) = \frac{j!}{2\pi i} \int_{\mathit{\Gamma}} f(z)(zI - A)^{-1}[E(zI - A)^{-1}]^{j+1} dz. \tag{17}$$

Proof. For any choice of s_i and E_i, we can write $f(A + s_1E_1 + \cdots + s_jE_j)$ as a Cauchy integral by using the standard Cauchy integral definition of a matrix function and choosing a contour $\tilde{\mathit{\Gamma}}$ that encloses $\mathit{\Lambda}(A + s_1E_1 + \cdots + s_jE_j)$. Then (15) becomes

$$L_f^{(j)}(A, E_1, \ldots, E_j) = \\ \frac{\partial}{\partial s_1} \cdots \frac{\partial}{\partial s_j}\bigg|_{(s_1,\ldots,s_j)=0} \int_{\tilde{\mathit{\Gamma}}} f(z)(zI - (A + s_1E_1 + \cdots + s_jE_j))^{-1} dz.$$

By continuity, the differential operator

$$\frac{\partial}{\partial s_1} \cdots \frac{\partial}{\partial s_j}\bigg|_{(s_1,\ldots,s_j)=0}$$

can be brought inside the integral sign. The required integrand is then obtained by using the identity

$$\frac{d}{dx} U^{-1} = -U^{-1} \frac{dU}{dx} U^{-1}.$$

The result (16) follows by then restricting the contour to any closed curve $\mathit{\Gamma}$ containing $\mathit{\Lambda}(A)$. The second part of the theorem, (17), follows by setting $E_1 = \cdots = E_j$. □

Theorem 4.1 shows that, to first order, the kth remainder term in the Taylor series is simply the $(k + 1)$st derivative, as we might expect. Specifically, comparing (17) with (7) we find

$$R_k(A, E) = \frac{1}{(k + 1)!} D_f^{[k+1]}(A, E) + o(\|E\|^{k+2}).$$

In addition, Theorem 4.1 allows us to prove the following theorem, which uses the pseudospectrum of A to bound the norm of the jth order Fréchet derivative.

Theorem 4.2. *Let $\mathit{\Gamma}_\epsilon \subset \mathcal{D}$ be a closed contour enclosing the ϵ-pseudospectrum of A within which f is analytic. The jth order Fréchet derivative can be bounded by*

$$\|L_f^{(j)}(A, E_1, \ldots, E_j)\| \leq \frac{j!\, L_\epsilon}{2\pi \epsilon^{j+1}} \left(\max_{z \in \mathit{\Gamma}_\epsilon} |f(z)| \right) \prod_{i=1}^{j} \|E_i\|, \tag{18}$$

where L_ϵ is the length of $\mathit{\Gamma}_\epsilon$.

Proof. In (16), use the contour $\mathit{\Gamma}_\epsilon$, take norms and note that $\|(zI-A)^{-1}\| \le \epsilon^{-1}$ on $\mathit{\Gamma}_\epsilon$. □

It would be desirable to obtain a bound on the level-q condition number, by first bounding it in terms of the norm of the qth Fréchet derivative and then applying Theorem 4.2. However, in the general case such bounds prove to be far too weak to be of any interest. Instead we restrict ourselves to the case $q = 2$ and the level-2 condition number.

Lemma 4.3. *The level-*2 *condition number is bounded by*

$$\mathrm{cond}_{\mathrm{abs}}^{(2)}(f, A) \le \frac{L_\epsilon}{\pi\epsilon^3} \max_{z\in\mathit{\Gamma}_\epsilon} |f(z)|,$$

where L_ϵ is the length of a closed contour $\mathit{\Gamma}_\epsilon \subset \mathcal{D}$ enclosing $\mathit{\Lambda}_\epsilon(A)$ within which f is analytic. When a circular contour is applicable

$$\mathrm{cond}_{\mathrm{abs}}^{(2)}(f, A) \le \frac{2\rho_\epsilon}{\epsilon^3} \max_{\theta\in[0,2\pi]} |f(\rho_\epsilon e^{i\theta})|,$$

where ρ_ϵ is the pseudospectral radius.

Proof. Higham and Relton [11, Sec. 5] give an upper bound for the level-2 absolute condition number in terms of the norm of the 2nd Fréchet derivative

$$\mathrm{cond}_{\mathrm{abs}}^{(2)}(f, A) \le \max_{\|E_1\|=1} \max_{\|E_2\|=1} \|L_f^{(2)}(A, E_1, E_2)\|. \tag{19}$$

Substituting the bound from (18) into (19) gives the required results. □

5. Numerical Experiments

In this section we show how the pseudospectral bounds on the condition number, (12) and (13), can be used to estimate the condition number of matrix powers at extremely low cost. Due this low cost, one might use the pseudospectral bound as a quick estimate of the condition number and, if it is unsatisfactorily large, use existing methods to estimate it more accurately.

Recall from (12) that

$$\mathrm{cond}_{\mathrm{abs}}(f, A) \le \frac{L_\epsilon}{2\pi\epsilon^2} \max_{z\in\mathit{\Gamma}_\epsilon} |f(z)|,$$

where $\mathit{\Gamma}_\epsilon$ is a closed contour of length L_ϵ that encloses the spectrum of A and within which $f(z)$ is analytic. Recall also that the relative condition number $\mathrm{cond}_{\mathrm{rel}}(f, A)$, is given by

$$\mathrm{cond}_{\mathrm{rel}}(f, A) = \mathrm{cond}_{\mathrm{abs}}(f, A)\frac{\|A\|}{\|f(A)\|}.$$

Combining these two results allows us to bound the relative condition number from above. This bound will be cheap to compute provided that the cost of computing L_ϵ and $\max_{z\in\Gamma_\epsilon} |f(z)|$ is sufficiently small.

In this section we will focus on matrix powers A^t for $t \in (0,1)$, so that $f(z) = z^t$. This restriction will enable us to easily compute $\max_{z\in\Gamma_\epsilon} |f(z)|$ provided a suitable contour is chosen. Fractional powers of a matrix arise in a number of applications such as Markov chain models from healthcare and finance [3], [13].

The functions $f(z) = z^t$ have a branch cut, conventionally taken to be along the closed negative real axis. If we choose ϵ such that $\Lambda_\epsilon(A)$ does not contain any segment of the branch cut, then the contour Γ_ϵ can be taken to be a "keyhole" contour, enclosing $\Lambda_\epsilon(A)$ but avoiding the closed negative real axis. We can take ρ_ϵ, the ϵ-pseudospectral radius, to be the radius of the outer circle of the keyhole. As the inner circle becomes infinitesimally small, the overall length of Γ_ϵ is $L_\epsilon = 2(\pi + 1)\rho_\epsilon$. It is also easy to see that $\max_{z\in\Gamma_\epsilon} |f(z)| = \rho_\epsilon^t$, and therefore

$$\operatorname{cond}_{\mathrm{rel}}(f, A) \leq \frac{2(\pi+1)\rho_\epsilon^{1+t}}{2\pi\epsilon^2}\frac{\|A\|}{\|f(A)\|}. \tag{20}$$

It remains to choose ϵ, ensuring that the ϵ-pseudospectrum does not cross the branch cut. One heuristic way to do this is to find the closest point on the branch cut for each eigenvalue of A and calculate the value of ϵ for which the boundary of $\Lambda_\epsilon(A)$ would intersect this point. Any value of ϵ less than these would be permissible. This leads to the following algorithm.

```
Compute the eigenvalues $\lambda_1, \dots, \lambda_n$ of $A$.
for $i = 1 : n$
    Find $z_i$, the nearest point to $\lambda_i$ on the negative real line.
    resnorm($i$) = $\|(A - z_i I)^{-1}\|$
end for
$\epsilon_{\max} = 1/\max(\text{resnorm})$
```

We can then select $\epsilon \in [0, \epsilon_{\max}]$ in the upper bound (20).

Note that, up to this point, our analysis has been largely independent of the matrix norm used. In order to compare this upper bound against the exact condition number we will need to choose a specific norm. It is known that, in the Frobenius norm, $\operatorname{cond}_{\mathrm{rel}}(f, A) = \|K_f(A)\|_2\|A\|_2/\|f(A)\|_2$, where $K_f(A)$ is the Kronecker form of the matrix function [9, Alg. 3.17].

In order to use our upper bound (20) in the Frobenius norm we will need to compute the corresponding ϵ-pseudospectral radius in this norm. The 2-norm pseudospectral radius can be computed at a very low cost using an algorithm by Guglielmi and Overton [7] so we can instead bound the pseudospectral radius in the Frobenius norm by that in the 2-norm. Since all norms over $\mathbb{C}^{n\times n}$ are equivalent, it is simple to show that ρ_ϵ in the Frobenius norm is less than $\rho_{\epsilon\sqrt{n}}$ in the 2-norm and so we have the following practical bound.

Lemma 5.1. *The relative condition number in the Frobenius norm of the matrix function $f(A) = A^t$, where $0 < t < 1$, can be bounded above by*

$$\operatorname{cond}_{\mathrm{rel}}(f, A) \leq \frac{2(\pi + 1)\rho_{\epsilon\sqrt{n}}^{1+t}}{2\pi\epsilon^2} \frac{\|A\|_F}{\|f(A)\|_F}, \tag{21}$$

where $\rho_{\epsilon\sqrt{n}}$ is the $\epsilon\sqrt{n}$ pseudospectral radius computed in the 2-norm.

Proof. See above. □

Our experiments will compare our estimate of the Frobenius norm condition number in Lemma 5.1, henceforth referred to as `CN Pseudo`, against two other algorithms. The first of these computes the exact condition number in the Frobenius norm by explicitly forming the Kronecker form $K_f(A)$ and taking its 2-norm at a cost of $O(n^5)$ flops. This requires computing multiple Fréchet derivatives of the matrix function, which we perform using an algorithm of Higham and Lin [10]. We refer to this method as `CN Exact`. Secondly we compare against the current state-of-the-art method which, using a block 1-norm estimator of Higham and Tisseur [12] estimates $\|K_f(A)\|_1$ in only $O(n^3)$ flops to approximate the 1-norm condition number. This method avoids computing $K_f(A)$ explicitly and needs only matrix-vector products with $K_f(A)$ which are given by $K_f(A)\operatorname{vec}(E) = L_f(A, E)$. The vec operator stacks the columns of a matrix vertically from left to right. We refer to this method as `CN Normest`.

Note that although this means we will be comparing condition numbers computed in different norms (since the Frobenius and 1-norms differ by at most a factor $\sqrt{n}$ and we use $n = 20$ in our experiments) this is not a completely unfair comparison. Indeed in Figure 2 as part of our first experiment we shall see that the values returned by `CN Exact` and `CN Normest` are indistinguishable on our graphs.

Our experiments use four test matrices whose pseudospectra are shown in Figure 1. Clockwise from the top-left these are the "airy" matrix from EIGTOOL, a matrix with eigenvalues sampled randomly from a Uniform$(-1, 0)$ distribution (with a small perturbation of the order 1e-3 to each eigenvalue to avoid the branch cut), the 1D Laplace operator known as "tridiag" in the MATLAB gallery, and the "grcar" matrix from the MATLAB gallery, all of dimension $n = 20$. All experiments are performed in MATLAB 2014b.

Our first experiment in Figure 2 shows, for each of our test matrices, the upper bound `CN Pseudo` against `CN Exact` and `CN Normest` as ϵ varies from 10^{-8} to 10, unless $\epsilon_{\max}$ is reached first. The lines corresponding to `CN Normest` and `CN Exact` overlap almost exactly on each plot. The distance between the upper bound and the exact condition number decreases almost linearly on the log-log plot as ϵ increases (until $\epsilon_{\max}$ is reached), suggesting that our upper bound behaves like $C\epsilon^m$ for some constants C and m. Generally the upper bound is 1–2 orders of magnitude above the exact condition number for $\epsilon = 0.1$, though the performance is worse for the upper-right plot corresponding to the matrix with eigenvalues close to the branch cut, forcing $\epsilon_{\max}$ to be rather small.

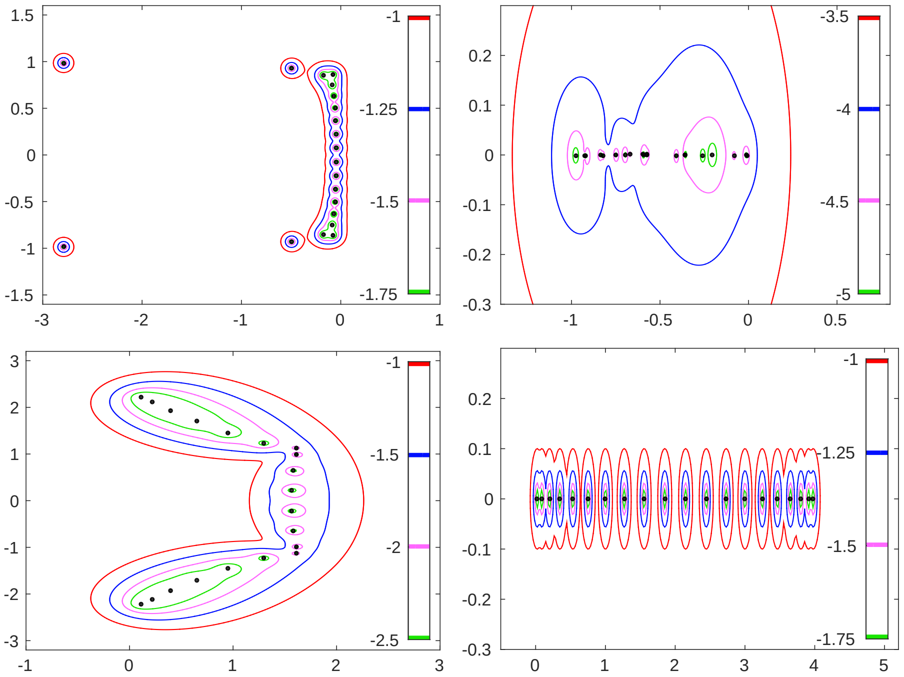

Figure 1: Pseudospectra of the four test matrices. Clockwise from top left: the "airy" matrix, a matrix with eigenvalues chosen randomly but close to the branch cut, the 1D Laplace operator, and the "grcar" matrix.

The blank spaces in the lower-right plot are where the code to calculate the pseudospectral radius failed for these values of ϵ.

Our next experiment investigates the reliability of `CN Pseudo` multiplied by the unit roundoff as a bound on the relative error. More precisely, let $\widehat{F}$ denote our computed value of A^t then—since the algorithm we use to compute A^t is backward stable in exact arithmetic—it is reasonable to expect that the relative error will approximately satisfy

$$\frac{\|A^t - \widehat{F}\|_F}{\|A^t\|_F} \leq \operatorname{cond}_{\mathrm{rel}}(x^t, A)u,$$

where $u = 2^{-53}$ is the unit roundoff in IEEE double precision arithmetic. Indeed the forward stability of this algorithm was observed in [10]. We can compute the relative error by obtaining an "exact" value for A^t.

The "exact" value of A^t is computed by using 250 digit arithmetic from the MATLAB symbolic toolbox. We make a random perturbation of norm 10^{-125} to A to ensure that, with probability 1, it has distinct eigenvalues. Following this we perform the diagonalization $A = XDX^{-1}$ and then $f(A) = Xf(D)X^{-1}$ where $f(D)$ is diagonal and $f(D)_{ii} = f(d_{ii})$. This idea was introduced by Davies [4], though not for high precision arithmetic.

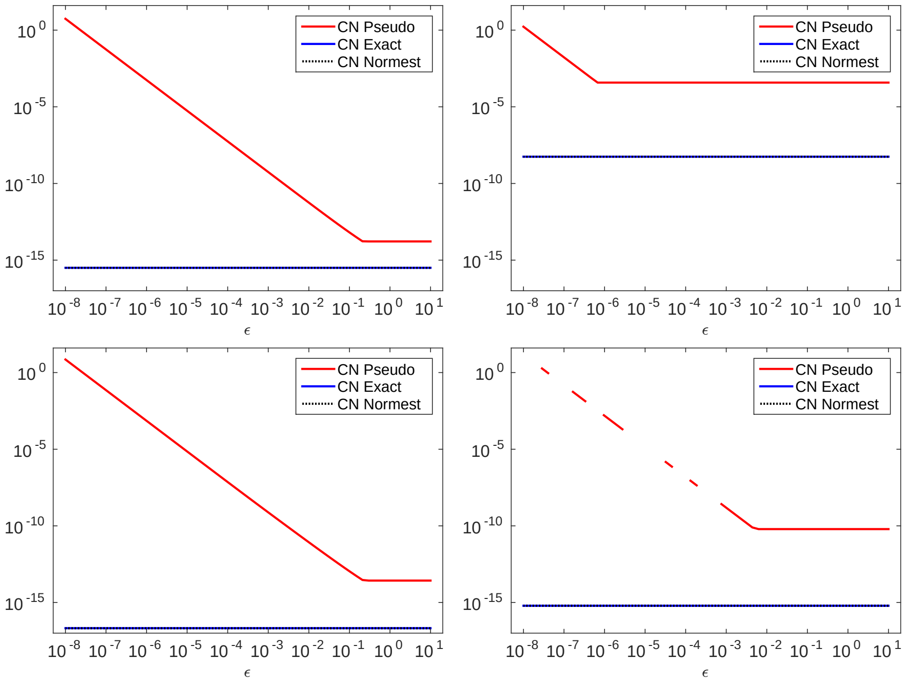

Figure 2: Condition number bounds for the test problems detailed in Figure 1 for $\epsilon \in [0, \epsilon_{\max}]$. `CN Pseudo` denotes the bound (21), `CN Exact` denotes the condition number in the Frobenius norm, and `CN Normest` denotes the 1-norm condition number estimate. In each plot `CN Normest` and `CN Exact` overlap almost exactly.

In Table 1, for each of our four test matrices, we list the relative error and the values of `CN Exact`, `CN Normest`, and `CN Pseudo` with the latter using the parameter $\epsilon = \min\{0.1, \epsilon_{\max}\}$. In each case we see that `CN Pseudo` is an upper bound on the condition number, and can be used to bound the relative error of our computed matrix function, though it can be pessimistic in some cases. Since our next experiment shows that the computation of `CN Pseudo` is far cheaper than the alternatives, we recommend trying it initially before using one of the alternatives if the upper bound on the relative error that `CN Pseudo` returns is larger than desired.

Our final experiment, shown in Figure 3, explores how the running time for computing `CN Pseudo` compares with `CN Normest`. We omit `CN Exact` for this experiment since its $O(n^5)$ flop count makes it prohibitively expensive for larger matrices. We consider $t \in \{1/5, 1/10, 1/15\}$ and use random matrices with n varying between 10 and 200 with each element selected from the Normal$(0, 1)$ distribution.

The plot on the left of Figure 3 shows the time taken to compute the condition number using the two methods. The three lines corresponding to `CN Pseudo` overlap almost exactly at the bottom of the plot. As n increases both

Table 1: Condition numbers and condition number estimates for the test problems detailed in Figure 1, multiplied by the unit roundoff $u = 2^{-53}$, where `CN Pseudo` is run with $\epsilon = \min\{0.1, \epsilon_{\max}\}$. The relative errors are calculated by comparing the code in [10] against the functions computed in 250 digit arithmetic using the MATLAB symbolic toolbox.

Matrix	airy	random	laplace	grcar
Rel. Err.	1e-14	4e-10	1e-14	4e-15
`CN Exact`	3e-16	6e-9	2e-17	6e-16
`CN Pseudo`	3e-14	4e-4	5e-14	6e-11
`CN Normest`	3e-16	6e-9	2e-17	6e-16

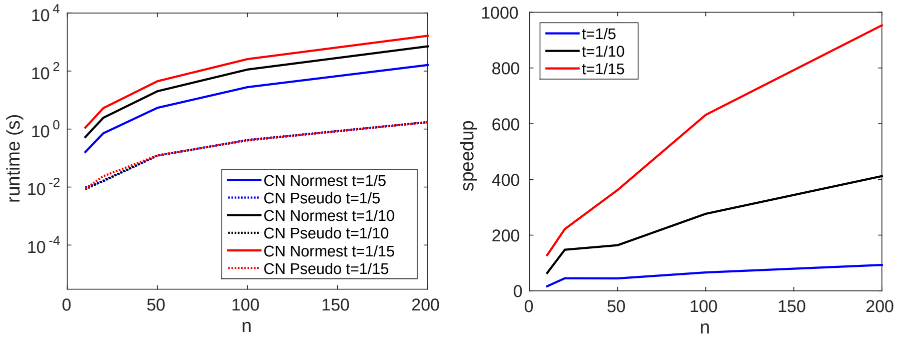

Figure 3: Run times and corresponding speedups for condition number estimation of A^t for random matrices of varying size n, and for $t \in \{1/5, 1/10, 1/15\}$. `CN Pseudo` denotes the bound (21) and `CN Normest` denotes the 1-norm condition number estimate. The three lines corresponding to `CN Pseudo` overlap almost exactly at the bottom of the plot on the left.

methods find it more difficult to estimate the condition number. However, increases in t have a small effect on our upper bound whilst they slow the norm estimation method considerably. This is due to the increased cost of calculating the required Fréchet derivatives. The plot on the right shows the speedup obtained by computing the upper bound instead. For larger values of t speedups of over 900x can be obtained.

It appears that, although our upper bound on the condition number can be loose, it is exceptionally cheap to compute. If a user needs to compute A^t to some desired accuracy, often only a few digits in many applications, then `CN Pseudo` offers a fast way to check this has been obtained. However, if the upper bound determines that the error might be unacceptably large, the user can continue to calculate the condition number more accurately using `CN Normest`.

We end this section by explaining why the methods above are not of practical use when applied to the matrix inverse. The function $f(z) = z^{-1}$ is not analytic at the origin so a keyhole contour can again be used to excise this point. The term $\max_\Gamma |f(z)|$ takes its maximum on the inner circle of the contour. To make the bound as tight as possible we should therefore take the inner circle to be as large as possible. The maximum possible radius for the inner circle is $|\lambda_{\min}|$ where $\lambda_{\min}$ is the eigenvalue of smallest magnitude. Combining these results, the bound (12) becomes

$$\mathrm{cond}_{\mathrm{rel}}(z^{-1}, A) \leq \frac{(\pi\rho_\epsilon + \pi|\lambda_{\min}| + \rho_\epsilon - |\lambda_{\min}|)}{\pi\epsilon^2|\lambda_{\min}|} \frac{\|A\|}{\|A^{-1}\|}.$$

Since the exact condition number can be obtained by estimating $\|A\|\|A^{-1}\|$, there is no computational advantage to be gained from using our estimate.

6. Conclusions

The main results in this paper are as follows. We have obtained an explicit expression for the remainder term of a matrix function Taylor polynomial (Theorem 2.2). Combining this with use of the pseudospectrum of A leads to upper bounds on the higher-order condition numbers of $f(A)$. In the case $f(A) = A^t, t \in (0, 1)$, we demonstrated how the bound on the level-1 condition number can be computed very efficiently and far more cheaply than standard condition number estimation methods. Our bounds could be used as a quick estimate of the condition number. If this estimate is too large, for example if the estimate suggests that an insufficient number of correct significant figures might be obtained in computing $f(A)$, then existing methods can be used to obtain the condition number more accurately.

Our results may also have further useful applications in the development of matrix function algorithms, by allowing us to estimate the size of remainder terms for Padé approximants. This will be the subject of future work.

Acknowledgements

We are very grateful to Nick Higham for numerous helpful comments on earlier drafts of this work.

References

[1] Lars V. Ahlfors. *Complex Analysis.* McGraw-Hill, New York, third edition, 1979. ISBN 978-0-0700-0657-7.

[2] Awad H. Al-Mohy and Nicholas J. Higham. The complex step approximation to the Fréchet derivative of a matrix function. *Numer. Algorithms*, 53 (1):133–148, 2010. doi: 10.1007/s11075-009-9323-y.

[3] Theodore Charitos, Peter R. de Waal, and Linda C. van der Gaag. Computing short-interval transition matrices of a discrete-time Markov chain from partially observed data. *Stat. Med.*, 27(6):905–921, 2008. doi: 10.1002/sim.2970.

[4] E. B. Davies. Approximate diagonalization. *SIAM J. Matrix Anal. Appl.*, 29(4):1051–1064, 2007. doi: 10.1137/060659909.

[5] Philip I. Davies and Nicholas J. Higham. A Schur–Parlett algorithm for computing matrix functions. *SIAM J. Matrix Anal. Appl.*, 25(2):464–485, 2003. doi: 10.1137/S0895479802410815.

[6] David Elliott. Truncation errors in Padé approximations to certain functions: An alternative approach. *Math. Comp.*, 21(99):398–406, 1967. ISSN 00255718.

[7] Nicola Guglielmi and Michael L. Overton. Fast algorithms for the approximation of the pseudospectal radius of a matrix. *SIAM J. Matrix Anal. Appl.*, 32(4):1166–1192, 2011.

[8] Kurt Hensel. Über Potenzreihen von Matrizen. *J. Reine Angew. Math.*, 155(42):100–110, 1926.

[9] Nicholas J. Higham. *Functions of Matrices: Theory and Computation.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008. ISBN 978-0-898716-46-7. doi: 10.1137/1.9780898717778.

[10] Nicholas J. Higham and Lijing Lin. An improved Schur–Padé algorithm for fractional powers of a matrix and their Fréchet derivatives. *SIAM J. Matrix Anal. Appl.*, 34(3):1341–1360, 2013. doi: 10.1137/130906118.

[11] Nicholas J. Higham and Samuel D. Relton. Higher order Fréchet derivatives of matrix functions and the level-2 condition number. *SIAM J. Matrix Anal. Appl.*, 35(3):1019–1037, 2014. doi: 10.1137/130945259.

[12] Nicholas J. Higham and Françoise Tisseur. A block algorithm for matrix 1-norm estimation, with an application to 1-norm pseudospectra. *SIAM J. Matrix Anal. Appl.*, 21(4):1185–1201, 2000. doi: 10.1137/S0895479899356080.

[13] R. B. Israel, J. S. Rosenthal, and J. Z. Wei. Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings. *Math. Finance*, 11:245–265, 2001.

[14] S.-H. Lui. A pseudospectral mapping theorem. *Math. Comp.*, 72(244): 1841–1854, 2003.

[15] Roy Mathias. Approximation of matrix-valued functions. *SIAM J. Matrix Anal. Appl.*, 14(4):1061–1063, 1993.

[16] Walter Rudin. *Real and Complex Analysis.* McGraw-Hill, New York, third edition, 1986. ISBN 0070542341.

[17] Lloyd Nicholas Trefethen and Mark Embree. *Spectra and pseudospectra: the behavior of nonnormal matrices and operators.* Princeton University Press, 2005.

[18] Edouard Weyr. Note sur la théorie de quantités complexes formées avec n unités principales. *Bull. Sci. Math. II*, 11:205–215, 1887.