COMPUTING THE STRUCTURED PSEUDOSPECTRUM OF A TOEPLITZ MATRIX AND ITS EXTREME POINTS

P. BUTTÀ†, N. GUGLIELMI‡, AND S. NOSCHESE†

Abstract. The computation of the structured pseudospectral abscissa and radius (with respect to the Frobenius norm) of a Toeplitz matrix is discussed and two algorithms based on a low-rank property to construct extremal perturbations are presented. The algorithms are inspired by those considered in [N. Guglielmi and M. Overton, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 1166–1192] for the unstructured case, but their extension to structured pseudospectra and analysis presents several difficulties. Natural generalizations of the algorithms, allowing us to draw significant sections of the structured pseudospectra in proximity of extremal points, are also discussed. Since no algorithms are available in the literature to draw such structured pseudospectra, the approach we present seems promising to extend existing software tools (Eigtool [T. G. Wright, Eigtool: A Graphical Tool for Nonsymmetric Eigenproblems, http://www.comlab.ox.ac.uk/pseudospectra/eigtool (2002)], Seigtool [M. Karow, E. Kokiopoulou, and D. Kressner, Systems Control Lett., 59 (2010), pp. 122–129] to structured pseudospectra representation for Toeplitz matrices. We discuss local convergence properties of the algorithms and show some applications to a few illustrative examples.

Key words. pseudospectrum, structured pseudospectrum, eigenvalue, spectral abscissa, spectral radius, Toeplitz structure

AMS subject classifications. 65F15, 65L07

DOI. 10.1137/120864349

1. Introduction. There is a growing development of structure-preserving algorithms for structured problems. Toeplitz matrices arise in many applications, including the solution of ordinary differential equations; thus it is meaningful to investigate the sensitivity of the eigenvalues of a Toeplitz matrix with respect to finite structure-preserving perturbations and, mainly, the sensitivity of the rightmost eigenvalue. The structure is given by the location of the nonzero diagonals of the matrix.

We add the structure requirement to the classical definition of $\varepsilon$-pseudospectrum; see, e.g., [TE05]. Given $\varepsilon > 0$, the structured $\varepsilon$-pseudospectrum of a given Toeplitz matrix $A \in \mathbb{C}^{n \times n}$ is the set of all eigenvalues of $A + \varepsilon E$ for some Toeplitz matrix $E \in \mathbb{C}^{n \times n}$ with unitary norm and with the same sparsity structure as $A$. As an example, if $A$ is a tridiagonal Toeplitz matrix, we consider all tridiagonal Toeplitz perturbation matrices of norm equal to $\varepsilon$. The structured pseudospectral abscissa is the maximal real part of points in the structured pseudospectrum.

We remark that the notion of $\varepsilon$-pseudospectrum depends on the choice of the matrix norm. In the literature, the spectral norm has been largely used also in the structured case [BGK01, G06, R06]. Guglielmi and Overton presented in [GO11] an efficient algorithm for computing the pseudospectral abscissa in the spectral norm. For our purposes, the Frobenius norm turns out to be the most appropriate. Since the points in a structured pseudospectrum are exact eigenvalues of some nearby Toeplitz matrix with the same structure diagonals as $A$, we are in a position to use results

---

*Received by the editors February 1, 2012; accepted for publication (in revised form) by Nicola Mastronardi September 24, 2012; published electronically DATE. This work was supported by the Italian MIUR and GNCS.

†Dipartimento di Matematica, SAPIENZA Università di Roma, P.le Aldo Moro 5, 00185 Roma, Italy (butta@mat.uniroma1.it, noschese@mat.uniroma1.it).

‡Dipartimento di Matematica Pura e Applicata, Università de L’Aquila, I-67100 L’Aquila, Italy (guglielm@univaq.it).
from the literature concerning the eigenvalue sensitivity to machine perturbations, that is, infinitely small structured perturbations. The structured condition number of an eigenvalue \( \lambda \in A \) is indeed a first-order measure of the worst-case effect on \( \lambda \) of perturbations of the same structure as \( A \). The structured conditioning measures we deal with can be computed endowing the subspace of matrices with the Frobenius norm; see, e.g., [HH92, KKT06, NP07, K10] and references therein.

Here we are concerned with the computation of the rightmost points in the structured pseudospectrum of a Toeplitz matrix. Since we are limiting finite perturbations to a given Toeplitz structure, it is not surprising that the main difference in our extension of the algorithm in [GO11] is the replacement of the classical eigenprojection for a simple eigenvalue with its structured analogue (normalized in the Frobenius norm).

We remark that we may generalize the above statements to nonreal Hankel matrices, considering antidiagonals in place of diagonals. Similarly, other symmetry-pattern nonnormal matrices can be treated (a symmetry-pattern being a structure that exhibits a kind of symmetry, like reflection or translation [NP07]), for instance, general persymmetric, skew-persymmetric, complex symmetric, or complex skew-symmetric matrices. In all cases, matrix perturbations with the given sparsity and symmetry-pattern have to be considered.

In this paper we thoroughly investigate the tridiagonal Toeplitz structure. The motivation is that the eigenvalues and eigenvectors of tridiagonal Toeplitz matrices are known in closed form, and all ingredients of our analysis are easily computable [NPR11]. Additionally, it is well known that the shape of the \( \varepsilon \)-pseudospectrum of a tridiagonal Toeplitz matrix approaches that of an ellipse, as \( \varepsilon \) tends to zero and the dimension \( n \) goes to infinity [RT92]. A slightly modified version of the algorithm in [GO11] for plotting the boundary of the \( \varepsilon \)-pseudospectrum constructs in fact an ellipse in the tridiagonal Toeplitz case. Analogously, we adapt the new algorithm in order to investigate the boundary of the structured \( \varepsilon \)-pseudospectrum in the Frobenius norm.

The paper is organized as follows. In section 2 we define the algorithm and show how to modify it to compute also the pseudospectral radius and partially draw the pseudospectral boundary. In section 3 we derive a local convergence analysis, establishing that the algorithm is linearly convergent to local maximizers of the structured pseudospectral abscissa. In section 4 the algorithms are tested on some examples. Finally, in the appendix we characterize the fixed points of the algorithm in the tridiagonal case.

2. The algorithm. We start with some notation and definitions. Given a Toeplitz matrix \( A \in \mathbb{C}^{n \times n} \) we denote by \( S \) the subspace of all Toeplitz matrices in \( \mathbb{C}^{n \times n} \) with the same sparsity structure as \( A \). We denote by \( M\vert_{S} \) the matrix in \( S \) closest to \( M \in \mathbb{C}^{n \times n} \) with respect to the Frobenius norm. It is straightforward to verify that \( M\vert_{S} \) is obtained by replacing in each structure diagonal all the entries of \( M \) with their arithmetic mean. We also define the normalized projection, where \( \| \cdot \|_{F} \) stands for the Frobenius norm,

\[
M\vert_{T} := \frac{M\vert_{S}}{\|M\vert_{S}\|_{F}}.
\]

If \( \lambda \) is a simple eigenvalue of a matrix \( M \in \mathbb{C}^{n \times n} \), a corresponding pair of right and left eigenvectors \( x \) and \( y \) are said normalized to be RP-compatible if \( \|x\|_{2} = \|y\|_{2} = 1 \) and \( y^{\ast}x \) is real and positive.

**Lemma 2.1** (see [NP07]). Let \( \lambda \) be a simple eigenvalue of a Toeplitz matrix \( A \) with corresponding right and left eigenvectors \( x \) and \( y \) normalized to be RP-compatible.
Given any Toeplitz matrix $E$ with $\|E\|_F = 1$, let $\lambda_E(t)$ be an eigenvalue of $A + tE$ converging to $\lambda$ as $t \to 0$. Then,

$$|\dot{\lambda}_E(0)| \leq \max \left\{ \left| \frac{y^* G x}{y^* x} \right| : G \in \mathcal{S}, \|G\|_F = 1 \right\} = \frac{\|yx^*|_S\|_F}{y^* x}$$

and

$$\dot{\lambda}_E(0) = \frac{\|yx^*|_S\|_F}{y^* x} > 0 \quad \text{if} \quad E = yx^*|_T.$$ 

**Remark 2.2.** If the right and left eigenvectors are normalized so that $\|x\|_2 = \|y\|_2 = 1$ and $\arg(y^* x) = -\theta$, then $\arg(\dot{\lambda}_E(0)) = \theta$ if $E = yx^*|_T$. Indeed, since $y^*(yx^*)|s x = \|yx^*|_S\|_F^2$ (see [NP07, Lemma 3.2]),

$$\dot{\lambda}_E(0) = \frac{y^*(yx^*)|s x}{\|yx^*|_S\|_F} \frac{1}{y^* x} = \frac{\|yx^*|_S\|_F}{y^* x}.$$

Lemma 2.1 allows us to extend the algorithm introduced in [GO11] to the case of Toeplitz structure.

**2.1. Pseudospectral abscissa.** We define

$$\alpha_T^\varepsilon(A) = \max \{ \Re(\lambda) : \lambda \in \Lambda_T^\varepsilon(A) \},$$

the structured pseudospectral abscissa, where

$$\Lambda_T^\varepsilon(A) = \{ \lambda \in \mathbb{C} : \lambda \in \Lambda(A + E) \quad \text{with} \quad E \in \mathcal{S}, \|E\|_F \leq \varepsilon \}.$$

The following algorithm allows us to compute locally rightmost points of the $\varepsilon$-pseudospectrum.

**Algorithm 1.** Let $\lambda_0$ be a rightmost eigenvalue of a given Toeplitz matrix $A \in \mathbb{C}^{n \times n}$ with corresponding right and left eigenvectors $x_0$ and $y_0$ normalized to be RP-compatible. Set $B_1 = A + \varepsilon y_0 x_0^*|_T$.

For $k = 1, 2, \ldots$, let $\lambda_k$ be a rightmost eigenvalue of $B_k$ closest to $\lambda_{k-1}$. (If it is not unique, choose, e.g., the one with the largest imaginary part.) Let $x_k$ and $y_k$ be corresponding right and left eigenvectors normalized to be RP-compatible. Set $B_{k+1} = A + \varepsilon y_k x_k^*|_T$.

We denote by $M_{\varepsilon}$ the iteration map associated to Algorithm 1, i.e.,

$$y_{k-1} x_{k-1}^*|_T \xrightarrow{M_{\varepsilon}} y_k x_k^*|_T.$$ 

By the definition of the algorithm it follows immediately that the fixed points of $M_{\varepsilon}$ are given by the pairs $(x, y)$ solution to

$$\begin{cases} y^*(A + \varepsilon y x^*|_T) = \lambda y^* , \\ (A + \varepsilon y x^*|_T) x = \lambda x , \end{cases}$$

where $\lambda$ is the rightmost eigenvalue of $A + \varepsilon y x^*|_T$. 
2.2. Local maxima and stationary points of Algorithm 1. We are now interested in relating locally rightmost points of the $\varepsilon$-pseudospectrum to stationary points of our algorithm, that is, fixed points of the map $M_\varepsilon$.

We shall obtain this after constructing a continuous dynamics for $E(t)$ in the manifold

$$\mathcal{M} = \{ E \in \mathbb{C}^{n \times n} : E \in \mathcal{S}, \|E\|_F = 1 \}$$

such that $A + \varepsilon E(t)$ evolves in a way to make as large as possible the real part of the derivative of its rightmost eigenvalue, as $t$ increases.

We exploit the following result, whose proof is very similar to that given in [GL12] for real matrices.

**Lemma 2.3.** Let $E \in \mathcal{M}$, let $x, y \in \mathbb{C}^n$ be given nonzero complex vectors, and denote by $(\cdot, \cdot)$ the Frobenius inner product, i.e., $(F, G) = \text{trace}(F^*G)$. The solution of the optimization problem

\begin{equation}
Z_\ast = \arg \max_{Z \in \mathcal{M}, \langle E, Z \rangle = 0} \Re(y^* Z x)
\end{equation}

is given by

\begin{equation}
\mu Z_\ast = y x^* \big|_\mathcal{T} - (E, y x^* \big|_\mathcal{T}) E,
\end{equation}

where $\mu$ is the Frobenius norm of the matrix on the right-hand side.

Let $y(t), x(t)$ be, respectively, the left and right eigenvectors associated to the rightmost eigenvalue $\lambda(t)$ of $A + \varepsilon E(t)$, which we assume to be simple, normalized such that $y(t)^* x(t) > 0$. By Lemma 2.3, the maximal increase of $\Re(\dot{\lambda}(t))$ is carried out by a vector field proportional to

$$y(t)x(t)^* \big|_\mathcal{T} - (E(t), y(t)x(t)^* \big|_\mathcal{T}) E(t),$$

which is the projection of the steepest ascent direction $y(t)x(t)^* \big|_\mathcal{T}$ on the tangent hyperplane to $\mathcal{M}$ at $E(t)$.

As a consequence, the differential equation

\begin{equation}
\begin{cases}
\dot{E}(t) = y(t)x(t)^* \big|_\mathcal{T} - (E(t), y(t)x(t)^* \big|_\mathcal{T}) E(t), \\
E(0) = E_0 \in \mathcal{M},
\end{cases}
\end{equation}

has the properties that $E(t) \in \mathcal{M}$ for all $t$ and $\Re(\dot{\lambda}(t))$ is maximal.

Note that (2.4) is similar to the differential equation analyzed in [GL11, GL12] in the case of standard complex and real pseudospectra.

**Lemma 2.4 (equilibria).** If $x$ and $y$ fulfill (2.1) with $y^* x \neq 0$, then $E = yx^* \big|_\mathcal{T}$ is an equilibrium of (2.4). Conversely, if $E$ is an equilibrium of (2.4), then $E = yx^* \big|_\mathcal{T}$.

**Proof.** The first statement is immediate. For the second one, as $E(t) \in \mathcal{M}$, by the Cauchy–Schwarz inequality we have

\begin{equation}
| \langle E(t), y(t)x(t)^* \big|_\mathcal{T} \rangle | \leq 1,
\end{equation}

where equality occurs if and only if $E(t) = y(t)x(t)^* \big|_\mathcal{T}$. Therefore, $E(t) = \dot{E}$ is an equilibrium of (2.4) if and only if $E = yx^* \big|_\mathcal{T}$, which implies that $x$ and $y$ fulfill (2.1). \( \square \)
Lemma 2.5 (monotonicity property of the flow). The solution of the differential equation (2.4) is characterized by the following property for the rightmost eigenvalue $\lambda(t)$ of $A + \varepsilon E(t)$:

$$\Re(\dot{\lambda}(t)) \geq 0 \quad \forall t \geq 0,$$

where equality occurs if and only if $E(t) = \bar{E}$ with $\bar{E}$ an equilibrium.

Proof. Having assumed $y(t)^*x(t) > 0$ for $t \geq 0$, we have to show that

$$\Re \left( y(t)^* \dot{E}(t)x(t) \right) \geq 0.$$

By multiplying the first equation in (2.4) by $y(t)^*$ from the left and $x(t)$ from the right and observing that

$$\Re \left( y(t)^* \left( y(t)^*x(t) \right) \right) = \|y(t)x(t)^*\|_F,$$

we get the result.

Theorem 2.6. Assume that $\lambda$ is a locally rightmost point of $\Lambda_T^\varepsilon(A)$; then $\lambda \in \Lambda(A + \varepsilon E)$, where $E = yx^*|_\tau$ with $x$ and $y$ satisfying (2.1).

Proof. We argue by contradiction and assume $\lambda \in \Lambda(A + \varepsilon E)$ with $E \neq yx^*|_\tau$.

By Lemma 2.4, this implies that $E$ is not an equilibrium of (2.4). Denoting by $E(t)$ the solution to (2.4) with initial datum $E_0 = E$, by Lemma 2.5 we get the strict inequality $\Re(\dot{\lambda}(0)) > 0$, which would imply that $\lambda$ is not a local maximum.

Remark 2.7. The converse of Theorem 2.6 is not true. In fact, with the exception of the tridiagonal case treated in the appendix, we cannot exclude the existence of other solutions to (2.1) for which $\lambda$ is not a local maximum on $\partial \Lambda_T^\varepsilon(A)$. In particular, the algorithm may converge to such other fixed points in some cases.

2.3. Pseudospectral radius. We define

$$\rho_T^\varepsilon(A) = \max\{ |\lambda| : \lambda \in \Lambda_T^\varepsilon(A) \},$$

the structured pseudospectral radius.

The following simple variant of Algorithm 1, which allows us to compute locally extremal points of the $\varepsilon$-pseudospectrum, with maximal modulus, is based on the following extension of Lemma 2.3.

Lemma 2.8. Let $E \in M$ and let $x, y \in \mathbb{C}^n$ be given nonzero complex vectors and $\lambda \in \mathbb{C}$. The solution of the optimization problem

$$Z_* = \arg \max_{Z \in \mathcal{M}, \langle E, Z \rangle = 0} \Re \left( \bar{\lambda} y^* Z x \right)$$

is given by

$$\mu Z_* = e^{i \arg(\lambda)} y x^*|_\tau - \langle E, e^{i \arg(\lambda)} y x^*|_\tau \rangle E,$$

where $\mu$ is the Frobenius norm of the matrix on the right-hand side.

Algorithm 2. Let $\lambda_0$ be an eigenvalue with largest modulus of a Toeplitz matrix $A \in \mathbb{C}^{n \times n}$ with corresponding right and left eigenvectors $x_0$ and $y_0$ normalized to be $RP$-compatible. Set $B_1 = A + \varepsilon e^{i \arg(\lambda_0)} y_0 x_0^*|_\tau$. 
For \( k = 1, 2, \ldots \), let \( \lambda_k \) be an eigenvalue with largest modulus of \( B_k \) closest to \( \lambda_{k-1} \). Let \( x_k \) and \( y_k \) be corresponding right and left eigenvectors normalized to be RP-compatible. Set \( B_{k+1} = A + \varepsilon e^{i\arg(\lambda_k)} y_k x_k^*|_{\mathcal{T}} \).

By the definition of Algorithm 2, it follows immediately that the fixed points of the associated map are given by the pairs \((x, y)\) solution to

\[
\begin{align*}
y^* (A + \varepsilon e^{i\arg(\lambda)} y x^*|_{\mathcal{T}}) &= \lambda y^*, \\
(A + \varepsilon e^{i\arg(\lambda)} y x^*|_{\mathcal{T}}) x &= \lambda x,
\end{align*}
\]

where \( \lambda \) is the eigenvalue of \( A + \varepsilon e^{i\arg(\lambda)} y x^*|_{\mathcal{T}} \) with largest modulus.

We next introduce the differential equation for \( E(t) \in \mathcal{M} \) obtained by replacing \( y(t) \) by \( e^{i\arg(\lambda(t))} y(t) \) in the right-hand side of (2.4). It is straightforward that the analogue of Lemma 2.4 applies in the present context. Moreover, as

\[
\frac{d(|\lambda(t)|^2)}{dt} = 2 \Re(\dot{\lambda}(t) \dot{\lambda}(t)),
\]

a monotonicity property for \(|\lambda(t)|\) holds for such flow. Therefore, arguing analogously to the proof of Theorem 2.6, we can conclude that every point \( \lambda \in \Lambda^T(A) \) which locally maximizes \(|\lambda|\) has to be a stationary point of Algorithm 2.

### 2.4. Rotated computation.

In order to partially compute the boundary of the pseudospectrum, we can apply Algorithm 1 to a rotated matrix \( e^{-i\theta} A \) to reach the boundary along the direction with angle \( \theta \). Indeed we are able to compute rightmost points of the rotated pseudospectrum by our algorithm and draw them after a rotation back. This allows us to represent some convex sections of the boundary and draw a set which includes the pseudospectrum. (See section 4 for some illustrative examples.)

### 2.5. Boundary of the \( \varepsilon \)-pseudospectrum.

We are interested in investigating whether the number of rightmost points of the \( \varepsilon \)-pseudospectrum has to be finite. This is done rigorously in the appendix for the case of tridiagonal Toeplitz matrices. Concerning the general case, we are able to show that such a number is finite at least when \( \varepsilon \) is small enough. Indeed, Theorem 3.6 shows that if \( \varepsilon \) is sufficiently small, then the algorithm locally converges to its fixed points, whence they have to be isolated.

### 3. Local error analysis.

The aim of this section is to provide for a Toeplitz matrix \( A \) with an arbitrary banded structure, a local error analysis of Algorithm 1 close to a simple locally rightmost point.

The analysis presents similarities but also some additional difficulties with respect to that given in [GO11] for the unstructured case. In order to proceed we recall the definition of group inverse, which we need in the analysis.

**Definition 3.1.** The group inverse of a matrix \( C \), denoted \( C^# \), is the unique matrix \( G \) satisfying \( CG = GC \), \( GCG = G \), and \( CGC = C \).

The following result is important for the error analysis of Algorithm 1.

**Theorem 3.2.** Suppose that \((x, y)\) is a boundary fixed point of the map \( M_\varepsilon \) corresponding to a simple rightmost eigenvalue \( \lambda \) of \( B = A + \varepsilon y x^*|_{\mathcal{T}} \). Let the sequence \( B_k \) and \( L_k = y_k x_k^*|_{\mathcal{T}} \) be defined as in the map \( M_\varepsilon \) and let \( L = y x^*|_{\mathcal{T}} \) be the fixed point. Set

\[
F_k = \varepsilon (y_{k-1} x_{k-1}^* - y x^*),
\]

\[
E_k = \varepsilon (y_{k-1} x_{k-1}^*|_{\mathcal{T}} - y x^*|_{\mathcal{T}}) = \varepsilon (L_{k-1} - L)
\]
for \( k = 1, 2, \ldots \) and let \( \delta_k = \|E_k\|_2 \). Then we have
\[
F_{k+1} = \varepsilon \left( \mathcal{R}(\beta_k + \gamma_k)L - LE_k^*G^* - G^*E_k^*L \right) + O(\delta_k^2),
\]
where
\[
G = (B - \lambda I)^#, \quad \beta_k = x^*GE_k x, \quad \gamma_k = y^*E_k G y.
\]

**Proof.** We omit the proof, which is similar to that of Guglielmi and Overton (see [GO11, Theorem 5.3]) and is based on the perturbation analysis of eigenvectors in [MS88].

Nevertheless there is an important difference with respect to the result given in [GO11]. Observe that in the result given there we have \( E_{k+1} \) replacing \( F_{k+1} \) in the left-hand side of (3.1) so that it is possible to study directly the map \( E_{k+1}(E_k) \).

In the present case, however, since \( E_k \neq E_k|_\tau \), the result (3.1) has to be further elaborated. In particular, in order to proceed, we need the following lemma.

**Lemma 3.3.** Let \( x, y \) and \( \hat{x}, \hat{y} \) be RP-compatible and
\[
\|y x^* - \hat{y} \hat{x}^*\|_2 \leq \delta
\]
with \( \delta \) sufficiently small. Then
\[
\|y x^*|_\tau - \hat{y} \hat{x}^*|_\tau\|_2 \leq \mathrm{const} \, \|y x^* - \hat{y} \hat{x}^*\|_2,
\]
where \( \mathrm{const} \) is a constant not depending on \( \delta \).

**Proof.** We recall the bounds
\[
\|M\|_2 \leq \|M\|_F \leq \sqrt{n}\|M\|_2, \quad \|M\|_S\|_F \leq \|M\|_F,
\]
valid for any \( M \in \mathbb{C}^{n \times n} \), and observe that by neglecting the off-diagonal terms’ contribution to the Frobenius norm,
\[
\|yx^*|_S\|_F \geq \frac{y^*x}{\sqrt{n}}.
\]
Therefore,
\[
\|\hat{y} \hat{x}^*|_S\|_F \geq \|yx^*|_S\|_F - \|yx^* - \hat{y} \hat{x}^*|_S\|_F \geq \frac{y^*x}{\sqrt{n}} - \|yx^* - \hat{y} \hat{x}^*\|_F
\]
\[
\geq \frac{y^*x}{\sqrt{n}} - \delta \sqrt{n}.
\]
Moreover,
\[
\|yx^*|_\tau - \hat{y} \hat{x}^*|_\tau\|_2 = \left\| \frac{yx^*|_S}{\|yx^*\|_F} - \frac{\hat{y} \hat{x}^*|_S}{\|\hat{y} \hat{x}^*\|_F} \right\|_2
\]
\[
= \left\| \frac{yx^*|_S - \hat{y} \hat{x}^*|_S}{\|yx^*\|_F} + \hat{y} \hat{x}^*|_S \left( \frac{1}{\|yx^*|_F} - \frac{1}{\|\hat{y} \hat{x}^*\|_F} \right) \right\|_2
\]
\[
\leq \left\| \frac{yx^*|_S - \hat{y} \hat{x}^*|_S}{\|yx^*\|_F} \right\|_2 + \left\| \frac{\hat{y} \hat{x}^*|_S}{\|yx^*|_F} \right\|_2 \left\| \frac{\hat{y} \hat{x}^*|_S}{\|yx^*|_F} \right\| \left\| \frac{yx^*|_S}{\|yx^*|_F} \right\|_F
\]
\[
\leq \left\| \frac{yx^*|_S - \hat{y} \hat{x}^*|_S}{\|yx^*\|_F} \right\|_2 + \left\| \frac{\hat{y} \hat{x}^*|_S}{\|yx^*|_F} \right\|_2 \left\| \frac{yx^*|_S - \hat{y} \hat{x}^*|_S}{\|yx^*|_F} \right\| \left\| \frac{yx^*|_S}{\|yx^*|_F} \right\|_F
\]
\[
\leq \left(1 + \frac{\|yx^*|_S^2}{\|yx^*|_F^2} \right) \frac{\|yx^*|_S^2}{\|yx^*|_F^2} \left(1 + \frac{\|yx^*|_S^2}{\|yx^*|_F^2} \right).
\]
By (3.2) it follows that
\[
\frac{\|\hat{y} \hat{x}^* S\|_2}{\|\hat{y} \hat{x}^* S\|_F} \leq 1, \quad \|yx^* S - \hat{y} \hat{x}^* S\|_F \leq \|yx^* - \hat{y} \hat{x}^*\|_F \leq \sqrt{n} \|yx^* - \hat{y} \hat{x}^*\|_2.
\]

Therefore,
\[
\|yx^*|_T - \hat{y} \hat{x}^*|_T\|_2 \leq \frac{2\sqrt{n}}{\|yx^*|_S\|_F} \|yx^* - \hat{y} \hat{x}^*\|_2 \leq \frac{2n}{y^*x} \|yx^* - \hat{y} \hat{x}^*\|_2. \quad \square
\]

The following theorem establishes a useful formula for the group inverse of a singular matrix C.

**Theorem 3.4** (see [GGO12]). Suppose that C is singular and has a simple zero eigenvalue. Let the two vectors \( x \in \ker(C) \) and \( y \in \ker(C^*) \) be normalized so that \( \|x\|_2 = \|y\|_2 = 1 \). Let \( C = USV^* \), where \( S = \text{diag}(s_1, \ldots, s_{n-1}, 0) \), i.e., \( s_i = \sigma_i(C) \), \( i = 1 : n - 1 \), and \( U, V \) are unitary matrices. Then
\[
G = C^\# = (I - wy^*) V \Xi U^* (I - wy^*),
\]
where \( w = qx \) and \( q = 1/y^*x \), so \( y^*w = 1 \) and
\[
\Xi = \text{diag}(s_1^{-1}, \ldots, s_{n-1}^{-1}, 0).
\]

Moreover, the following estimate holds:
\[
(3.3) \quad \|G\|_2 \leq \frac{\varrho^2}{\sigma_{n-1}(C)}.
\]

We can now establish a sufficient condition for local convergence.

**Theorem 3.5.** Suppose that \( (x, y) \) is a boundary fixed point of the map \( M_\varepsilon \) corresponding to a simple rightmost eigenvalue \( \lambda \) of \( B = A + \varepsilon yx^*|_T \). Define
\[
(3.4) \quad r = \frac{4}{\sigma_{n-1}(A + \varepsilon yx^*|_T - \lambda I)} \sum_{j \leq \varepsilon \varrho^2 \varepsilon} \text{const}, \quad \text{where } q = \frac{1}{y^*x}
\]
and (const) is the constant in Lemma 3.3. Then, if \( r < 1 \), and if \( \delta_k = \|E_k\|_2 \) is sufficiently small, then \( \lim_{j \to \infty} \lambda_{k+j} = \lambda \). Convergence is at least linear with a rate less than or equal to \( r \).

**Proof.** Assume \( \delta_k \) is sufficiently small. According to Theorem 3.2, for studying local convergence we consider the map \( N_\varepsilon \) defined by
\[
F_{k+1} = N_\varepsilon(F_k) = \varepsilon \left( \Re(\delta_k + \gamma_k)L - L E_k^* G^* - G^* E_k^* L \right)
\]
with \( \delta_k = x^* G \beta_k x \) and \( \gamma_k = y^* E_k y \), where \( E_k \) depends on \( F_k \).

Since \( \|L\|_2 = 1 \), Lemma 3.3 and Theorem 3.4 yield
\[
\|N_\varepsilon(F_k)\|_2 \leq \frac{4\varrho^2 \varepsilon}{\sigma_{n-1}(A + \varepsilon yx^*|_T - \lambda I)} \|E_k\|_2 \leq r\|F_k\|_2.
\]

The convergence of \( F_k \) clearly implies that of \( E_k \).

So, if \( r < 1 \), the map \( N_\varepsilon \) is a contraction, and the sequence \( \{\lambda_k\} \) converges to \( \lambda \) with a linear rate bounded above by \( r \). \( \square \)
Theorem 3.6. Assume that \( \lambda(0) \) is a simple rightmost eigenvalue of \( A \) and \( \lambda(\varepsilon) \) is a path of boundary fixed points of the map \( M_\varepsilon \). Then the bound \( r(\varepsilon) \) in (3.4) is such that \( \lim_{\varepsilon \to 0} r(\varepsilon) = 0 \).

Proof. We put in evidence the dependence of the fixed point on \( \varepsilon \) and denote by \( x(\varepsilon) \) and \( y(\varepsilon) \) the eigenvectors associated to the fixed point \( \lambda(\varepsilon) \).

Observe that \( \lim_{\varepsilon \to 0} \varrho(\varepsilon) = 1/y(0)^\ast x(0) > 0 \) and

\[
\lim_{\varepsilon \to 0} \sigma_{n-1} \left( A + \varepsilon y(\varepsilon)^\ast |_T - \lambda(\varepsilon)I \right) = \sigma_{n-1} (A - \lambda(0)I) > 0
\]

by the simplicity assumption for the rightmost eigenvalue of \( A \), which extends to \( \lambda(\varepsilon) \) for sufficiently small \( \varepsilon \) by a continuity argument.

Theorem 3.6 implies that for sufficiently small \( \varepsilon \), Algorithm 1 converges at least linearly with a rate \( O(\varepsilon) \). Numerical experiments show that the method converges also for large values of \( \varepsilon \).

Remark 3.7. We notice that the parameter \( \varrho \) appearing in Theorem 3.6 is the condition number of \( \lambda \). In the case of tridiagonal Toeplitz matrices the condition number is computed in [NPR11, eq. (20)]. Hereafter in the paper, we shall use the notation \( T(s,d,t) \) for the tridiagonal Toeplitz matrix with \( s,d,t \) as subdiagonal, diagonal, and superdiagonal entries, respectively. Calling

\[
m = \min \left\{ |s|, |t| \right\} / \max \left\{ |s|, |t| \right\},
\]

one deduces that \( \varrho \to 1 \) as \( m \to 1 \) and \( \varrho \to \infty \) as \( m \to 0 \). Moreover, in the latter case the asymptotics

\[
\varrho \approx 1 - \cos \left( \frac{2h\pi}{n+1} \right) \left( \frac{1}{m} \right)^{\frac{n}{n-1}}
\]

holds with \( h = 1 \) or \( h = n \) depending on the displacement of the spectrum of \( A \).

4. Examples. We provide here some illustrative examples.

Example 1. We consider the \( 12 \times 12 \) tridiagonal Toeplitz matrix

\[
A = T(s,d,t), \quad s = \frac{-1 + i}{10}, \quad d = \frac{-3 + 4i}{10}, \quad t = 2 + i.
\]

In Figure 4.1, left, we plot the unstructured \( \varepsilon \)-pseudospectrum and the computed section of the structured \( \varepsilon \)-pseudospectrum for \( \varepsilon = 0.5 \), together with its convex hull. (Red points are boundary points computed by the rotated variant of Algorithm 1 discussed in section 2.4; thin red lines give the convex hull.)

The behavior of Algorithm 1 is shown in Table 4.1 and in Figure 4.2, right, where the iterates rapidly converge to the rightmost point. The estimated linear convergence rate is \( r \approx 0.085 \). In Figure 4.1, middle, we plot a further section of \( \Lambda^T_\varepsilon (A) \) in the following way. Using the simple property

\[
\Lambda^T_\varepsilon (A - \mu I) = \Lambda^T_\varepsilon (A) - \mu = \{ z = \lambda - \mu : \lambda \in \Lambda^T_\varepsilon (A) \},
\]

we are able to use a variant of Algorithm 2 which converges to the point of minimal modulus of the \( \varepsilon \)-pseudospectrum of \( A - \mu I, \mu \in \mathbb{C} \) being a point external to \( \Lambda^T_\varepsilon (A) \). The obtained value \( \lambda_{\text{min}}(\mu) \) is then shifted by \( \mu \) and gives a point on the boundary of the \( \varepsilon \)-pseudospectrum.
Example 2. We consider the $30 \times 30$ pentadiagonal matrix

$$A = T(0, 10/19, 0, 0, 10/19)$$

generated by the symbol $\alpha(t) = \frac{10}{19}(t + t^{-2})$.

In Figure 4.3 we show both the structured pseudospectrum and sections of the unstructured pseudospectrum. (The drawn blue section is not continuous, because the
Fig. 4.3. Left: the unstructured $\varepsilon$-pseudospectrum of matrix (4.2) is drawn with the section of the computed structured $\varepsilon$-pseudospectrum for $\varepsilon = 0.5$. Right: the structured $\varepsilon$-pseudospectrum with the circle of radius $\rho_\varepsilon^T(A)$.

In Figure 4.4 we zoom the iterates generated by Algorithms 1 and 2, respectively, to a rightmost point $\lambda_1$, that is, $\Re(\lambda_1) = \alpha_\varepsilon^T(A)$, and to a point $\lambda_2$ of maximal modulus, that is, $|\lambda_2| = \rho_\varepsilon^T(A)$.

4.1. Extension to Hankel matrices. An extension to Hankel matrices is straightforward. We provide here an illustrative example, complementing Example 1.

Example 3. We consider the antitridiagonal $12 \times 12$ Hankel matrix

(4.3) \[ A = H(s, d, t), \quad s = \frac{-1 + i}{10}, \quad d = \frac{-3 + 4i}{10}, \quad t = 2 + i, \]

that is, the matrix with elements

\[ a_{i,n+1-i} = d, \quad i = 1, \ldots, n, \]
\[ a_{i,n-i} = s, \quad i = 1, \ldots, n - 1, \]
\[ a_{i+1,n+1-i} = t, \quad i = 1, \ldots, n - 1. \]

The red section of the structured pseudospectrum is computed by the rotated implementation of the basic algorithm to compute the pseudospectral abscissa. The
blue section is computed by a variant of the method to compute the pseudospectral radius.

**Appendix. The tridiagonal case.** In this section we study the system (2.1) in the simpler case of tridiagonal Toeplitz matrices. Recall that the spectrum of $T = T(s, d, t)$ is given by

$$
\lambda_h(T) = d + 2 \sqrt{|st|} e^{i (\arg s + \arg t)/2} \cos \frac{h \pi}{n+1}, \quad h = 1 : n .
$$

Let $A = T(\sigma_0, \delta_0, \tau_0) \in \mathbb{C}^{n \times n}$ with

$$
(A.1) \quad \sigma_0 \tau_0 \neq 0, \quad \frac{\arg \sigma_0 + \arg \tau_0}{2} \neq \pm \frac{\pi}{2} .
$$

We remark that the assumptions (A.1) guarantee $A$ has $n$ simple eigenvalues lying on a nonvertical segment. If a pair $(x, y)$ is solution to (2.1), then they are the right and left eigenvectors of a rightmost eigenvalue of a tridiagonal Toeplitz matrix, say, $T(\sigma, \delta, \tau)$. Therefore, by setting

$$
\sigma = |\sigma| e^{i \alpha}, \quad \tau = |\tau| e^{i \beta} ,
$$

the vectors $x, y$ have components

$$
(A.2) \quad \begin{cases}
    x_k = e^{i k (\alpha - \beta)/2} \left( \frac{\sigma}{\tau} \right)^{k/2} \sin \frac{k \pi r}{n+1}, & k = 1 : n , \\
    y_k = e^{i k (\alpha - \beta)/2} \left( \frac{\tau}{\sigma} \right)^{k/2} \sin \frac{k \pi r}{n+1},
\end{cases}
$$

where, by (A.1), $r = 1$ or $r = n$ depending on which one between the extremal eigenvalues of $A$ has the largest real part; indeed we tacitly assume the parameter $\varepsilon$ to be small enough to not affect this property. Therefore,

$$
(A.3) \quad (yx^*)_{k, h} = e^{i (k - h) (\alpha - \beta)/2} \left( \frac{\tau}{\sigma} \right)^{(k - h)/2} \sin \frac{k \pi r}{n+1} \sin \frac{h \pi r}{n+1}, \quad k, h = 1 : n .
$$

---

**Fig. 4.5.** Left: boundary of the unstructured pseudospectrum (in gray) and section of the boundary of the structured $\varepsilon$-pseudospectrum (in red and blue), with $\varepsilon = 1$, of the Hankel matrix (4.3). Black points are the spectra of 1000 randomly selected structured perturbation matrices of norm 1. Right: zoom.
We notice that \(yx^*|_S = T(\sigma_1, \delta_1, \tau_1)\) with
\[
\sigma_1 = e^{i(\alpha - \beta)/2} \sqrt{\frac{\tau}{\sigma}} \frac{n + 1}{2(n - 1)} \cos \frac{\pi r}{n + 1} ; \quad \delta_1 = \frac{n + 1}{2n},
\]
\[
\tau_1 = e^{i(\beta - \alpha)/2} \sqrt{\frac{\sigma}{\tau}} \frac{n + 1}{2(n - 1)} \cos \frac{\pi r}{n + 1} .
\]
Indeed, the arithmetic means of the diagonal terms in (A.3) can be explicitly computed. More precisely,
\[
\frac{1}{n} \sum_{k=1}^{n} \sin^2 \frac{k\pi r}{n + 1} = \frac{1}{2n} \Re \left\{ \sum_{k=1}^{n} \left[ 1 - \left( e^{\frac{i\pi r}{n + 1}} \right)^k \right] \right\} = \frac{n + 1}{2n},
\]
\[
\frac{1}{n - 1} \sum_{k=1}^{n-1} \sin \frac{k\pi r}{n + 1} \sin \frac{(k + 1)\pi r}{n + 1} = \frac{1}{2} \cos \frac{\pi r}{n + 1} - \frac{1}{2(n - 1)} \Re \left\{ e^{i\frac{\pi r}{n + 1}} \sum_{k=1}^{n-1} (e^{i\frac{\pi r}{n + 1}})^k \right\}
\]
\[
= \frac{1}{2} \cos \frac{\pi r}{n + 1} + \frac{1}{2(n - 1)} \left( \sin \frac{\pi r}{n + 1} \right)^{-1} \sin \frac{2\pi r}{n + 1}
\]
\[
= \frac{n + 1}{2(n - 1)} \cos \frac{\pi r}{n + 1} .
\]
Moreover,
\[
||yx^*||_F = \sqrt{n|\delta_1|^2 + (n - 1)(|\sigma_1|^2 + |\tau_1|^2)}
\]
\[
= \frac{n + 1}{2} \sqrt{\frac{1}{n} + \frac{1}{n - 1} \left( \left| \frac{\sigma}{\tau} \right| + \left| \frac{\tau}{\sigma} \right| \right) \cos^2 \frac{\pi r}{n + 1}}.
\]
In conclusion,
\[
A + \varepsilon yx^*|_T = T(\sigma_0 + \varepsilon \hat{\sigma}, \delta_0 + \varepsilon \hat{\delta}, \tau_0 + \varepsilon \hat{\tau}) ,
\]
where
\[
(A.4) \quad \hat{\sigma} = e^{i(\alpha - \beta)/2} \frac{n - 1}{\sqrt{\frac{\tau}{\sigma}}} \cos \frac{\pi r}{n + 1} \left[ \frac{1}{n} + \frac{1}{n - 1} \left( \left| \frac{\sigma}{\tau} \right| + \left| \frac{\tau}{\sigma} \right| \right) \cos^2 \frac{\pi r}{n + 1} \right]^{-1/2} ,
\]
\[
\hat{\delta} = \frac{1}{n} \left[ \frac{1}{n} + \frac{1}{n - 1} \left( \left| \frac{\sigma}{\tau} \right| + \left| \frac{\tau}{\sigma} \right| \right) \cos^2 \frac{\pi r}{n + 1} \right]^{-1/2} ,
\]
\[
(A.5) \quad \hat{\tau} = e^{i(\beta - \alpha)/2} \frac{n - 1}{\sqrt{\frac{\sigma}{\tau}}} \cos \frac{\pi r}{n + 1} \left[ \frac{1}{n} + \frac{1}{n - 1} \left( \left| \frac{\sigma}{\tau} \right| + \left| \frac{\tau}{\sigma} \right| \right) \cos^2 \frac{\pi r}{n + 1} \right]^{-1/2} .
\]
By the characterization (A.2) of the eigenvectors of a tridiagonal Toeplitz matrix, for \((x, y)\) to be a solution to (2.1), the parameters \(\sigma\) and \(\tau\) have to satisfy the following
relations:
\[
\begin{cases}
\sqrt{|\sigma_0 + \varepsilon \hat{\sigma}|} = \sqrt{|\sigma_0 + \varepsilon \hat{\sigma}|}, \\
\exp \left( \frac{i \arg(\sigma_0 + \varepsilon \hat{\sigma}) - \arg(\tau_0 + \varepsilon \hat{\tau})}{2} \right) = \exp \left( \frac{i \alpha - \beta}{2} \right).
\end{cases}
\]

The system (A.6) can be analyzed by considering the complex equation
\[
\frac{\sigma_0 + \varepsilon \hat{\sigma}}{\tau_0 + \varepsilon \hat{\tau}} = \frac{|\sigma_0 + \varepsilon \hat{\sigma}|}{|\tau_0 + \varepsilon \hat{\tau}|} e^{i(\alpha - \beta)},
\]
whose solutions solve either system (A.6) or
\[
\begin{cases}
\sqrt{|\sigma_0 + \varepsilon \hat{\sigma}|} = \sqrt{|\sigma_0 + \varepsilon \hat{\sigma}|}, \\
\exp \left( \frac{i \arg(\sigma_0 + \varepsilon \hat{\sigma}) - \arg(\tau_0 + \varepsilon \hat{\tau})}{2} \right) = - \exp \left( \frac{i \alpha - \beta}{2} \right).
\end{cases}
\]

Therefore, it suffices to solve (A.7) with the constraint
\[
\exp \left( \frac{i \arg(\sigma_0 + \varepsilon \hat{\sigma}) - \arg(\tau_0 + \varepsilon \hat{\tau})}{2} \right) = \exp \left( \frac{i \alpha - \beta}{2} \right).
\]

Setting
\[
\varrho = \frac{|\sigma_0 + \varepsilon \hat{\sigma}|}{|\tau_0 + \varepsilon \hat{\tau}|}, \quad \varphi = \frac{\alpha - \beta}{2}, \quad a = \frac{\varepsilon \sqrt{n}}{n - 1} \cos \frac{\pi r}{n + 1}, \quad b = \frac{n}{n - 1} \cos^2 \frac{\pi r}{n + 1},
\]
by (A.4) and (A.5) we have
\[
\varepsilon \hat{\sigma} = \frac{ae^{i\varphi}}{\sqrt{\varrho + b(1 + \varrho^2)}}, \quad \varepsilon \hat{\tau} = \frac{ae^{-i\varphi}}{\sqrt{\varrho + b(1 + \varrho^2)}}.
\]

Substituting in (A.7), after some easy computations the latter reads
\[
\varrho \tau_0 e^{2i\varphi} + \frac{a(\varrho^2 - 1)}{\varrho + b(1 + \varrho^2)} e^{i\varphi} - \sigma_0 = 0.
\]

Hence, defining
\[
G(\varrho) = \frac{a(1 - \varrho^2)}{2 \varrho \sqrt{\varrho + b(1 + \varrho^2)}},
\]
(A.11) becomes
\[
e^{i\varphi} = \frac{G(\varrho)}{\tau_0} \pm \frac{1}{\tau_0} \sqrt{G(\varrho)^2 + \frac{\sigma_0 \tau_0}{\varrho}}.
\]

**Theorem A.1.** For each $\sigma_0, \tau_0 \in \mathbb{C}$ satisfying (A.1) and for each $n > 1$ there exists $\varepsilon_n = \varepsilon_n(\sigma_0, \tau_0)$ such that for any $\varepsilon \in [0, \varepsilon_n)$ there is a unique pair $(\varrho_*, \varphi_*)$
We study separately the following two cases.

Moreover, the following hold.

(A.14) \[ |a| \leq \frac{2\varepsilon}{\sqrt{n}}, \quad \frac{1}{2} \leq b \leq \frac{3}{2} \quad \forall n > 1, \]

we can analyze (A.13) under the assumptions that \( b \in \left[ \frac{1}{2}, \frac{3}{2} \right] \) and \( |a| \) is small enough. Theorem A.1 proves to be an easy corollary of the proposition below.

**Proposition A.2.** Let

(A.15) \[ F_\pm(\varrho) = \frac{G(\varrho)}{\tau_0} \pm \frac{1}{\tau_0} \sqrt{G(\varrho)^2 + \frac{\sigma_0 \tau_0}{\varrho}}. \]

For each \( \sigma_0, \tau_0 \in \mathbb{C} \) there exists \( a_0 > 0 \) such that for any \( |a| \in (0, a_0) \) and \( b \in \left[ \frac{1}{2}, \frac{3}{2} \right] \) there is a unique positive \( \varrho^+ \) (resp., \( \varrho^- \)) such that \( |F_+(\varrho^+)| = 1 \) (resp., \( |F_-(\varrho^-)| = 1 \)). Moreover, the following hold.

1. If \( \Re(\sigma_0 \tau_0) > -|\sigma_0||\tau_0| \), then

(A.16) \[ r = 1 \quad \Rightarrow \quad \begin{cases} \varrho_- < \varrho_+ < 1 & \text{if } |\sigma_0| < |\tau_0|, \\ 1 < \varrho_+ < \varrho_- & \text{if } |\sigma_0| > |\tau_0|. \end{cases} \]

(A.17) \[ r = n \quad \Rightarrow \quad \begin{cases} \varrho_+ < \varrho_- < 1 & \text{if } |\sigma_0| < |\tau_0|, \\ 1 < \varrho_- < \varrho_+ & \text{if } |\sigma_0| > |\tau_0|. \end{cases} \]

Finally, if \( |\sigma_0| = |\tau_0| = 1 \), then \( \varrho_+ = \varrho_- = 1 \).

2. If \( \Re(\sigma_0 \tau_0) = -|\sigma_0||\tau_0| \), then \( \varrho_+ = \varrho_- = \frac{|\sigma_0|}{|\tau_0|} \).

**Proof.** The square root appearing in (A.15) is intended to be the principal one. Otherwise stated, if \( z = r e^{i\theta} \) with \( \theta \in (-\pi, \pi] \), then \( \sqrt{z} = \sqrt{r} e^{i\theta/2} \). In particular, this implies

\[ \sqrt{z} = \begin{cases} \sqrt{z} & \text{if } \Re(z) \neq -|z|, \\ -i\sqrt{|z|} & \text{if } \Re(z) = -|z|. \end{cases} \]

We study separately the following two cases.

Case (1): \( \Re(\sigma_0 \tau_0) \neq -|\sigma_0||\tau_0| \). We have

\[ |F_\pm(\varrho)|^2 = \frac{1}{|\tau_0|^2} \left( G(\varrho) \pm \sqrt{G(\varrho)^2 + \frac{\sigma_0 \tau_0}{\varrho}} \right) \left( G(\varrho) \pm \sqrt{G(\varrho)^2 + \frac{\sigma_0 \tau_0}{\varrho}} \right), \]

where the equation \( |F_\pm(\varrho)|^2 = 1 \) reads

\[ G(\varrho) \pm \sqrt{G(\varrho)^2 + \frac{\sigma_0 \tau_0}{\varrho}} = |\tau_0|^2 \left( G(\varrho) \pm \sqrt{G(\varrho)^2 + \frac{\sigma_0 \tau_0}{\varrho}} \right)^{-1}, \]

which can be recast into the form

\[ G(\varrho) \pm \sqrt{G(\varrho)^2 + \frac{\sigma_0 \tau_0}{\varrho}} = \frac{\varrho \tau_0}{\sigma_0} \left( G(\varrho) \pm \sqrt{G(\varrho)^2 + \frac{\sigma_0 \tau_0}{\varrho}} \right), \]
that is,
\[(A.18) \quad \left(1+\frac{\sigma_0}{\sigma}\right) G(\varrho) = \pm \frac{\sigma_0}{\sigma_0} \sqrt{G(\varrho)^2 + \frac{\sigma_0}{\sigma}} \mp \sqrt{G(\varrho)^2 + \frac{\sigma_0}{\sigma}} .\]

Rationalizing, after some computations we obtain
\[4\varrho^2 G(\varrho)^2 = \frac{(|\tau|^2 \varrho^2 - |\sigma|^2)^2}{|\tau|^2 \varrho^2 + 2\Re(\sigma_0 \tau_0) \varrho + |\sigma|^2} .\]

Plugging in the definition (A.12) of \(G(\varrho)\), we finally get
\[(A.19) \quad \frac{a^2(1-\varrho^2)^2}{b\varrho^2 + \varrho + b} = \frac{(|\tau_0|^2 \varrho^2 - |\sigma|^2)^2}{|\tau|^2 \varrho^2 + 2\Re(\sigma_0 \tau_0) \varrho + |\sigma|^2} ,\]

Let us consider the functions appearing in (A.19), that is,
\[
f_1(\varrho) = \frac{a^2(1-\varrho^2)^2}{b\varrho^2 + \varrho + b} , \quad f_2(\varrho) = \frac{(|\tau_0|^2 \varrho^2 - |\sigma|^2)^2}{|\tau|^2 \varrho^2 + 2\Re(\sigma_0 \tau_0) \varrho + |\sigma|^2} ,
\]

restricted on the domain of interest \(\{\varrho > 0\}\). We claim that if \(|\sigma| \neq |\tau_0|\), then the corresponding graphs intersect each other in two points whose abscissae \(\varrho_1, \varrho_2\) are such that
\[
\begin{aligned}
\varrho_1 < \frac{|\sigma_0|}{|\tau_0|} < \varrho_2 & \text{ if } |\sigma_0| < |\tau_0| , \\
1 < \varrho_1 < \frac{|\sigma_0|}{|\tau_0|} < \varrho_2 & \text{ if } |\sigma_0| > |\tau_0| ,
\end{aligned}
\]

while if \(|\sigma_0| = |\tau_0|\) such graphs intersect each other solely in the point \((1, 0)\).

We start by noticing that \(f_1(\varrho)\) reaches its minimum value uniquely in \(\varrho = 1\). More precisely, it decreases in \([0, 1]\), from \(f_1(0) = a^2/b\) to \(f_1(1) = 0\), and increases in \([1, +\infty)\), diverging with \(\varrho^{-2} f_1(\varrho) \to a^2/b\) as \(\varrho \to \infty\).

Concerning \(f_2(\varrho)\), if \(|a|\) is small enough, then
\[(A.21) \quad f_2(0) = |\sigma_0|^2 > f_1(0) , \quad \lim_{\varrho \to +\infty} \varrho^{-2} f_2(\varrho) = |\tau_0|^2 > \lim_{\varrho \to +\infty} \varrho^{-2} f_1(\varrho) .\]

We now distinguish the cases \(\Im(\sigma_0 \tau_0) = 0\) and \(\Im(\sigma_0 \tau_0) \neq 0\). In the first case \(\Re(\sigma_0 \tau_0) = |\sigma_0||\tau_0|\), whence
\[
f_2(\varrho) = \left(|\tau_0|\varrho - |\sigma_0|\right)^2 ,
\]

which is the law of a parabola with vertex in \(\left(\frac{\sigma_0}{\tau_0}, 0\right)\). Therefore, by (A.21), the claim is straightforward. In the second case we have
\[
f_2(\varrho) = \frac{(|\tau_0|^2 \varrho^2 - |\sigma|^2)^2}{(|\tau_0|^2 \varrho^2 + 2\Re(\sigma_0 \tau_0)/|\tau_0|)^2 + \kappa^2}
\]

with \(\kappa^2 = |\sigma_0|^2 - \Re(\sigma_0 \tau_0)^2/|\tau_0| > 0\). As a function on the whole line, \(f_2\) has two absolute minima for \(\varrho = \pm \frac{\sigma_0}{|\tau_0|}\). In the interval \(|a| < |\sigma_0|/|\tau_0|\) there can be one or three local extrema. But in any cases, by choosing \(|a|\) small enough, the claim is easily verified.
We are left with showing that \( \varrho_1 \) and \( \varrho_2 \) solve \( |F_+(\varrho_1)| = |F_- (\varrho_2)| = 1 \) or \( |F_+ (\varrho_2)| = |F_- (\varrho_1)| = 1 \), thus proving the proposition. (In the case \( |\sigma_0| = |\tau_0| \) the identity \( F_{\pm} (1) = 1 \) is immediate.) Taking the real part of (A.18) with \( \varrho = \varrho_1 \) we have

\[
(A.22) \quad (1 + \gamma_1 \cos \theta) G(\varrho_1) = \pm (\gamma_1 \cos \theta \Re (H_1) - \gamma_1 \sin \theta \Im (H_1) - \Re (H_1)) ,
\]

where

\[
\gamma_1 = \varrho_1 \frac{|\tau_0|}{|\sigma_0|}, \quad \theta = \arg \sigma_0 + \arg \tau_0, \quad H_1 = \sqrt{G(\varrho_1)^2 + \frac{\sigma_0 \tau_0}{\varrho_1}} .
\]

By (A.12), (A.20), and recalling that by the definition of \( \gamma \)

\[
\sin \theta \quad \text{we have}
\]

\[
(\text{A.23}) \quad (1 + \gamma_1 \cos \theta) G(\varrho_1) = \pm (\gamma_1 \cos \theta \Re (H_1) - \gamma_1 \sin \theta \Im (H_1) - \Re (H_1)) ,
\]

where

\[
\gamma_1 = \varrho_1 \frac{|\tau_0|}{|\sigma_0|}, \quad \theta = \arg \sigma_0 + \arg \tau_0, \quad H_1 = \sqrt{G(\varrho_1)^2 + \frac{\sigma_0 \tau_0}{\varrho_1}} .
\]

By (A.12), (A.20), and recalling that by the definition of \( a \) (see (A.9)) we have \( a > 0 \) if \( r = 1 \) and \( a < 0 \) if \( r = n \), we conclude that

\[
G(\varrho_1) \begin{cases} 
> 0 & \text{if } r = 1 \text{ and } |\sigma_0| < |\tau_0| \text{ or } r = n \text{ and } |\sigma_0| > |\tau_0| , \\
< 0 & \text{if } r = 1 \text{ and } |\sigma_0| > |\tau_0| \text{ or } r = n \text{ and } |\sigma_0| < |\tau_0| . 
\end{cases}
\]

Moreover, as \( \gamma_1 < 1 \), the left-hand side of (A.22) has the same sign as \( G(\varrho_1) \). On the other hand, since \( \Re (H_1^2) \) has the same sign as \( \sin \theta \), we have \( \Re (H_1) > 0 \) and \( \sin \theta \Im (H_1) \geq 0 \), so that

\[
\gamma_1 \cos \theta \Re (H_1) - \gamma_1 \sin \theta \Im (H_1) - \Re (H_1) \leq (\gamma_1 \cos \theta - 1) \Re (H_1) < 0 .
\]

By (A.22) it follows that

(i) \( \varrho_1 \) cannot be a solution of \( |F_+ (\varrho)| = 1 \) if \( r = 1 \) and \( |\sigma_0| < |\tau_0| \) or \( r = n \) and \( |\sigma_0| > |\tau_0| \), and therefore \( \varrho_- = \varrho_1 \) and \( \varrho_+ = \varrho_2 \) in these cases;

(ii) \( \varrho_1 \) cannot be a solution of \( |F_- (\varrho)| = 1 \) if \( r = 1 \) and \( |\sigma_0| > |\tau_0| \) or \( r = n \) and \( |\sigma_0| < |\tau_0| \), and therefore \( \varrho_- = \varrho_2 \) and \( \varrho_+ = \varrho_1 \) in these cases.

Case (2): \( \Re (\sigma_0 \tau_0) = -|\sigma_0||\tau_0| \). We have

\[
F_\pm (\varrho) = \frac{1}{|\tau_0|} \left( G(\varrho) \pm \sqrt{G(\varrho)^2 - \frac{|\sigma_0||\tau_0|}{\varrho}} \right) ,
\]

where

\[
\sqrt{G(\varrho)^2 - \frac{|\sigma_0||\tau_0|}{\varrho}} = \begin{cases} 
\sqrt{|G(\varrho)^2 - \frac{|\sigma_0||\tau_0|}{\varrho}} & \text{if } G(\varrho)^2 \leq \frac{|\sigma_0||\tau_0|}{\varrho} , \\
\sqrt{|G(\varrho)^2 - \frac{|\sigma_0||\tau_0|}{\varrho}} & \text{if } G(\varrho)^2 > \frac{|\sigma_0||\tau_0|}{\varrho} .
\end{cases}
\]

Therefore

\[
(A.23) \quad |F_\pm (\varrho)|^2 = \frac{|\sigma_0|}{|\tau_0|} \chi_{(G(\varrho)^2 < \frac{|\sigma_0||\tau_0|}} + K_\pm (\varrho) \chi_{(G(\varrho)^2 > \frac{|\sigma_0||\tau_0|}} ,
\]

where \( \chi \) denotes the characteristic set function, and

\[
K_\pm (\varrho) = \frac{1}{|\tau_0|^2} \left( 2G(\varrho)^2 - \frac{|\sigma_0||\tau_0|}{\varrho} \pm 2G(\varrho) \sqrt{G(\varrho)^2 - \frac{|\sigma_0||\tau_0|}{\varrho}} \right) .
\]
We now observe that the equation $K_\pm(q) = 1$ can be written in the form
\[ |\tau_0|^2 - 2G(q)^2 + \frac{\sigma_0|\tau_0|}{q} = \pm 2G(q)\sqrt{G(q)^2 - \frac{\sigma_0|\tau_0|}{q}}, \]
from which, recalling the definition of $f_1(q)$, we get
\[ f_1(q) = (|\tau_0|q + |\sigma_0|)^2. \]

By the previous qualitative analysis of $f_1(q)$ we easily deduce that such an equation does not have positive solutions for $|a|$ small. On the other hand, if $|a|$ is sufficiently small, then the condition $G(q)^2 \leq |\sigma_0||\tau_0|$ is fulfilled by $q = \frac{|\sigma_0|}{|\tau_0|}$, whence by (A.23) we get the result.

Remark A.3. It is worth noticing that even the case $\Re(\sigma_0\tau_0) = |\sigma_0||\tau_0|$ is quite explicit. Indeed, since $a > 0$ and $4q^2G(q)^2 = (|\tau_0|q - |\sigma_0|)^2$,
\[ G(q) = \begin{cases} \frac{|\sigma_0| - |\tau_0||\tau_1|}{2\theta_1}, & \text{if } i = 1 \text{ and } |\sigma_0| < |\tau_0| \text{ or } i = 2 \text{ and } |\sigma_0| > |\tau_0|, \\ \frac{|\tau_0||\tau_1| - |\sigma_0|}{2\theta_1}, & \text{if } i = 1 \text{ and } |\sigma_0| > |\tau_0| \text{ or } i = 2 \text{ and } |\sigma_0| < |\tau_0|. \end{cases} \]

Plugging these values into (A.15), since $\sigma_0\tau_0 = |\sigma_0||\tau_0|$, in the first case we easily get
\[ F_\pm(q_i) = \frac{|\sigma_0| - |\tau_0||\tau_1|}{2\tau_0\theta_1} \pm \frac{|\sigma_0| + |\tau_0||\tau_1|}{2\tau_0\theta_1} = \begin{cases} \frac{|\sigma_0|}{\tau_0\theta_1}, & \text{if } +, \\ -\frac{|\tau_0|}{\tau_0}, & \text{if } -. \end{cases} \]
while in the second case,
\[ F_\pm(q_i) = \frac{|\tau_0||\tau_1| - |\sigma_0|}{2\tau_0\theta_1} \pm \frac{|\tau_0||\tau_1| + |\sigma_0|}{2\tau_0\theta_1} = \begin{cases} \frac{|\tau_0|}{\tau_0}, & \text{if } +, \\ -\frac{|\sigma_0|}{\tau_0\theta_1}, & \text{if } -. \end{cases} \]

Proof of Theorem A.1. We show that, setting $\varphi_\pm = \arg F_\pm(q_\pm)$, for any $|a|$ small enough we have
\[ \exp\left(\frac{i\arg(\sigma_0 + \varepsilon\hat{\sigma}_\pm) - \arg(\tau_0 + \varepsilon\hat{\tau}_\pm)}{2}\right) = \begin{cases} \pm \exp(i\varphi_\pm) & \text{if } r = 1, \\ \mp \exp(i\varphi_\pm) & \text{if } r = n, \end{cases} \]
where $\hat{\sigma}_\pm, \hat{\tau}_\pm$ are defined by $\hat{\sigma}, \hat{\tau}$ as in (A.10) and evaluated for $(q, \varphi) = (q_\pm, \varphi_\pm)$. By (A.26) the statement of the theorem follows with $(q_+, \varphi_+) = (q_+, \varphi_+)$ if $r = 1$ and $(q_-, \varphi_-) = (q_-, \varphi_-)$ if $r = n$. Moreover, since $|a| \leq 2\varepsilon/\sqrt{n}$, this also shows that the threshold $\varepsilon_n$ can be chosen arbitrarily large increasing the dimension $n$.

To prove (A.26) it suffices to observe that
\[ q_\pm = \frac{|\sigma_0|}{|\tau_0|} + o(1), \quad G(q_\pm) = o(1), \quad \varepsilon\hat{\sigma}_\pm = o(1), \quad \varepsilon\hat{\tau}_\pm = o(1), \]
where $o(1)$ stands for a generic function vanishing as $|a| \to 0$. Therefore, by (A.15) it follows
\[
F_\pm(\varrho_\pm) = \pm \frac{1}{\tau_0} \sqrt{\frac{\sigma_0}{|\sigma_0|} \tau_0 + o(1)} = \pm \exp \left( i \frac{\arg(\sigma_0 \tau_0) - 2 \arg \tau_0}{2} \right) + o(1).
\]
On the other hand,
\[
\exp \left( i \frac{\arg(\sigma_0 + \varepsilon \hat{\sigma}_\pm) - \arg(\tau_0 + \varepsilon \hat{\tau}_\pm)}{2} \right) = \exp \left( i \frac{\arg \sigma_0 - \arg \tau_0}{2} \right) + o(1).
\]
Equation (A.26) now follows by noticing that $\arg(\sigma_0 \tau_0) = \arg \sigma_0 + \arg \tau_0$ if $r = 1$ while $\arg(\sigma_0 \tau_0) = \arg \sigma_0 + \arg \tau_0 \pm 2\pi$ if $r = n$.

**Remark A.4.** In case (2) of Proposition A.2, i.e., $\sigma_0 \tau_0 = -|\sigma_0||\tau_0|$, the second condition in (A.1) is not satisfied and the choice of $r$ cannot be established a priori. However, we observe that in such case $\varrho_+ = \varrho_- = \frac{|\sigma_0|}{|\tau_0|}$,
\[
F_\pm \left( \frac{|\sigma_0|}{|\tau_0|} \right) = \frac{e^{-i \arg \tau_0}}{|\tau_0|} \left( G \left( \frac{|\sigma_0|}{|\tau_0|} \right) \pm i \sqrt{|\tau_0|^2 - G \left( \frac{|\sigma_0|}{|\tau_0|} \right)^2} \right),
\]
and $\arg \sigma_0 - \arg \tau_0 - \pi = -2 \arg \tau_0$. Therefore,
\[
F_\pm \left( \frac{|\sigma_0|}{|\tau_0|} \right) = \frac{e^{i(\arg \sigma_0 - \arg \tau_0)/2}}{|\tau_0|} \left( \pm \sqrt{|\tau_0|^2 - G \left( \frac{|\sigma_0|}{|\tau_0|} \right)^2} + i G \left( \frac{|\sigma_0|}{|\tau_0|} \right) \right).
\]
We conclude that Theorem A.1 holds also in this case and precisely with $(\varrho_+, \varphi_+) = \left( \frac{|\sigma_0|}{|\tau_0|}, \arg F_+ \left( \frac{|\sigma_0|}{|\tau_0|} \right) \right)$.

**Acknowledgment.** We thank the anonymous referees for valuable comments and suggestions.

**REFERENCES**


