CERFACS **Technical Report No. TR/PA/01/84**

Françoise Chaitin-Chatelin and Elisabeth Traviesas

Homotopic perturbation

Unfolding the field of singularities of a matrix by a complex parameter: a global geometric approach

Abstract This work discusses the meaning which emerges from the linear relation between two matrices A and B, done via the complex parameter t such that $A(t) = A + t(B - A)$. Such a meaning is an explanation drawn from an *outside* observation of A and B. One could possibly want to compare this explanation with the meaning which emerges from *inside*. Such a line of research is of *epistemological* importance.
The second idea which emerges from the theory of Inexact Computing that is presented is the importance of the duality of viewpoints which can be taken in mathematics with respect to singularities.
When continuity is the prevalent point of view, a finite set of singularities is *nongeneric* and can be ignored: this is the basis for the elimination of singularities for continuous subharmonic functions. This is a *global* point of view. However, *locally* the singularities do have an influence when computation is inexact, such as in finite precision. Their neighborhood is of *positive Lebesgue measure*. The homotopic unfolding of the resolvent field $z \to (A - zI)^{-1}$ is an extremely rich domain, where we are able to follow the subtle interplay between the continuous and the discrete, through the action of computation.

Key words. Homotopic unfolding, global, local, backward error, pseudospectrum, subharmonicity, synchronicity, knowledge acquisition, deviation, Inexact Computing, critical point

1. Introduction to homotopic perturbation

We briefly survey the use of homotopic perturbation in the context of spectral approximation of linear operators and matrices. This survey covers the work realized or supervised by one of us (FCC) during the period 1970 to date.

In her book **Spectral approximation of Linear Operators** [15], which appeared almost 20 years ago, Françoise Chatelin used the red thread of homotopic perturbations to analyze the numerical approximation, by T_n with finite rank, of a closed linear operator T in a Banach space. She used the identity $T = T_n + (T - T_n)$ and looked at the homotopic family of linear operators $T(t) = T_n + t(T - T_n)$, such that $T(0) = T_n$ and $T(1) = T$, depending linearly on the **complex** parameter $t \in \mathbb{C}$.

The notion that $\{T_n\}$ is a *radial approximation* of T as $n \to \infty$ ([15], p 245) permits the use of powerful analytic perturbation theory to get series expansion of the *exact* eigenelements in terms of the *computed* eigenelements and in terms of the perturbation $T - T_n$. This approach provided an impressively effective tool to present in a unified manner most of a posteriori error bounds known at that time for spectral computations on matrices as well as on linear operators, bounded or closed. The theory is quite general, it requires no assumption of normality or selfadjointness (symmetry).

Françoise Chaitin-Chatelin: Université Toulouse 1 and CERFACS, 42 av. G. Coriolis, 31057 Toulouse cedex - chatelin@cerfacs.fr

Elisabeth Traviesas: CERFACS, 42 av. G. Coriolis, 31057 Toulouse cedex - travies@cerfacs.fr

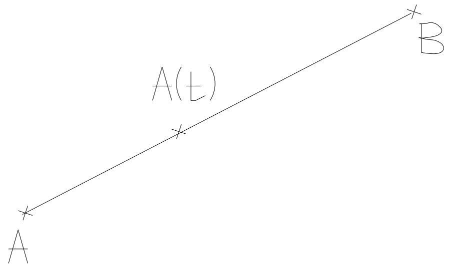


Figure 1.1. *The family* $A(t)$ *in* $\mathbb{R}^{n\times n}$, $0 \leq t \leq 1$

Five years later, in 1988 was published the book **Valeurs propres de matrices** [16] and an enlarged English translation appeared in 1993 [17]. These books were a sequel to the first one and focused only on matrices, that are linear operators on *finite* Banach spaces. However in her use of homotopic perturbation Françoise Chatelin introduced *two key new ideas*:

1. the notion of block reduced resolvent and partial inverse ([17], p. 76-79),
2. the use of Rayleigh-Schrödinger expansions ([17], p. 86-89) rather than Rellich-Kato expansions ([17], p.85-86) for iterative refinement.

The first idea allows to circumvent the discontinuity of the spectral projection associated with a simple computed eigenvalue when it approximates a multiple defective one. The second idea leads to much better behaved algorithms to refine the computed eigenvalues.

Then, in 1996, FCC coauthored with Valérie Frayssé **Lectures on Finite Precision Computations** [6]. This third book concentrates on the effect of roundoff on the behavior of numerical methods which are proved convergent in exact arithmetic. It shows how the powerful notion of backward error analysis can be easily recast in the framework of homotopic perturbation, with the important difference that the perturbation itself is not known in case of *inexact* arithmetic [10]. One has access only to its norm by means of the fundamental idea of the backward error. This leads to the key notion, for numerical methods run in finite precision, of *conditional convergence* with respect to the arithmetic of the computer.

This brief historical account of three decades of research on the approximation of the singularities of the resolvent field: $z \to (A - zI)^{-1}$, where A is a linear operator or a matrix and z is not an eigenvalue of A, puts in full light the ubiquitous role played by homotopic approximation in spectral approximation. It allows to relate information on A and $B = A + E$ by means of the homotopic family $A(t) = A + tE$, $t \in \mathbb{C}$. It has been argued that, according to the classical definition of an homotopic family, the parameter t should be *real* [22]. The adjective "homotopic" of Greek origin means "of the same spatial form", which refers to the fact that the structure of the perturbation tE stays homothetic to the matrix E. In other words, the structure of the perturbation is fixed. This is true whether t is real or complex. The reason to consider t complex is that the treatment of analyticity properties for $(A(t) - zI)^{-1}$, $z \in \mathbb{C}$, is much easier in that context. This reflects the difference between holomorphy in $\mathbb{C}$ and analyticity in $\mathbb{R}$. Going complex simplifies the analysis, as is the case for the fundamental theorem of Algebra.
Note that when A and B are real, there is an easy geometric interpretation of the family $A(t)$, for t real and $0 \leq t \leq 1$, in $\mathbb{R}^{n\times n}$, illustrated in Figure 1.1. Because $A(t) = A + tE = A + t(B - A) = tB + (1-t)A$, it is clear that $A(t)$ represents, for any t in $[0, 1]$, a point on the straight interval joining A to B.

In the report to follow we want to look at the notion of homotopic perturbation from a purely *information-theoretic* point of view: what sort of information on A and $B = A + E$ can be related by means of the homotopic family

$$A(t) = A + tE,\ t \in \mathbb{C}\ ?$$

The novelty of the work is fourfold:

i) it interprets the computing activity of well known algorithms to solve $Ax = b$ or $Ax = \lambda x$ in terms of knowledge acquisition, thus bridging a gap between numerical linear algebra and cognitive sciences.

ii) it structures the field of singularities of $(A(t) - zI)^{-1}$ by means of two families of curves
   a) the *singular rays* which are obtained for $t = re^{i\theta}$, with a fixed θ in $[0, 2\pi[$,
   b) the *singular orbits* which correspond to $t = re^{i\theta}$, with r fixed, $r > 0$.

iii) It gives a global synthetic view of the resolvent field $z \rightarrow (A(t) - zI)^{-1}$ for the family $A(t) = A + tE$, when $|\, t \,|$ varies in $]0; \infty[$. There is **no restriction** on E, such as $\|\, E \,\|$ small enough, as is usual in classical perturbation theory. The view relies on the analysis of the map $z \rightarrow \rho(E(A - zI)^{-1})$.

iv) It provides a theoretical framework to analyze the interplay between exact computation and finite precision results.

2. Knowledge acquisition via Inexact Computing

We are given two matrices A and E of order n such that

$$B = A + E.$$

We suppose that the *resolvent field* for A: $z \rightarrow (A - zI)^{-1}$ is known for any z not an eigenvalue of A. The set of such eigenvalues is denoted $\sigma(A)$, the *spectrum* of A. It is the *field of singularities* of A: the set of complex points z such that $(A - zI)^{-1}$ does not exist.
The resolvent field allows to solve the linear system

$$(A - zI)x = y \Leftrightarrow x = (A - zI)^{-1} y$$

for any $z \notin \sigma(A)$, and any vector y in $\mathbb{C}^n$.

Now we consider the similar system with B:

$$(B - zI)u = y. \tag{2.1}$$

If we knew $(B - zI)^{-1}$ (whenever it exists), we would write $u = (B - zI)^{-1} y$ and the problem would be solved. However, the rule of the learning game is that we are allowed only to use E and the resolvent field $z \rightarrow (A - zI)^{-1}$ to solve (2.1). Such a game is called **Inexact Computing**.

2.1. Knowledge acquisition

How can we play the learning game? We introduce the homotopic family $A(t) = A + tE$, $t \in \mathbb{C}$, such that $A(0) = A$ and $A(1) = B$. For any $z \notin \sigma(A)$,

$$\begin{aligned} A(t) - zI &= A + tE - zI \\ &= (A - zI)(I + t(A - zI)^{-1}E) \\ &= (I + tE(A - zI)^{-1})(A - zI) \end{aligned}$$

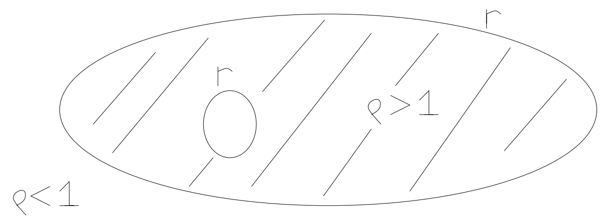


Figure 2.1. *Partition of* $\mathbb{C}$*, where the shaded area is the inexact pseudo-spectrum* $\Psi\sigma$

As recalled in the Annex, the Neumann series

$$(A(t)-zI)^{-1} = (I+\sum_{k=1}^{\infty}(-t(A-zI)^{-1}E)^k)(A-zI)^{-1}$$

converges iff $\rho((A-zI)^{-1}E)<1$. Note that $\rho((A-zI)^{-1}E)=\rho(E(A-zI)^{-1})$. Indeed,

$$E(A-zI)^{-1}x=\lambda x,\ x\neq 0 \Leftrightarrow (A-zI)^{-1}Ey=\lambda y \ \text{ with } \ y=(A-zI)^{-1}x\neq 0.$$

Therefore the solution $x(t)$ of $(A(t)-zI)x(t)=y$ is obtained as the limit, when $k\rightarrow\infty$, of the series of vectors

$$\begin{aligned} x_k(t) &= (I+\textstyle\sum_{j=1}^k(-t(A-zI)^{-1}E)^j)x \\ &= (A-zI)^{-1}(I+\textstyle\sum_{j=1}^k(-tE(A-zI)^{-1})^j)y, \end{aligned} \tag{2.2}$$

for any t such that $|\,t\,|\leq 1$. The rate of convergence of $x_k(1)$ toward $u=(B-zI)^{-1}y$ depends on the value of $\rho(E(A-zI)^{-1})$ compared to 1.

When z varies in $\mathbb{C}$ outside of $\sigma(A)$, the map $z\rightarrow\rho(E(A-zI)^{-1})$ is an upper semi-continuous function of z (proposition 2.2.7 in [16,17]). Therefore the set

$$\Gamma(A,E)=\{z;\rho(E(A-zI)^{-1})=1\} \tag{2.3}$$

is a compact set in $\mathbb{C}$ consisting of a finite number of closed curves, certain of which enclose eigenvalues of A. See in Section 5 a proof of this assertion. It is the boundary of the domain of convergence of the Neumann series (2.2) for $|\,t\,|<1$. The set $\Gamma=\Gamma(A,E)$ divides the complex plane into two complementary sets:

i) the inexact *resolvent set*, which is an unbounded open set defined by

$$\mathrm{Re}=\{z;\ u=(B-zI)^{-1}y=lim_{k\rightarrow\infty}x_k(1)\}=\{z;\rho(E(A-zI)^{-1})<1\}$$

ii) the *inexact pseudo-spectrum* which is a bounded closed set defined by

$$\begin{aligned} \Psi\sigma &= \{z;\ B-zI \text{ is not invertible with the available computing means (Neumann series)}\}. \\ &= \{z;\rho(E(A-zI)^{-1})\geq 1\} \end{aligned}$$

See Figure 2.1. In general, $\Psi\sigma$ is a finite collection of connected regions which is not itself necessarily simply connected. The regions are such that one or several eigenvalues of A lie inside.
From the point of view of **Inexact Computation** Re is the set of complex points where (2.1) can be

solved by means of a convergent Neumann series. On the other hand, $\Psi\sigma$ is the set of points where (2.1) cannot be solved, be it because this is impossible or because the solution exists but cannot be computed by a Neumann series. The pseudo-spectrum $\Psi\sigma$ can be seen as a *black hole* because no information on the resolvent field $z \to (B - zI)^{-1}$ can be obtained inside it.

This point of view, the possibility of knowledge acquisition outside of $\Psi\sigma$ via the resolvent field $z \to (B - zI)^{-1}$, will be complemented in the next Section by the dual point of view of the singularities of B, and more generally, the singularities of $A + tE$.

2.2. The speed of knowledge acquisition

This topic will be addressed specifically in Section 4, after the presentation of the dual point of view in Section 3.

3. Distance to singularity in Inexact Computing

The point of view in this section is, given z in $\mathbb{C} \setminus \sigma(A)$, can this point z be seen as an eigenvalue of $A + tE$ for one or more values of t?
This can be easily studied by means of the factorization, valid for $z \notin \sigma(A)$,

$$A + tE - zI = [I + tE(A - zI)^{-1}](A - zI).$$

Therefore z is an eigenvalue of $A+tE$ iff $1 + t\lambda_i = 0$, for λ_i, $i = 1, \ldots, n$, any eigenvalue of $E(A - zI)^{-1}$. We conclude that z can be seen as an eigenvalue of at most n matrices $A + t_i E$ such that $t_i = \frac{-1}{\lambda_i}$, $i = 1, \ldots, n$. This requires that both λ and t are nonzero.
Amongst these n matrices in the family $A(t) = A + tE$, there exists one which is **closest**, this is the one associated with $t_\star$ of minimum modulus which realizes

$$\mid t_\star \mid = \min_i \mid t_i \mid = \frac{1}{\max_i \mid \lambda_i \mid}.$$

The matrix $A(t_\star)$ is the closest matrix of the family $A(t)$ for which z is an eigenvalue, when the distance is measured by the modulus of t.

Definition 3.1. *The (*homotopic*) *distance to singularity *of z is*

$$\min(\mid t \mid, z \text{ is an eigenvalue of } A + tE).$$

It is also called (homotopic) backward error*, that we denote $BE(z)$.*

It can be computed by the formula

$$BE(z) = 1/\rho(E(A - zI)^{-1}). \tag{3.1}$$

The notion of distance to singularity allows to reinterpret the curve Γ defined by (2.3) as the curve of all points z for which the distance to singularity (or equivalently the backward error) is equal to 1. This explains the:

Definition 3.2. *The border $\Gamma = \partial\Psi\sigma$ of the pseudo-spectrum is called the* spectral curve *associated with A and E.*

We turn now to the task of characterizing the points z in $\mathbb{C}\setminus\sigma(A)$ as singularities which lie on Γ and inside $\Psi\sigma$. For the eigenvalues on Γ, we shall also specify at least one eigendirection, thanks to the

Lemma 3.1. *y is an eigenvector of $E(A-zI)^{-1}$ associated with $\lambda \neq 0$ iff $s=(A-zI)^{-1}y$ is an eigenvector of $A-\frac{1}{\lambda}E$ associated with z.*

Proof

$$E(A-zI)^{-1}y=\lambda y \Leftrightarrow \lambda(A-zI)s=Es \Leftrightarrow (A-\frac{1}{\lambda}E)s=zs$$

□

When E is of rank >1, there are several possible matrices associated with z as we saw. However when E is of rank 1, so is $E(A-zI)^{-1}$. Therefore it has only one eigenvalue λ which is possibly nonzero. To this unique $\lambda \neq 0$ is associated a finite $t=\frac{-1}{\lambda}$ and z is an eigenvalue of the unique matrix $A-\frac{1}{\lambda}E$. If $\lambda=0$ then t is unbounded: $|\, t \,| \to \infty$ as $\lambda \to 0$.

This discussion indicates that it is convenient to distinguish between the possible values of rank of E.

i) <u>E is of rank 1</u>
$E(A-zI)^{-1}$ is of rank 1 for $z \notin \sigma(A)$ and has exactly one possibly nonzero eigenvalue. Therefore z on $\Gamma \Leftrightarrow \rho=1 \Leftrightarrow \lambda=e^{i\theta} \Leftrightarrow t=e^{i(\pi-\theta)} \Leftrightarrow \Gamma$ is the set of all eigenvalues of $A+tE$, with $|\, t \,|=1$. Let us write $E=\alpha\beta^H$, where $\alpha,\ \beta \in \mathbb{C}^n$ are two nonzero vectors. Therefore α defines the unique eigendirection of E associated with the possibly nonzero eigenvalue $\beta^H\alpha$ depending whether $\beta^H\alpha \neq 0$ or not: $E\alpha=(\beta^H\alpha)\alpha$, $\alpha \neq 0$. The other eigenvalue 0 (of multiplicity $n-1$ if $\beta^H\alpha \neq 0$) is associated with the eigenspace Ker $E=\{\beta\}^{\perp}$.

From the uniqueness of the eigendirection α for $E(A-zI)^{-1}$ associated with $\lambda=e^{i\theta}=\beta^H(A-zI)^{-1}\alpha$, we deduce the uniqueness of the eigendirection $\gamma=(A-zI)^{-1}\alpha$ associated with the (not necessarily simple) eigenvalue z of $A'=A+e^{i(\pi-\theta)}E$. The unique eigendirection γ of A' associated with z, is deduced from the knowledge of the resolvent field $z \to (A-zI)^{-1}$ and of the range of the deviation matrix E: Im $E=\{\alpha\}$.
If we set $t=1$, we see in particular that z on Γ is an eigenvalue of $B=A+E$ iff $\lambda=-1$ is an eigenvalue of $E(A-zI)^{-1}$, that is $\beta^H(A-zI)^{-1}\alpha=-1$.

ii) <u>E is of rank >1</u>
The matrix $E(A-zI)^{-1}$ has at least 2 possibly nonzero eigenvalues. Therefore there exist eigenvalues of $A+tE$, $|\, t \,|=1$, which lie inside $\Psi\sigma$ and not on its boundary Γ.

In conclusion, if E is of rank 1, any z in $\Psi\sigma$ is an eigenvalue of $A+tE$ for $|\, t \,|\leq 1$, for which we know the unique eigendirection.
When E is of rank >1, one can only conclude that all eigenvalues of $A+tE$ with $|\, t \,|\leq 1$ belong to $\Psi\sigma$. The reciprocal does not hold, which means that $\Psi\sigma$ may contain eigenvalues of $A+tE$ for $|\, t \,|>1$ [5]. This will be illustrated later in Section 6.

The above discussion can be summarized by the following proposition:

Proposition 3.1. *i) When E is of rank 1, the curve Γ runs through the n eigenvalues of any matrix $A(t)=A+tE$, $|\, t \,|=1$. For each distinct eigenvalue z is associated at least one eigendirection* *defined by* $\gamma = (A - zI)^{-1}\alpha$.

ii) *When* rank $E > 1$, *the curve* Γ *runs through some (but not necessarily all) of the n eigenvalues of* $A(t)$, $|\, t \,| = 1$. *For such eigenvalues on* Γ, *at least one eigenvector is associated and defined by the same formula.*

Proof:

i) If E is of rank 1, there is only *one* possibly nonzero eigenvalue λ for $E(A - zI)^{-1}$. Therefore, the n eigenvalues z_i, $i = 1, \ldots, n$, of $A + tE$ with $|\, t \,| = 1$ define (at most) n matrices $E(A - z_i I)^{-1}$ which have the same non zero eigenvalue λ, with $\lambda t = -1$, hence $|\, \lambda \,| = 1$.

ii) If rank$E > 1$, the non-uniqueness of $\lambda \neq 0$ implies that not all of the n eigenvalues of $A + tE$ can always represented that way with $z \in \Gamma$, $|\, t \,| = 1$.

□

When E is of rank 1, all the eigenvalues of $A(t)$ reach the spectral curve Γ in a unit length interval $|\, t \,| = 1$.
When E is of rank > 1, a desynchronisation may occur since certain points *inside* $\Psi\sigma$ require a unit interval $|\, t \,| = 1$ to be reached, even though they are not on Γ. At such points, no knowledge can be acquired, because $\rho > 1$. This phenomenon will be illustrated in Section 6.

It is remarkable that the spectral curve Γ lends itself to two complementary interpretations:

1. on the one hand, it is the border of the exterior domain Re where knowledge acquisition is possible, because the *radius of convergence* is 1,
2. on the other hand, it is the border of the interior domain $\Psi\sigma$, a curve consisting of points for which the *distance to singularity* equals 1.

4. The rate of knowledge acquisition

We show that, at a finite number of points z, a superfast convergence can be achieved.

4.1. E is of rank 1

We examine the rate of convergence of the series expansion (2.2) for $x_k(t)$, for z outside Γ. And we suppose that $E = \alpha\beta^H$, with α, β nonzero vectors of $\mathbb{C}^n$. In this particular case, (2.2) can be conveniently rewritten by introducing the two scalars:

$$\xi = \beta^H (A - zI)^{-1} y \qquad \text{and} \qquad \mu = \beta^H (A - zI)^{-1}\alpha = \mu(z).$$

μ is the unique possibly nonzero eigenvalue of $E(A - zI)^{-1}$, so that $\rho(E(A - zI)^{-1}) = |\, \mu \,|$.

Lemma 4.1. *With* $E = \alpha\beta^H$ *of rank 1, the limit solution* $x(t)$ *is given by* $x(t) = (A - zI)^{-1}[y + s(t)\alpha]$, *where the scalar* $s(t) = \frac{-\xi t}{1 + t\mu}$ *is computed for* $|\, \mu t \,| < 1$, *as the limit of a geometric sequence* $s_k(t)$.

Proof:
From (2.2), follows that

$$x_k(t) = (A - zI)^{-1}[y + (\sum_{j=1}^{k}(-t)^j \mu^{j-1}\xi)\alpha].$$

Therefore

$$\begin{aligned} s(t) &= -\xi \textstyle\sum_{j=0}^{\infty}(-t)^j \mu^j \\ &= -\frac{\xi t}{1+t\mu}. \end{aligned}$$

$s(t) = \lim_{k\to\infty} s_k(t)$, where s_k is the sum of the first k terms of the series as long as $| t\mu |< 1$, which is the condition z outside Γ.

□

When $s(t)$ is computed by means of (2.2), the rate of convergence is bounded by $| t\mu |< 1$. Since $\mu = \mu(z)$ depends on z, it is conceivable that for certain z in Re, μ is zero. This can happen when the rank 1 matrix $F_z = E(A - zI)^{-1}$ is in addition $defective$, that is when $F_z^2 = 0 \Leftrightarrow \mu = \beta^H(A - zI)^{-1}\alpha = 0$. Indeed $\mu(z)$ is a rational fraction, whose denominator is the characteristic polynomial in z, of degree n, and the numerator is a polynomial in z of degree d, $0 \le d \le n-1$. So it has d roots in $\mathbb{C}$ ($d = 0$ if $n = 1$). At such points z in $\mathbb{C}$, the condition $| t\mu |< 1$ is satisfied for all t and the learning process is therefore **superfast** for any t: $s(t) = -t\xi$ and

$$x(t) = (A - zI)^{-1}(y - t\xi\alpha) = x - t\xi\gamma$$

if we set $\gamma = (A - zI)^{-1}\alpha$. At such points, $x(t)$ is *linear* in t for all t. We note that $\mu = \mu(z) \to 0$ as $| z |\to \infty$. We prove in [12] that when $A = J^T$ is the companion matrix of x^n and $E = -ae_n^T$, $a = (a_0, ..., a_{n-1})^T$, so that $A + E$ is the companion matrix associated with

$$p(x) = x^n + \sum_{i=0}^{n-1} a_i x^i = x^n + q(x)$$

then $\mu(z) = \frac{q(z)}{z^n}$ is zero at the roots of $q(z)$. This polynomial of degree $d \le n-1$ has $n-1$ roots in $\mathbb{C}$, unless $a_{n-1} = 0$. There are therefore $d \le n-1$ points z of superfast convergence, for which $p(z) = z^n$ and $x(t)$ is linear in t for any t. Moreover, the choice y parallel to α yields $\xi = 0$, $s(t) = 0$ and $x(t) = x$ for all t: the solution is stationary (independent of t).
This means that the solution $x = (A - zI)^{-1}y$ we get at the first iteration happens to be the exact solution of $(B - zI)u = y$ with $B = A + E$, which was our goal for the learning process. And this is true for any solution $x(t)$ of $(A + tE - zI)x(t) = y$, $t \in \mathbb{C}$.
This happens for the choice of a vector y parallel to the direction α which defines the range of the deviation matrix $E = \alpha\beta^H$. We set

$$(A - zI)\gamma = \alpha \Leftrightarrow \gamma = (A - zI)^{-1}\alpha.$$

Points z where $\mu(z) = 0$ are **special in two ways**. First, Lemma 3.1 indicates that a point z such that $\mu(z) = 0$ cannot be interpreted as an eigenvalue of $A + tE$ because t would be infinite. We have to interpret it as a regular point for $A + tE - zI$: the matrix is regular.
Second the inverse $(A + tE - zI)^{-1}$ can be represented more effectively than by the limit of an infinite Neumann series convergent for $| t | \rho < 1$. The behaviour is now *finite*. There is no question of convergence with respect to t anymore, because one has the *finite* expansion

$$(I + tE(A - zI)^{-1})^{-1} = I - tE(A - zI)^{-1}$$

valid for any t. This is in sharp contrast with the infinite Neumann series expansion which is required when $\mu(z) \neq 0$.

4.2. *And if rank $E > 1$?*

When E is of rank k, $n \geq k \geq 1$, we may write $E = UV^H$, where U and V are matrices in $\mathbb{C}^{n\times k}$, of rank k such that $V^H U$ is invertible. Therefore $F_z = UV^H(A - zI)^{-1}$ of rank k has the same k possibly non zero eigenvalues as the $k \times k$ matrix $M_z = V^H(A - zI)^{-1}U$. The elements of M_z are rational fractions in z, that is they are ratios of a polynomial of degree $\leq n-1$ to the characteristic polynomial of A of degree n.
An analysis, similar to the previous one done for $k = 1$, can be carried out to look for conditions of superconvergence. The fact that now M_z is a matrix of order $k > 1$, rather than the scalar $\mu(z) = \beta^H(A - zI)^{-1}\alpha$, allows more freedom and therefore richer conclusions can be achieved. The reader is referred to [12] for a complete analysis in the case $k = 2$.

5. The global geometry of the map $\varphi:\ z \to \rho(E(A - zI)^{-1})$

5.1. *A related problem: the map $\Psi : z \to \| (A - zI)^{-1} \|$*

During the decade of 90s, a lot of attention has been given to the map Ψ defined for $z \in \mathbb{C}$, $z \notin \sigma(A)$, into $\mathbb{R}^+$. Such a map Ψ is often encountered in stability issues related to the distribution and influence of eigenvalues of A. Theoretical insight can be gained by inspecting the 2D level curves of Ψ. It is easy to show ([6], p. 177-178) that the r-level curve = $\{z;\ \| (A - zI)^{-1} \| = \frac{1}{r}\}$, $r > 0$, is the border of the set of all the eigenvalues of $A + \Delta A$, for all matrices ΔA such that $\| \Delta A \| \leq r$. Such a set is known as the *normwise r-pseudo-spectrum* of A, where A can be a matrix or a linear operator [19,25,26].

As for the Γ curve, the 1-level curve $\{z; \| (A - zI)^{-1} \| = 1\}$ for Ψ can be interpreted by means of the (normwise) backward error or (normwise) distance to singularity of z [6].
The 1-level curves for $\Psi(z)$ and $\varphi(z)$ can be easily related [24] when $\| E \| = 1$ because of the inequality

$$\rho(E(A - zI)^{-1}) \leq \| E(A - zI)^{-1} \| \leq \| (A - zI)^{-1} \| .$$

An important application of normwise pseudo-spectra is the analysis of finite precision computation by means of scaled norms [6,10,24].

5.2. *Ψ and φ as subharmonic functions for z in $\mathbb{C}\setminus\sigma(A)$*

For any $z \notin \sigma(A)$, $(A - zI)^{-1}$ is an analytic function which can be written under the Cauchy integral formula: if the disk $\{s \in \mathbb{C},\ | s - z | \leq r\}$ contains no eigenvalue of A, then

$$(A - zI)^{-1} = \frac{1}{2i\pi}\int_{|s-z|=r} \frac{(A - sI)^{-1}}{s - z} ds = \frac{1}{2\pi}\int_0^{2\pi} [A - (z + re^{i\theta})I]^{-1} d\theta.$$

This powerful result is what is needed to prove that $\varphi(z) = \rho(E(A - zI)^{-1})$ and $\Psi(z) = \| (A - zI)^{-1} \|$ are *subharmonic* functions in $\mathbb{C}\setminus\sigma(A)$. We recall that a subharmonic function $f : \Omega \subset \mathbb{C} \to \mathbb{R}$, when Ω is an open domain, is continuous (or at least upper semi continuous) in Ω and satisfies, for any z in Ω such that $\{s,\ | s - z | < \alpha\} \subset \Omega$, the *inequality*

$$f(z) \leq \frac{1}{2\pi}\int_0^{2\pi} f(z+re^{i\theta})d\theta,$$

for any r, $0 < r \leq \alpha$.

The following is standard in Complex Analysis [2].

Proposition 5.1. *A subharmonic function f in Ω satisfies a Maximum Principle:*

$$\sup_{z\in\Omega} f(z) = \max_{s\in\partial\Omega} f(s).$$

From this follows that f cannot have a local maximum, unless it is constant locally. More can be said for φ and ψ because of their behaviour at infinity:

$$\lim_{|z|\to\infty} \rho(E(A-zI)^{-1}) \leq \|E\| \lim_{|z|\to\infty} \left\|(A-zI)^{-1}\right\| = 0$$

by Corollary 2.2.5 in [16, 17].

Theorem 5.1. *If φ and Ψ are locally constant, they have to be zero on $\mathbb{C}\setminus\sigma(A)$.*

Proof:

1. We begin with the map $z \to \Psi(z)$. Since

$$\| (A-zI)^{-1} \| = \max_{x\in\mathbb{C}^n} \frac{\| (A-zI)^{-1}x \|}{\| x \|},$$

   there exists a particular x, $\| x \| = 1$, such that $\| (A-zI)^{-1} \| = \| (A-zI)^{-1}x \|$. Now this vector norm can be represented by means of a scalar product: there exists y, $\| y \|_\star = 1$ $\|$ (y linear functional on $\mathbb{C}^n$ where $\| \,.\, \|_\star$ denotes the dual vector norm) such that

$$\| (A-zI)^{-1}x \| = y^H (A-zI)^{-1}x.$$

   The vectors x and y depend continuously on z, the function $g(z) = y^H(A-zI)^{-1}x$ is therefore *holomorphic* for z in $\mathbb{C}\setminus\sigma(A)$. It cannot be constant on an open domain unless it is constant everywhere in $\mathbb{C}\setminus\sigma(A)$. That this constant should be 0 follows from the fact that $\lim_{|z|\to\infty}\Psi(z) = 0$.

2. We turn to the map $z \to \varphi(z)$. We first remark that, in the particular case $E = \alpha\beta^H$ of rank 1, the above proof applies verbatim, where x (resp. y, g) becomes α (resp. β, φ).
   For a general E of rank > 1, we suppose that there exists an open domain Ω in $\mathbb{C}\setminus\sigma(A)$ on which $\varphi(z) = \rho(E(A-zI)^{-1})$ has the constant value $c \neq 0$. We choose a z in Ω, by definition it is an eigenvalue of a matrix $A(t) = A + tE$ with $| t | = \frac{1}{c} < \infty$. We consider a sequence $\{t_n\}$ in $\mathbb{C}$ such that $| t_n | < \frac{1}{c}$ for all n, and $t_n \to t$ as $n \to \infty$. For all n large enough, $n > N$, $A(t_n)$ is close enough to $A(t)$ that there exists at least one eigenvalue z_n of $A(t_n)$ in Ω. But such $z_n \in \Omega$ has a distance to singularity equal to $| t_n | < \frac{1}{c}$. This is impossible for all z in Ω. Hence the contradiction. The constant value c has to be zero for all z in $\mathbb{C}\setminus\sigma(A)$.

□

Corollary 5.1 *For z in $\mathbb{C}\setminus\sigma(A)$, one has only one of the two possibilities for φ (resp. ψ)*

i) the function φ (resp. ψ) is non constant on open subsets of $\mathbb{C}\setminus\sigma(A)$,
ii) φ (resp. ψ) is zero on $\mathbb{C}\setminus\sigma(A)$.

Proof: It follows from Theorem 5.1.

□

5.3. The geometric picture

Subharmonicity for φ (resp. ψ) has important consequences with respect to maximality properties. First at all, there cannot exist a local maximum unless the value is $+\infty$ (which happens at the eigenvalues in $\sigma(A)$). Second, because of corollary 5.1, level sets for φ (resp. ψ) have to be closed sets, and moreover closed *curves*.
The r-level curves enclose eigenvalues of A (local minimum at ∞) for r large enough. They can enclose a local minimum for r in a certain interval $[r_1, r_2] \subset [0, +\infty[$. Such examples can be found in [18, 24], and in the Venice example below.

For $r_1 = 0$, we have seen in Section 4, that the 0-level curve at finite distance, when it exists, is reduced to a finite number d, $0 \leq d \leq n-1$, of isolated points in $\mathbb{C}$, where we get superconvergence for (2.2). This is a particular case of the theorem ([2], théorème 6, p. 303) which states that the set of singularities (i.e. zero values) for a positive subharmonic function is non generic (its interior is empty). To distinguish theses points from the eigenvalues which are also singular points, we shall refer to the former as **critical** points.

6. Emergence of meaning via an increase of complexity: the interpretation with spectral rays and orbits

We wish to structure the pseudo-spectrum which is left unstructured by the point of view of solving (2.1) via series expansion (a “black hole”). Intuitively, we realize that we have to turn to the point of view of the singularities of the resolvent field in order to get an understanding.

In terms of **complexity**, this means changing from a system of n linear equations to the n roots of a characteristic polynomial, that is a polynomial of degree n. So the complexity of the new point of view (an eigenproblem) is **vastly increased** as compared to the original one (solving a linear system). The parameter t is the thread which will help us order $\Psi\sigma$, that is make sense of it. It will help us understand how it is to be interpreted, by means of t, from the data A and E in terms of knowledge acquisition as before, but of knowledge of **higher complexity**.

The eigenvalues $\lambda(t)$ of $A(t) = A + tE$ lie in the inexact spectrum $\Psi\sigma$ for $|\, t \,| \leq 1$. We shall analyze this domain of the complex plane by means of

i) the *spectral rays* $\Lambda(\theta)$, which are defined by

$$\Lambda(\theta) = \{\lambda(t) \text{ eigenvalue of } A(t) = A + tE \text{ for } t = re^{i\theta},\ 0 \leq r \leq 1 \text{ and } \theta \text{ fixed in } [0, 2\pi[\},$$

ii) the *spectral orbits* $\Sigma(r)$ which are defined by

$$\Sigma(r) = \{\lambda(t) \text{ eigenvalue of } A(t) = A + tE, \text{ for } t = re^{i\theta},\ 0 \leq \theta < 2\pi \text{ and } r \text{ fixed},\ r > 0\}.$$

The analysis of the structure of $\Psi\sigma$ is illustrated below on an example which is now described.

6.1. The structure of the inexact pseudo-spectrum with a companion matrix

A of order 8 is the companion matrix associated with the polynomial of degree 8:

$$\begin{aligned} p(x) &= (x-1)^3(x-3)^4(x-7) \\ &= x^8 - 22x^7 + 198x^6 - 958x^5 + 2728x^4 - 4674x^3 + 4698x^2 - 2538x + 567 \end{aligned}$$

We call **Venice** the associated companion matrix:

$$A = \begin{pmatrix} 0 & & & & & & & -567 \\ 1 & \ddots & & & & & & 2538 \\ & \ddots & \ddots & & & & & -4698 \\ & & \ddots & \ddots & & & & 4674 \\ & & & \ddots & \ddots & & & -2728 \\ & & & & \ddots & \ddots & & 958 \\ & & & & & \ddots & 0 & -198 \\ & & & & & & 1 & 22 \end{pmatrix}$$

It has 1 as triple eigenvalue, 3 as a quadruple eigenvalue and 7 as a simple one.
The deviation matrix E will be successively taken:

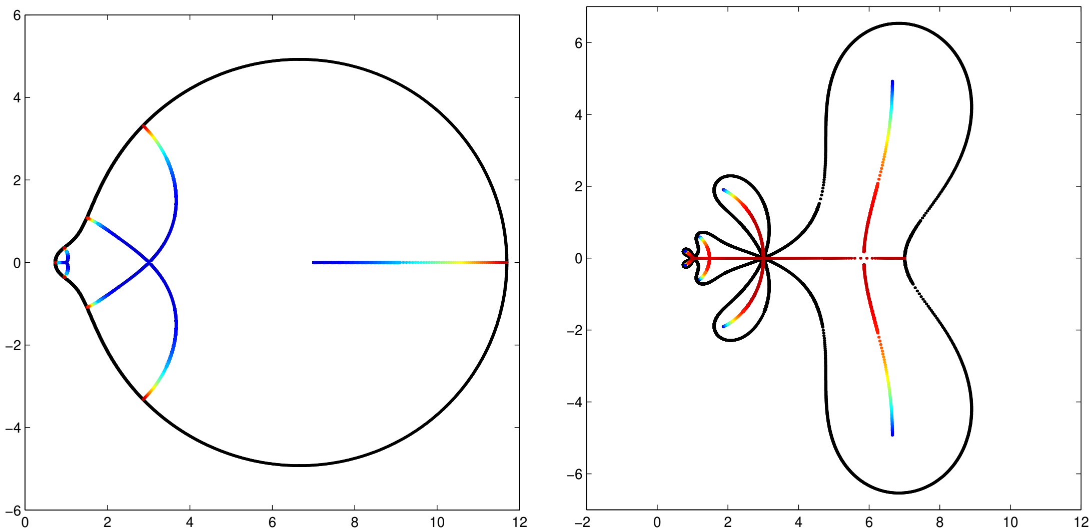
- of rank 1, as $E_1 = e\,e_8^T$, with $e = (1 \dots 1)^T$ and $e_8 = (0 \dots 01)^T$,
- of rank 2, as $E_2 = E_1 + e_1 e_6^T$,
- of rank 4, as $E_3 = E_2 + e_2 e_4^T + e_3 e_2^T$;

The same analysis will be conducted in parallel i) on A, E and ii) on $A - E, E$. In case i), the multiple eigenvalues are inside $\Gamma(A,E)$, whereas in case ii) they are on $\Gamma(A-E,E)$.

6.2. $E = E_1 = e\,e_8^T$ is of rank 1

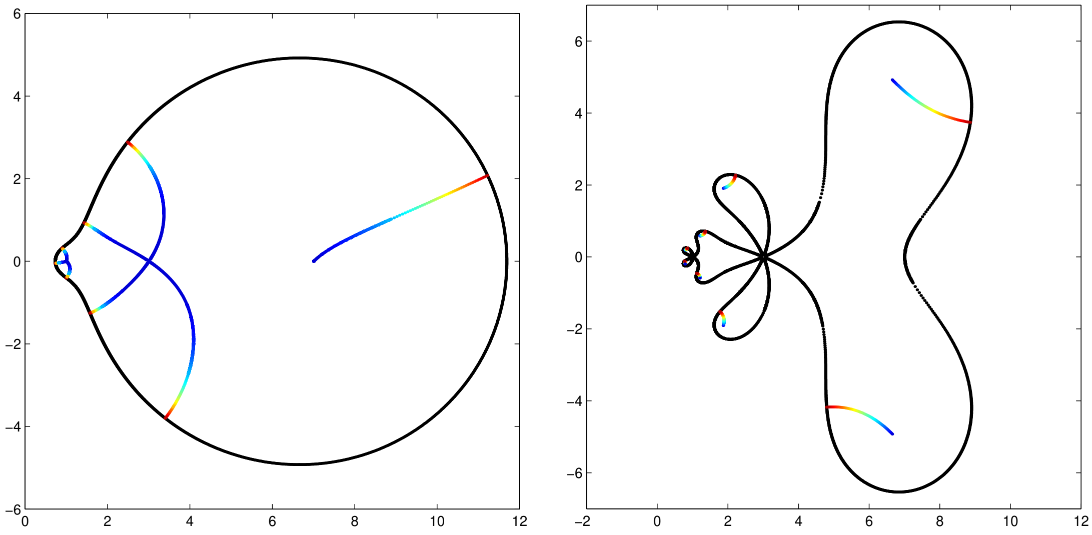
The spectral curve Γ (black) and 8 spectral rays $\Lambda(\theta)$ in black

Case i) A, E Case ii) $A - E, E$



$\theta = 0$

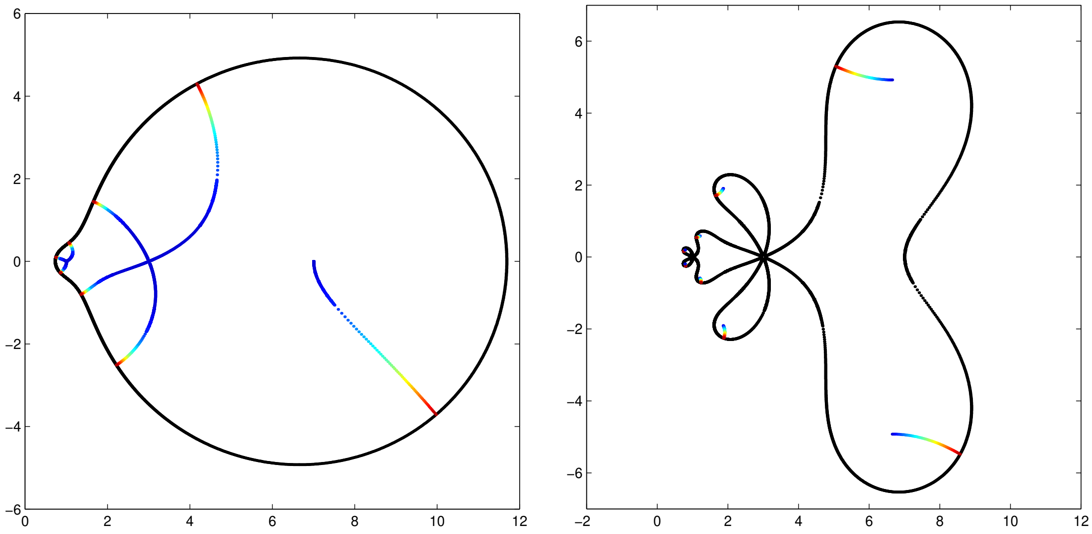
Case i) A, E Case ii) $A - E, E$



$\theta = \frac{\pi}{4}$

The spectral curve Γ (black) and 8 spectral rays $\Lambda(\theta)$ in black

Case i) A, E Case ii) $A - E, E$

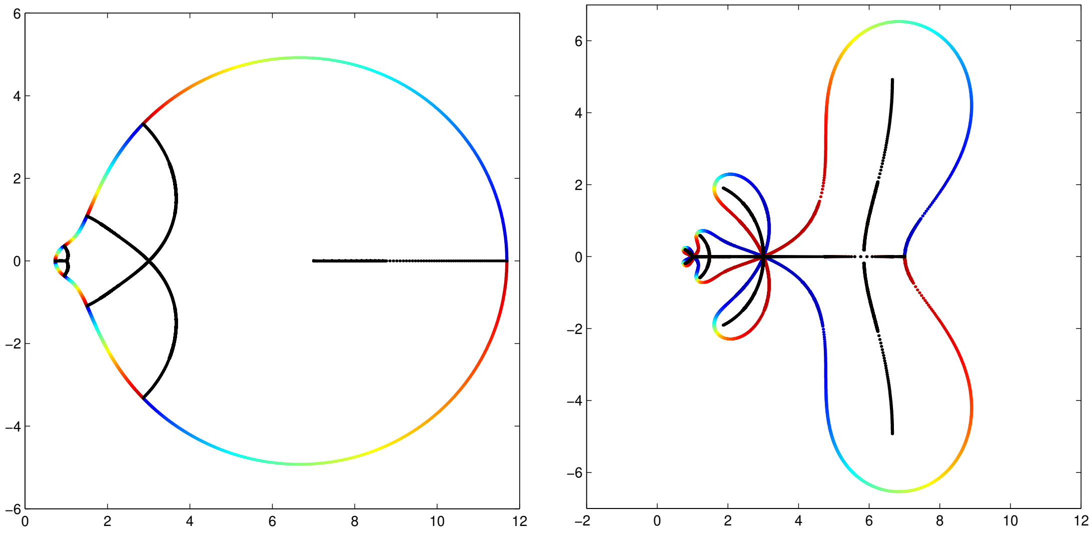


$\theta = \frac{3\pi}{2}$

As expected, all rays end on Γ. The color for $0 \leq r \leq 1$ is parameterized uniformly, see Section 7.

The spectral orbit $\Sigma(r)$ and 8 spectral rays $\Lambda(0)$ in black

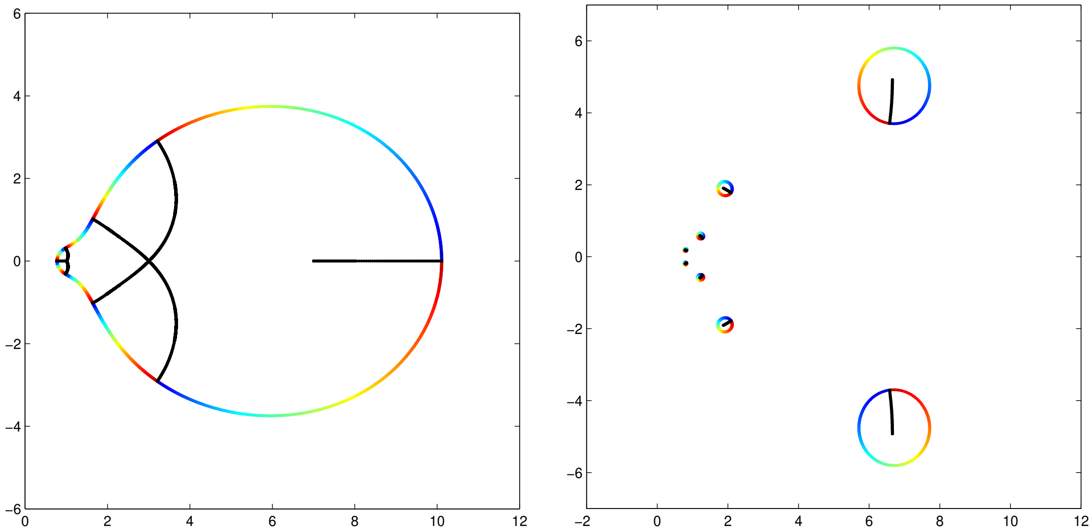
Case i) A, E Case ii) $A - E, E$



$r = 1$

The spectral orbit $\Sigma(r)$ and 8 spectral rays $\Lambda(0)$ in black

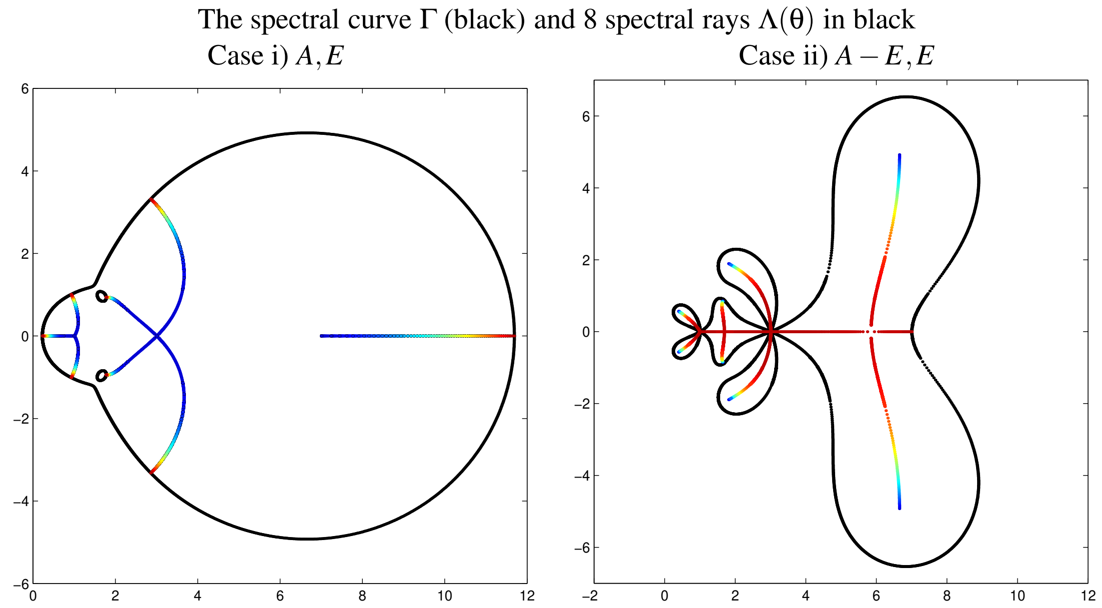
Case i) A, E Case ii) $A - E, E$



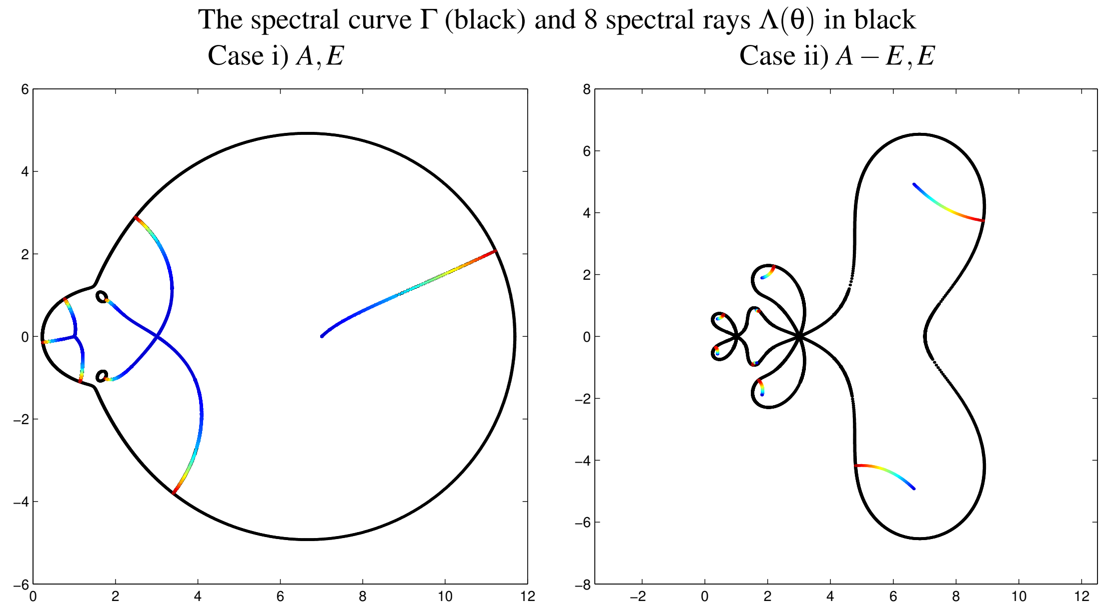
$r = \frac{1}{2}$

As expected, the spectral orbit $\Sigma(1)$ is identical to the spectral curve Γ. The 8 spectral rays $\Lambda(0)$ define in $\mathbb{C}$ 8 sectors, in which θ varies from 0 to 2π. This decomposes the spectral orbit $\Sigma(\theta)$ in 8 curvilinear segments, each of which has a full color span (blue to red) corresponding to $0 \leq \theta \leq 2\pi$. The color for $0 \leq \theta \leq 2\pi$ is parameterized uniformly, see section 7.

6.3. $E = E_2$ is of rank 2



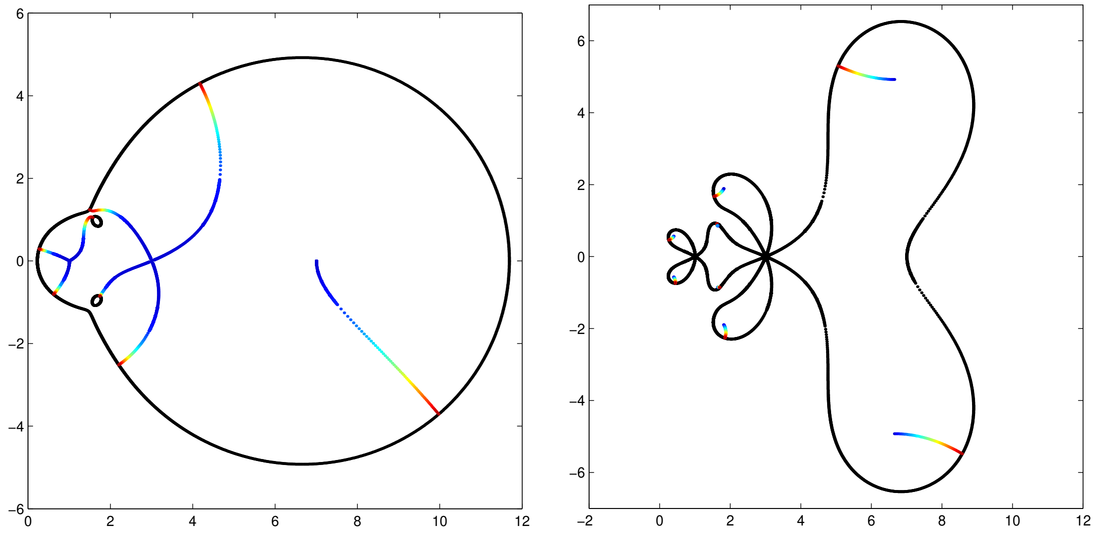
$\theta = 0$



$\theta = \frac{\pi}{4}$

The spectral curve Γ (black) and 8 spectral rays $\Lambda(\theta)$ in black

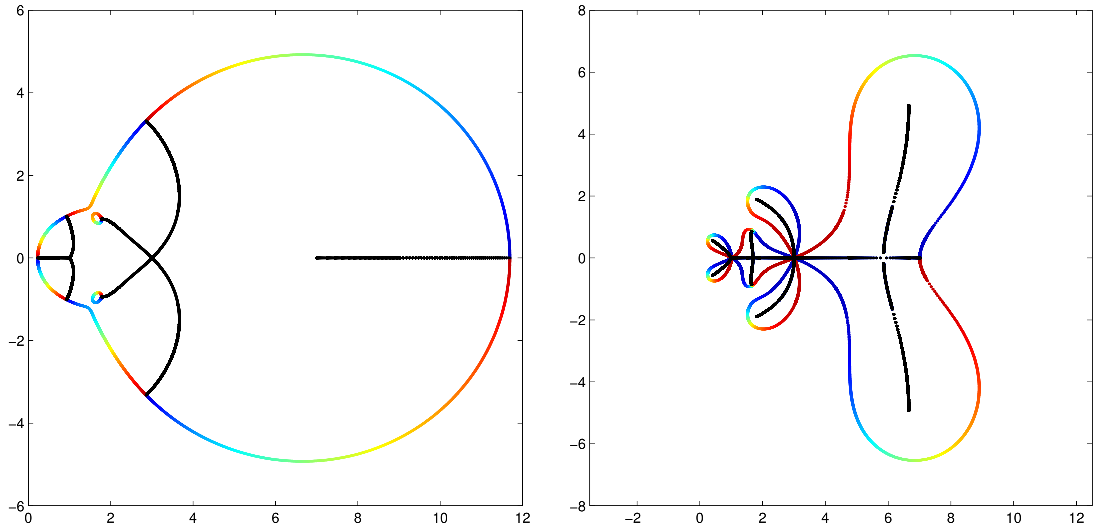
Case i) A, E Case ii) $A - E, E$



$\theta = \frac{3\pi}{2}$

The spectral orbit $\Sigma(r)$ and 8 spectral rays $\Lambda(0)$ in black

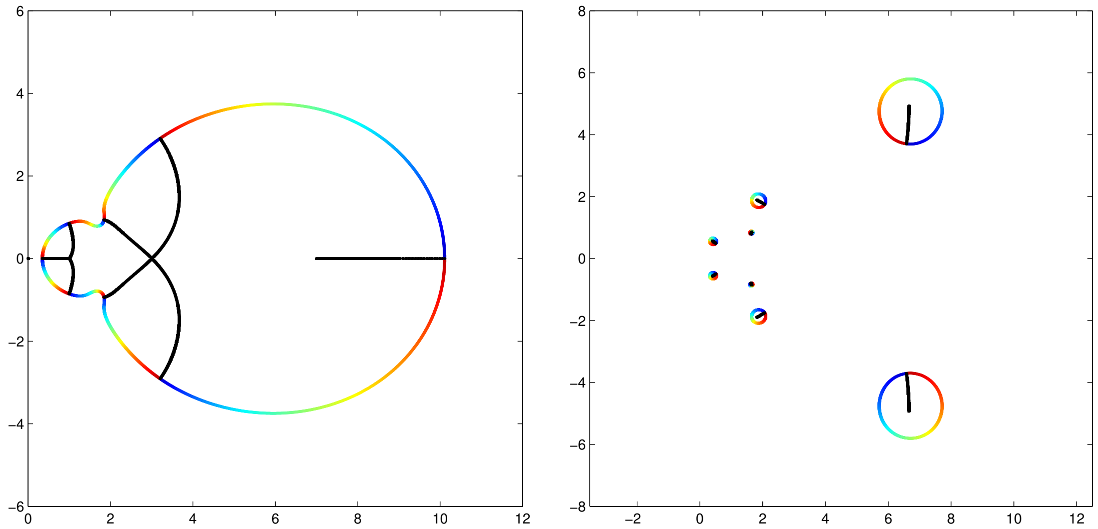
Case i) A, E Case ii) $A - E, E$



$r = 1$

The spectral orbit $\Sigma(r)$ and 8 spectral rays $\Lambda(0)$ in black

Case i) A, E Case ii) $A - E, E$



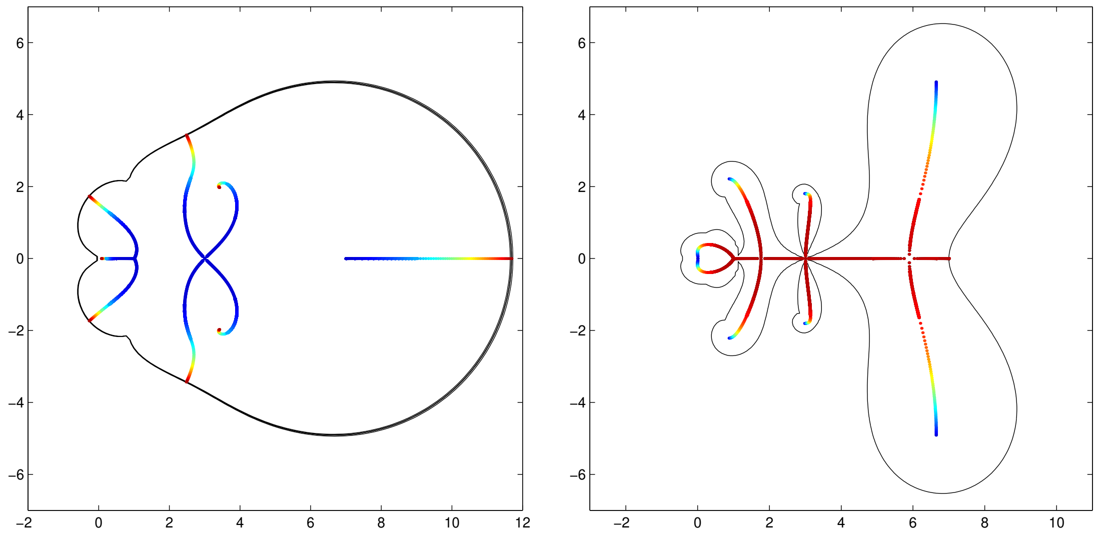
$r = \frac{1}{2}$

In this case of a rank 2 deviation, all rays still end on Γ which is identical to $\Sigma(1)$. The only important difference is that now, in case i), $\Sigma(1)$ *is not connected*: it consists of three closed curves, the two inner curves do not enclose any eigenvalue of A. However such a phenomenon is *not* related to the rank of E: it can happen when E is rank 1 [24]. As a consequence, $\Gamma \neq \Sigma(1)$ and there are only 6 curvilinear segments on Γ, each corresponding to $0 \leq \theta \leq 2\pi$. The two other curves are inside $\Psi\sigma$: they are not seen on Γ.

6.4. $E = E_3$ is of rank 4

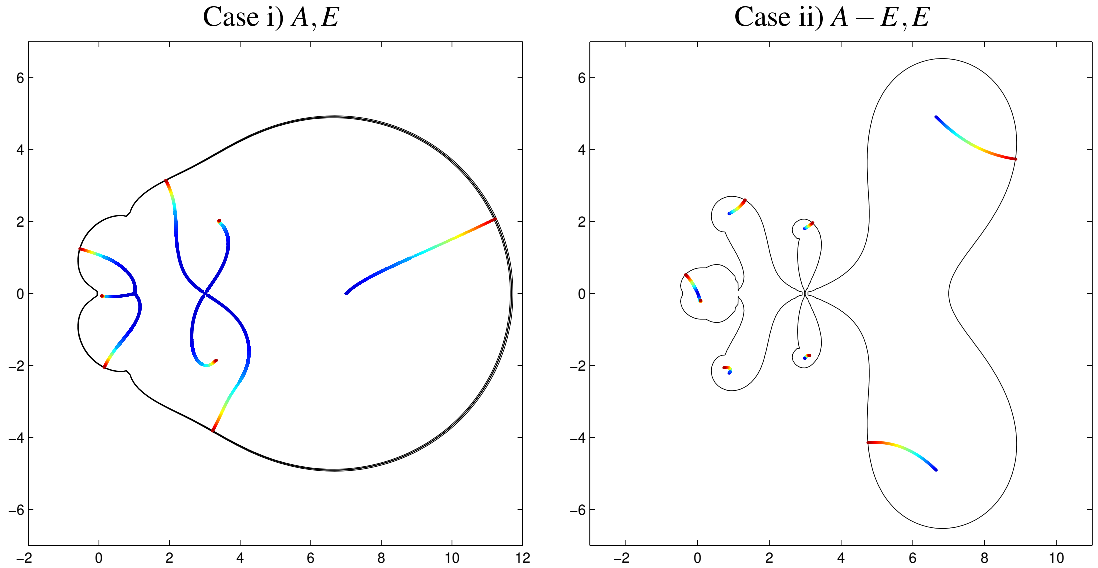
The spectral curve Γ (black) and 8 spectral rays $\Lambda(\theta)$ in black

Case i) A, E Case ii) $A - E, E$



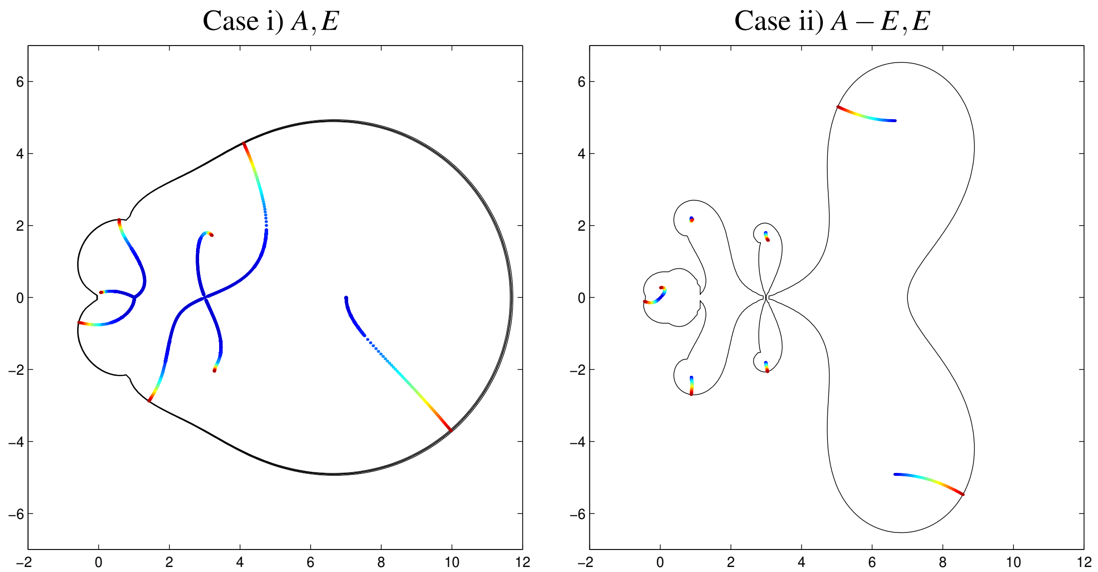
$\theta = 0$

The spectral curve Γ (black) and 8 spectral rays $\Lambda(\theta)$ in black



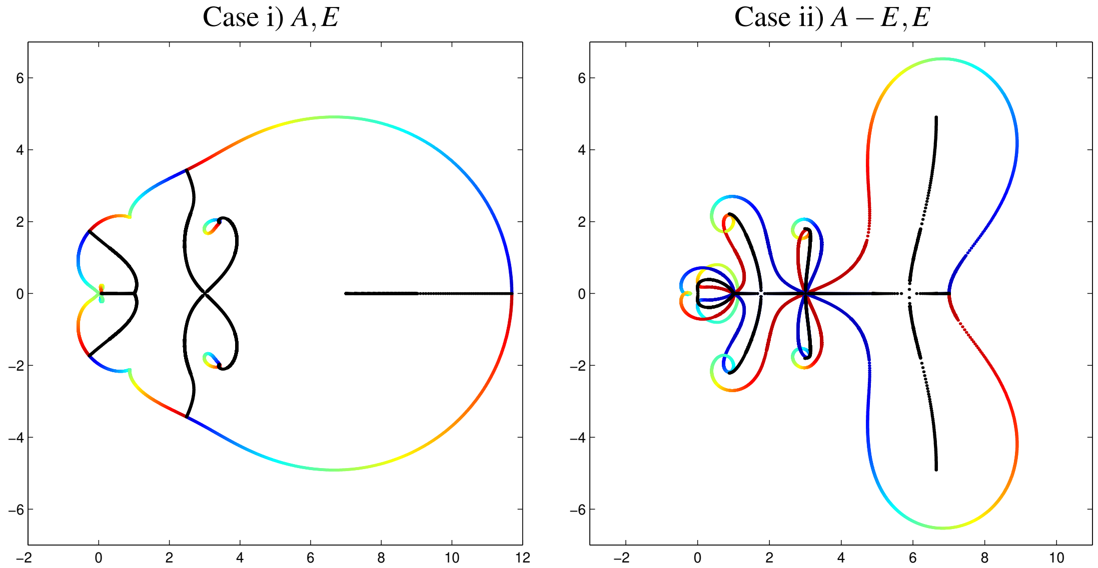
$\theta = \frac{\pi}{4}$

The spectral curve Γ (black) and 8 spectral rays $\Lambda(\theta)$ in black



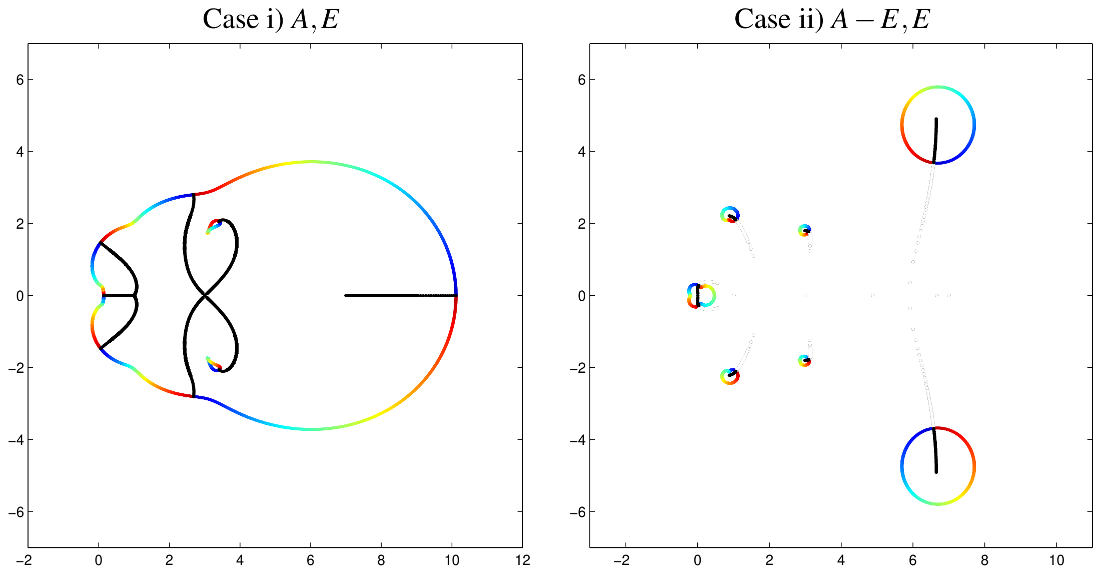
$\theta = \frac{3\pi}{2}$

The spectral orbit $\Sigma(r)$ and 8 spectral rays $\Lambda(0)$



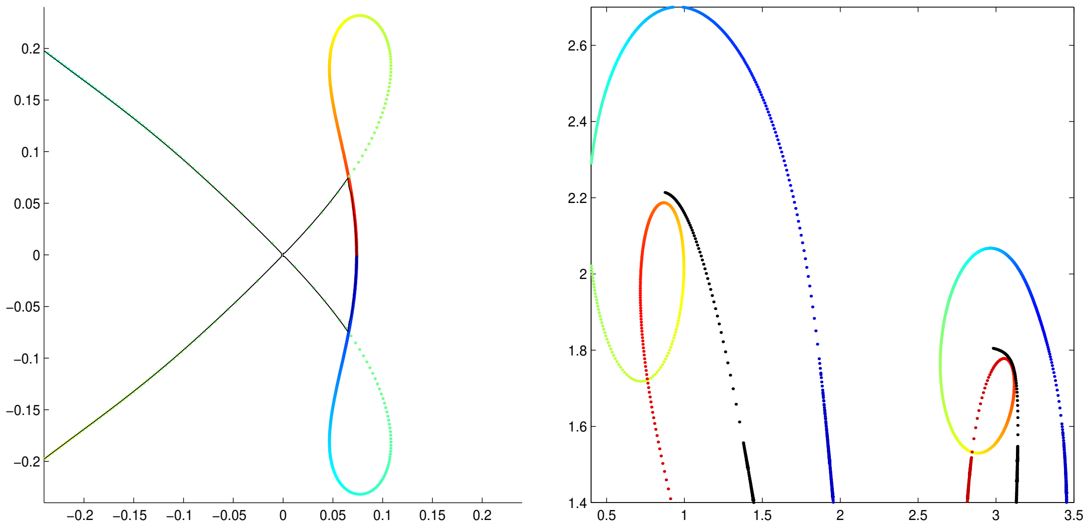
$r = 1$

The spectral orbit $\Sigma(r)$ and 8 spectral rays $\Lambda(0)$



$r = \frac{1}{2}$

The most striking phenomenon with this rank 4 deviation is the difference between Γ and $\Sigma(1)$ for A, E and even more so for $A - E, E$: in the latter case, the orbit tends to become “curly”. When a curl (or loop) of $\Sigma(1)$ happens to be inside Γ without enclosing any eigenvalue of A, or $A - E$, the spectral rays $\Lambda(\theta)$, which always end on $\Sigma(1)$, may end at a point inside $\Psi\sigma$ rather than on the border Γ. The significance of loops inside Γ versus outside Γ will be specifically addressed in Section 6.9.



Case i) A, E, rank(E)=4 ($r = 1$)
Zoom on [-0.24;0.24] $\times$ [-0.24;0.24]

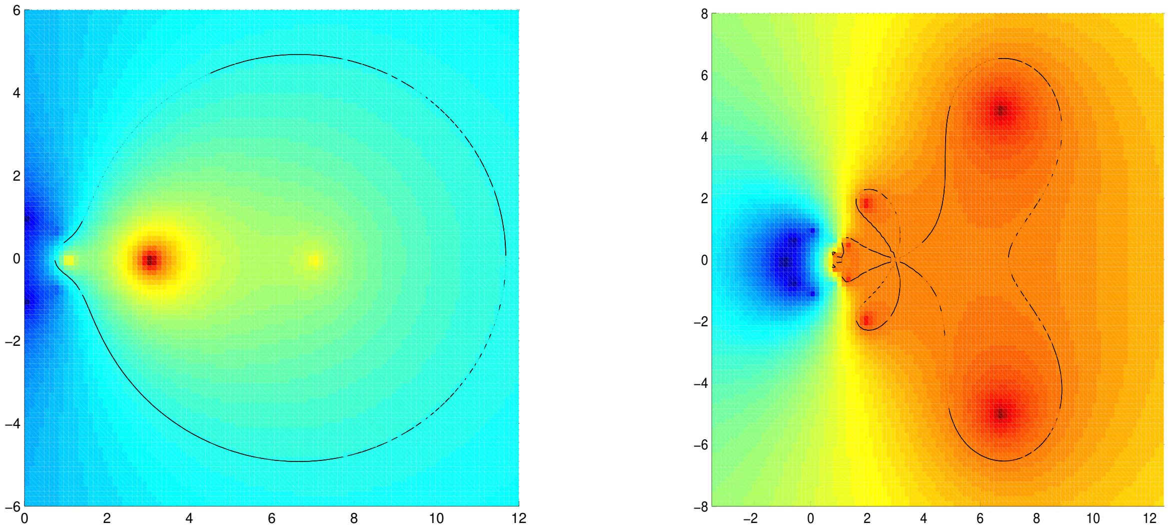
Case ii) $A - E, E$ ($r = 1$)
Zoom on [0,4;3,5]$\times$ [1,4;2,7]

6.5. The map $\varphi : z \longrightarrow \rho(E(A - zI)^{-1})$

It is more informative to consider a logarithmic scale for ρ, because of its wide range of variation in $]0;+\infty[$. The graphical display of $z \longrightarrow \log\rho(E(A - zI)^{-1})$ can be done in 2D with level curves, or rendered in 3D.

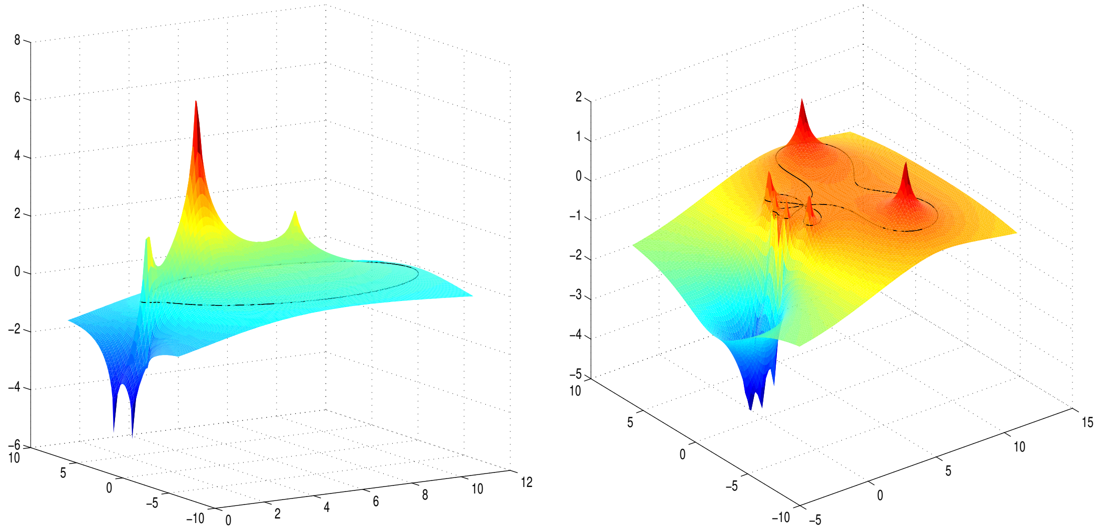
Case i) A, E

Case ii) $A - E, E$



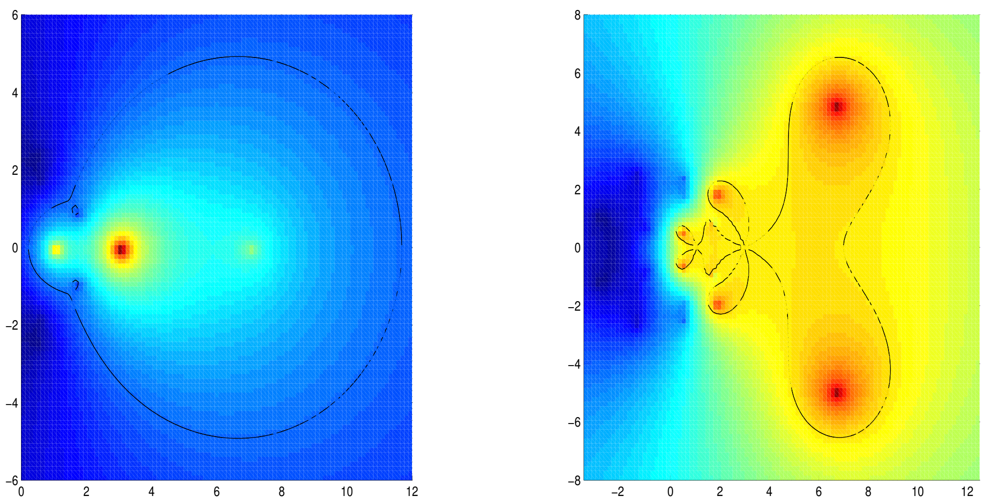
2D display for E of rank 1

Case i) A, E Case ii) $A - E, E$



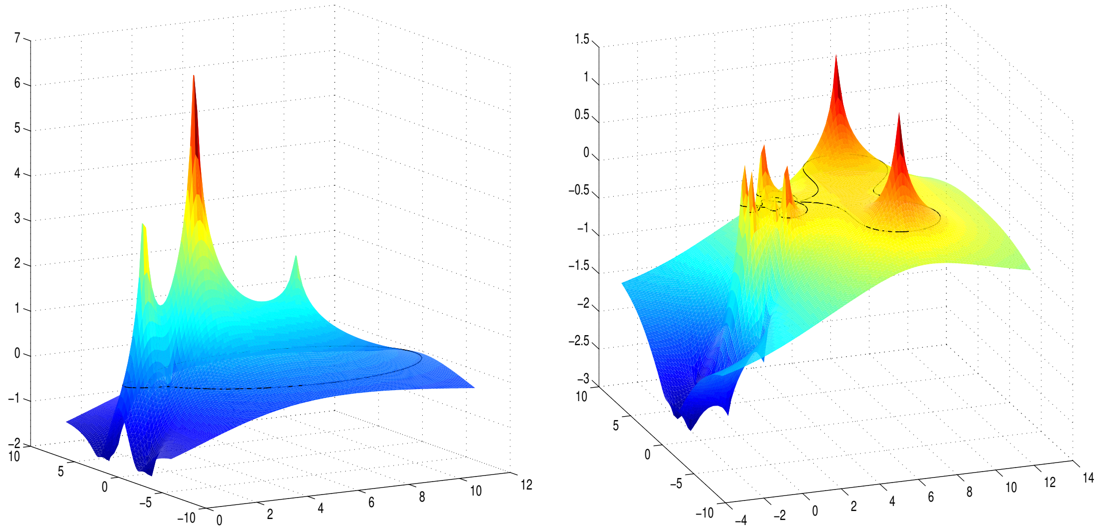
3D display for E of rank 1

Case i) A, E Case ii) $A - E, E$



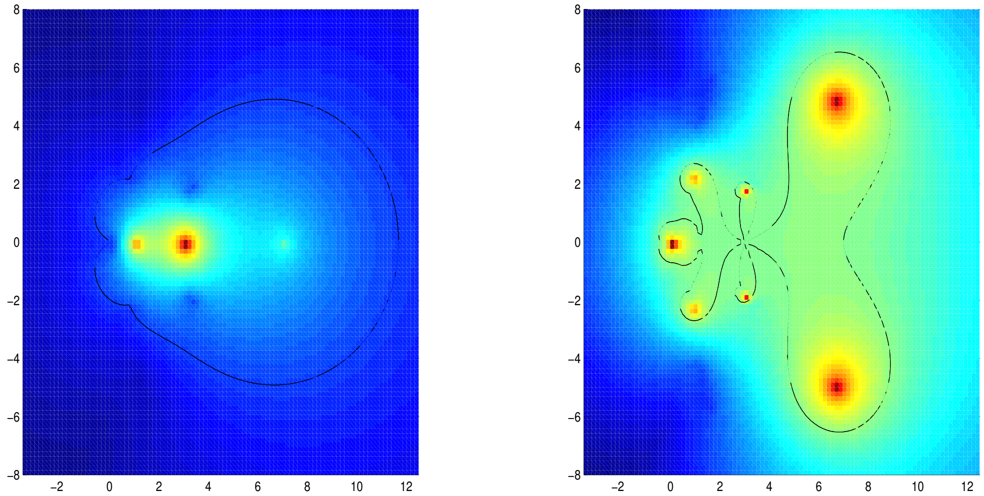
2D display for E of rank 2

Case i) A, E Case ii) $A - E, E$



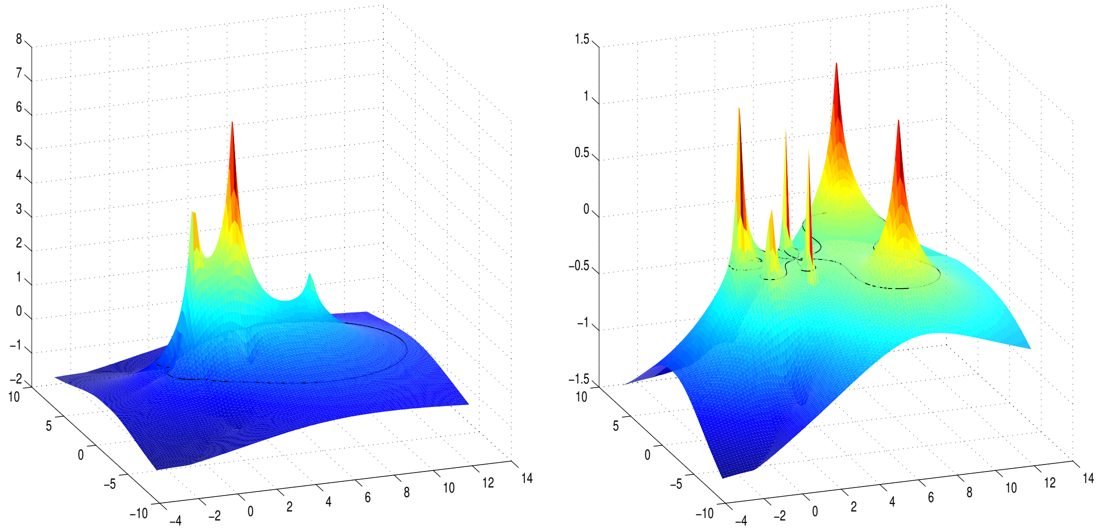
3D display for E of rank 2

Case i) A, E Case ii) $A - E, E$



2D display for E of rank 4

Case i) A, E Case ii) $A-E, E$



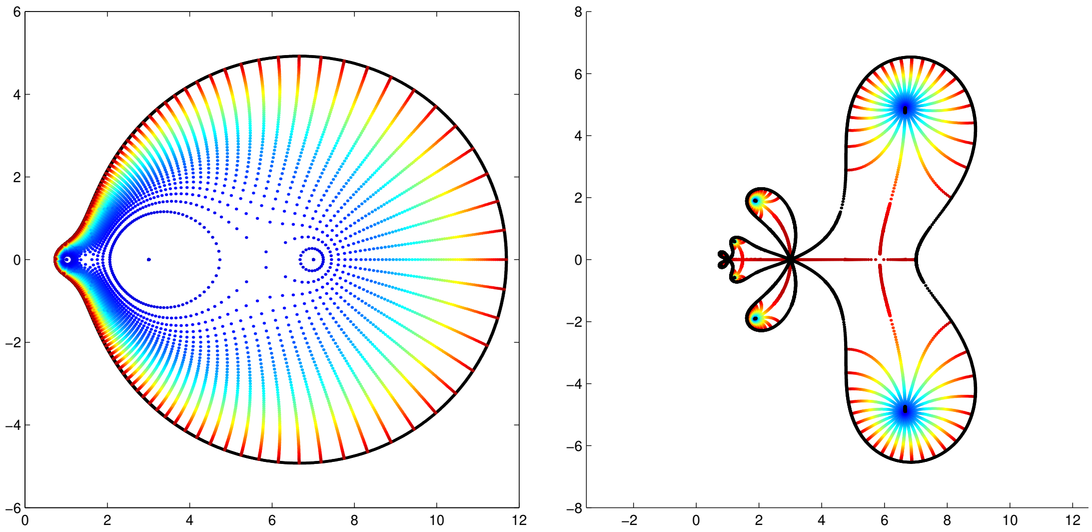
3D display for E of rank 4

6.6. A mesh of rays $\theta = 0, ..., 2\pi$

The segment $[0;2\pi]$ is evenly divided by 25 points which realize a discretization of size $\frac{\pi}{12}$. We plot the 24 rays $\Lambda(\theta_i)$ corresponding to the discrete values $\theta_i = i\frac{\pi}{12}$, $i = 0, 1, ..., 23$.

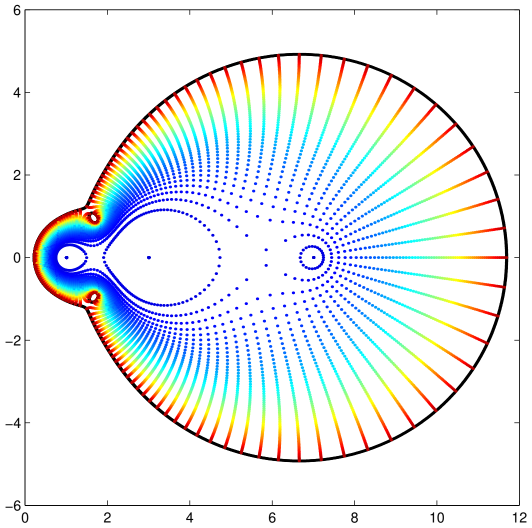
Case i) A, E Case ii) $A-E, E$

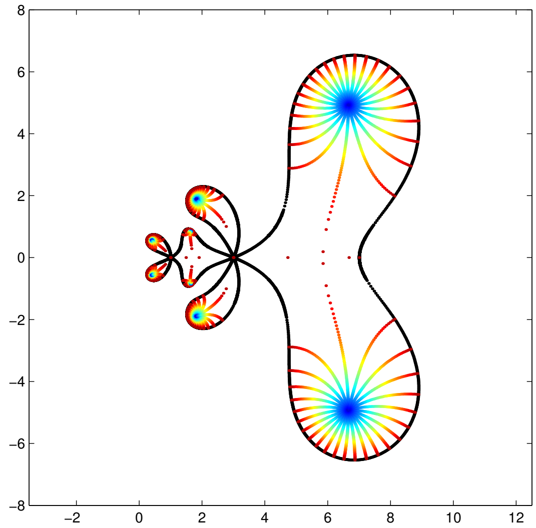
E of rank 1



Case i) A,E Case ii) $A-E,E$

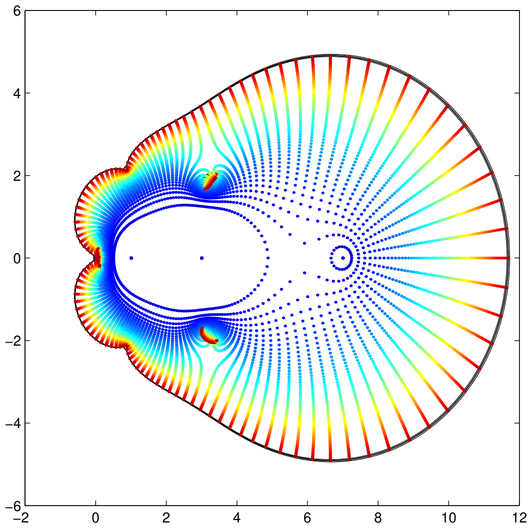
E of rank 2

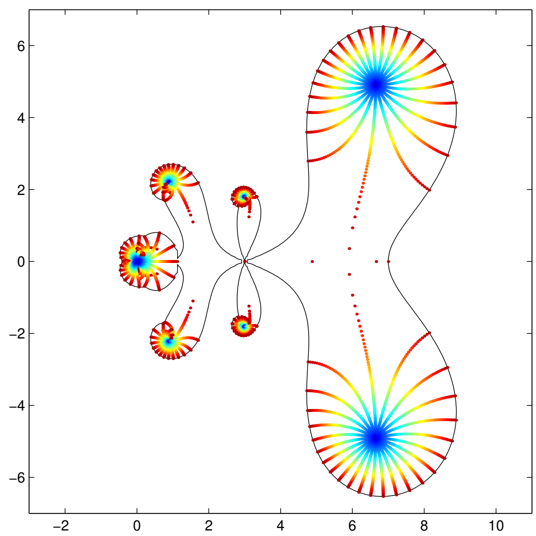




Case i) A,E Case ii) $A-E,E$

E of rank 4



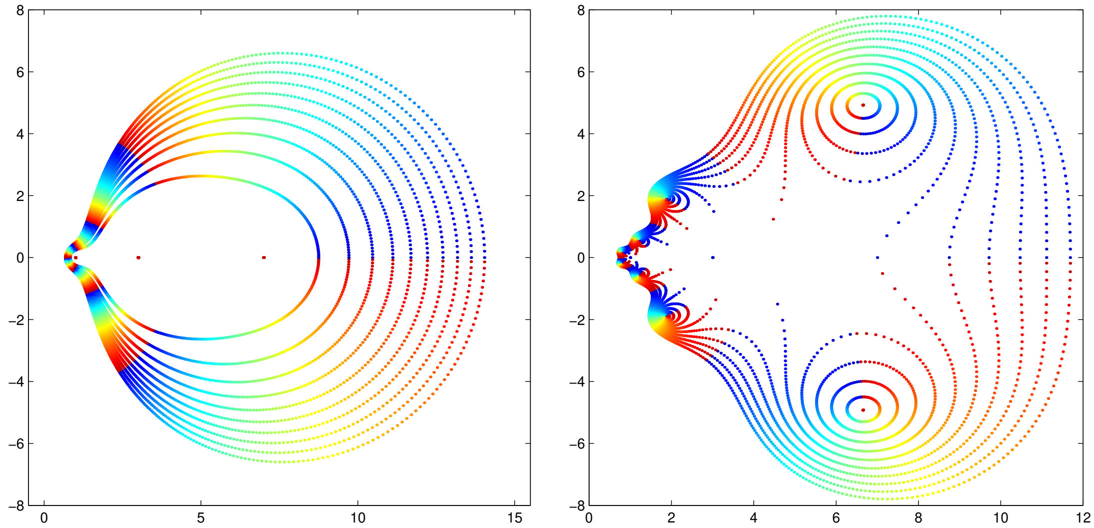


6.7. A mesh of orbits $r = 0, \ldots, 1, \ldots, 2$

The segment $[0;2]$ is evenly divided by 11 points such that the mesh size is 0.2. We plot the 10 orbits $\Sigma(r_i)$ corresponding to the discrete values $r_i = \frac{i}{5}$, $i = 1, 2, \ldots, 10$.

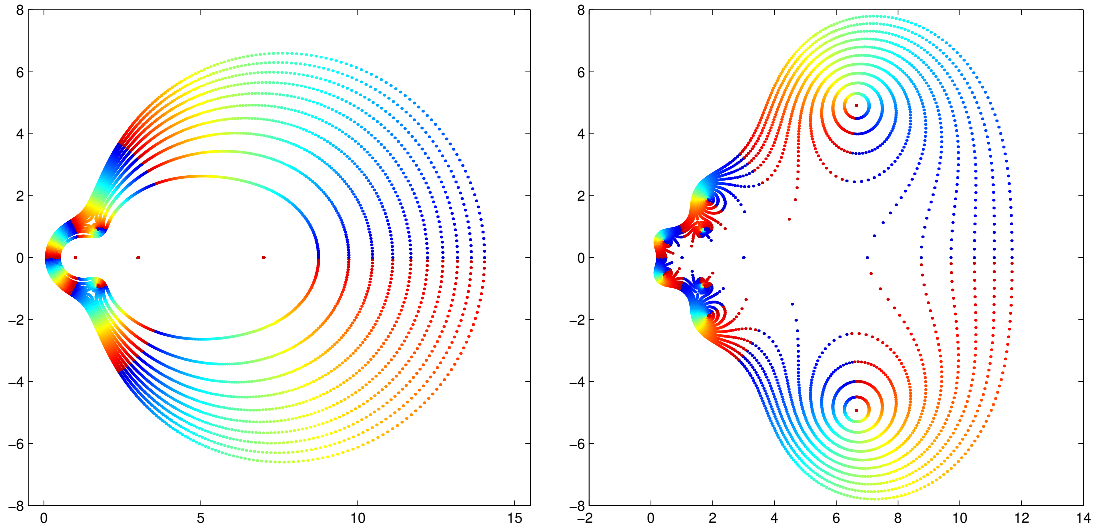
Case i) A, E Case ii) $A - E, E$

E of rank 1



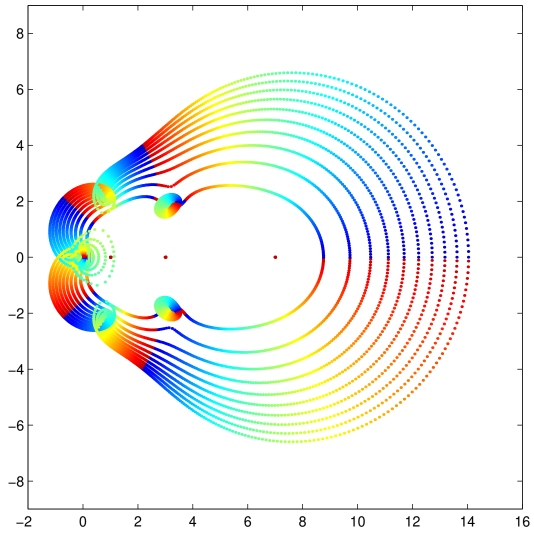
Case i) A, E Case ii) $A - E, E$

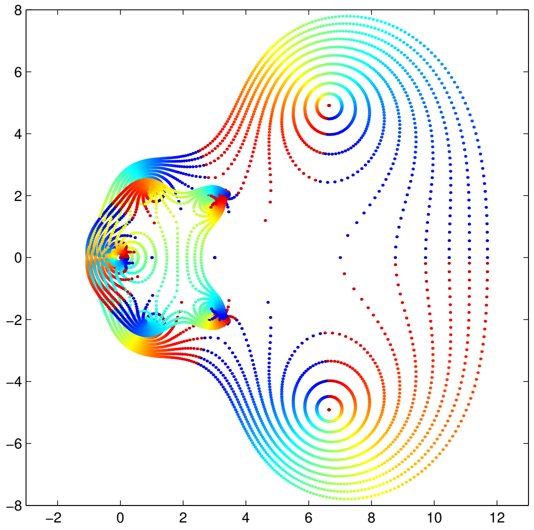
E of rank 2



Case i) A, E Case ii) $A-E, E$

E of rank 4



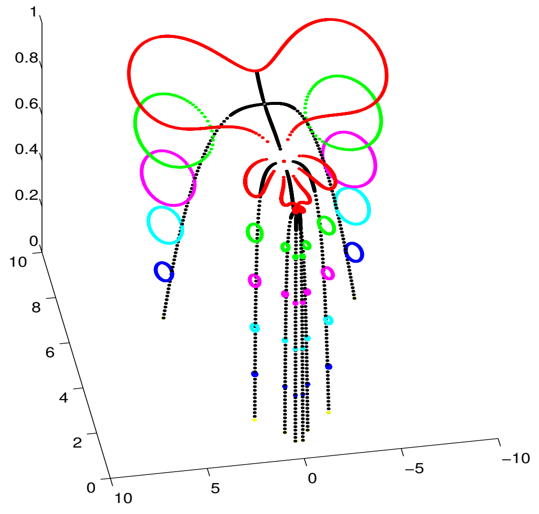


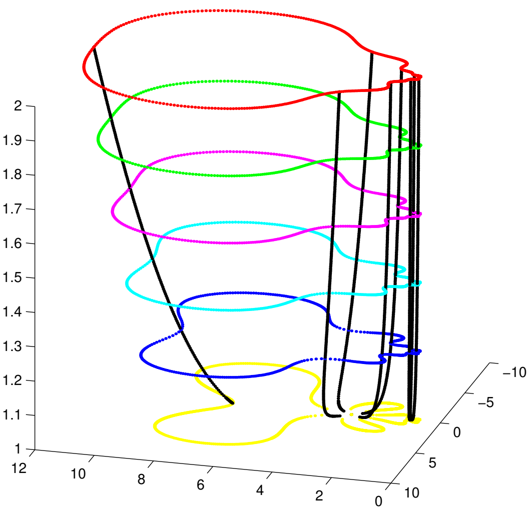
6.8. *Weaving rays and orbits*

We plot in 3D for r ranging in $[0;1]$ and $[1;2]$ 6 orbits along the 8 rays $\Lambda(0)$, corresponding to $\theta = 0$. The displays correspond to case ii) $A-E, E$.

$0 \leq r \leq 1$ $1 \leq r \leq 2$

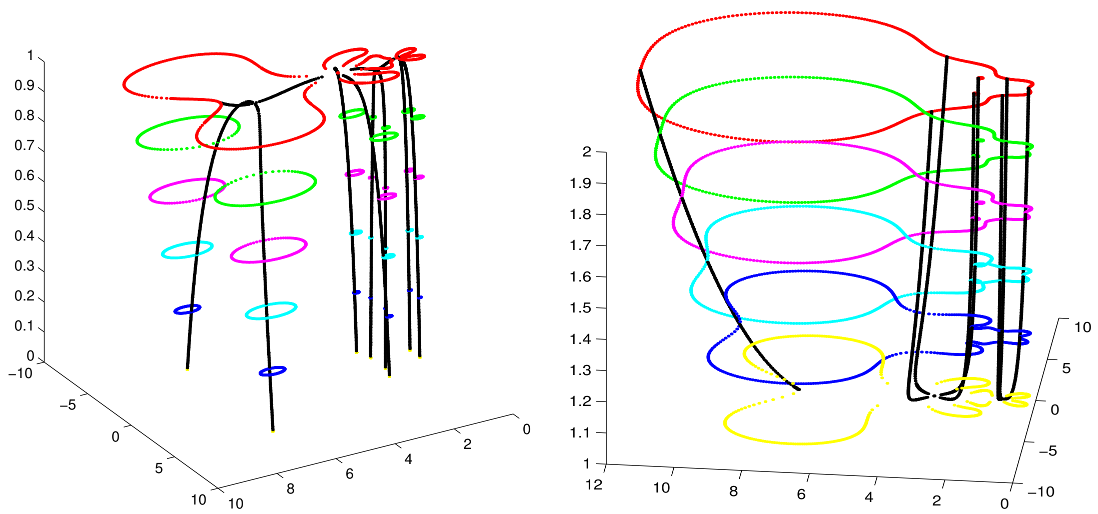
E of rank 1





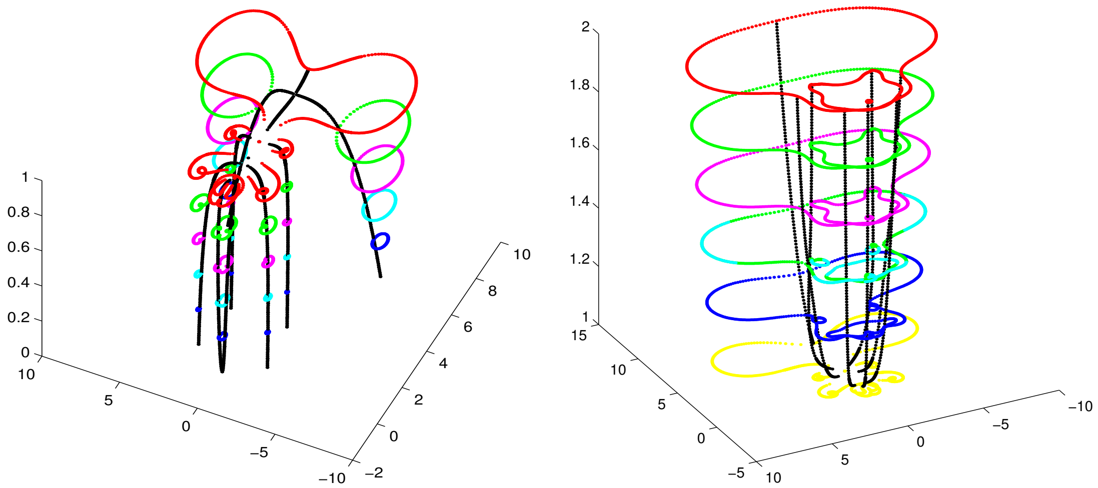
$0 \le r \le 1$ $\qquad$ $1 \le r \le 2$

E of rank 2



$0 \le r \le 1$ $\qquad$ $1 \le r \le 2$

E of rank 4



As θ varies in $[0;2\pi]$, the rays $\Lambda(\theta)$ create in 3D the surface whose r-level curves are the orbits $\Sigma(r)$.

6.9. A closer look at the loops of $\Sigma(1)$

We wish to draw the attention of the reader on the fact that, when $\Sigma(1)$ has loops, there is a *big difference* between an *internal loop* which is inside $\Psi\sigma$, but does not belong to Γ, and an *external loop* which is part of Γ. We study the 2 cases successively.

1. Γ *and* $\Sigma(1)$ *have an external loop*
   Such a case is exemplified by the matrix Venice in case ii) $A-E,E$, with E of all three ranks 1, 2

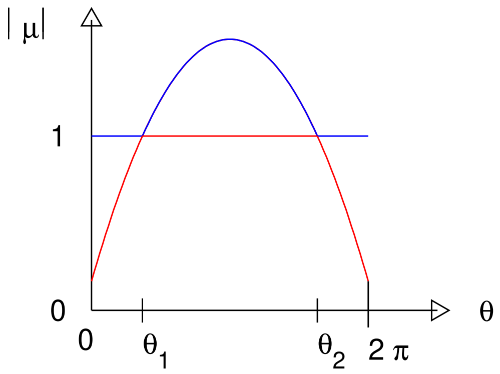


Figure 6.1. *Maps* $\theta \rightarrow \mid \mu_1(\theta) \mid$ *(blue),* $\mid \mu_2(\theta) \mid$ *(red)*

and 4. This is a consequence of the multiplicity of two external eigenvalues: the eigenvalues 1 and 3 of $(A-E)+E=A$ are of respective multiplicity 3 and 4. Each individual loop contains certain eigenvalues of $A-E$. It takes a multiple of 2π to describe such a loop.

2. $\Sigma(1)$ *has an internal loop which encloses no eigenvalue of A and is not a part of* Γ

This can happen only if rank $E>1$. Such an example is provided by Venice, case ii) $A-E, E$, for E *of rank 4*, where seven such loops are present. See also the zoom around 0 for case i) A, E. Such an occurrence is not anymore the result of the multiplicity of an eigenvalue of $A(\theta)=(A-E)+e^{i\theta}E$, but is rather due to the *coincidence* at z of an eigenvalue of a matrix $A(\theta_1)$ with that of a matrix $A(\theta_2)$, for $\theta_1 \neq \theta_2$.
This is easily explained with E of rank 2. Let $\mu_1(\theta)$ and $\mu_2(\theta)$ be the two eigenvalues of $F_\theta = e^{i\theta}E(A-zI)^{-1}$ which are not necessarily zero. We suppose that $\mid \mu_1(\theta) \mid \geq \mid \mu_2(\theta) \mid$. The representation $\theta \rightarrow \mid \mu_1(\theta) \mid, \mid \mu_2(\theta)$ is sketched in Figure 6.1. The equality $\mid \mu_1(\theta) \mid = \mid \mu_2(\theta) \mid$ for $\theta_1 \neq \theta_2$ (mod. 2π) is possible because E is of rank 2. This implies that $A(\theta_1)$ and $A(\theta_2)$ share the same eigenvalue z. The same analysis can be carried on any orbit $\Sigma(r)$, $r>0$. So we introduce the

Definition 6.1. *We call* synchronicity *the coincidence at z of two eigenvalues on the same orbit* $\Sigma(r)$ *corresponding to two different phases* $\theta_1 \neq \theta_2$ *(mod.* 2π*).*

Because $\mid \theta_1 - \theta_2 \mid < 2\pi$, it takes θ less than 2π to go around the internal loop.

The family $\Lambda(\theta)$ of n rays has the global period 2π in θ and it divides any orbit $\Sigma(r)$ into n curvilinear segments. Each individual component of an orbit $\Sigma(r)$, $r>0$, is a curve which is divided in as many curvilinear segments as there are eigenvalues of A (counting their multiplicity) present inside the individual component.

7. Computation in finite precision

7.1. *Inexact Computing in finite precision*

We expect finite precision to have a great influence on the solution of any system of matrix $A+tE-zI$, when z is in a region of $\mathbb{C}$ where lie the critical points. This is because in this region, $(A+tE-zI)^{-1}$ has a finite representation, rather than an asymptotic one.

Although the (at most) $n-1$ points of superconvergence form a *non generic* set of singularities in exact arithmetic, their influence in finite precision can be **important**. Indeed, finite precision computation expresses the point of view of **Lebesgue measure** rather than Baire category (genericity). We know that the neighborhood of any exact root of a polynomial is a set of positive Lebesgue measure, the greater the measure (with respect to machine precision), the more ill conditioned the root is.

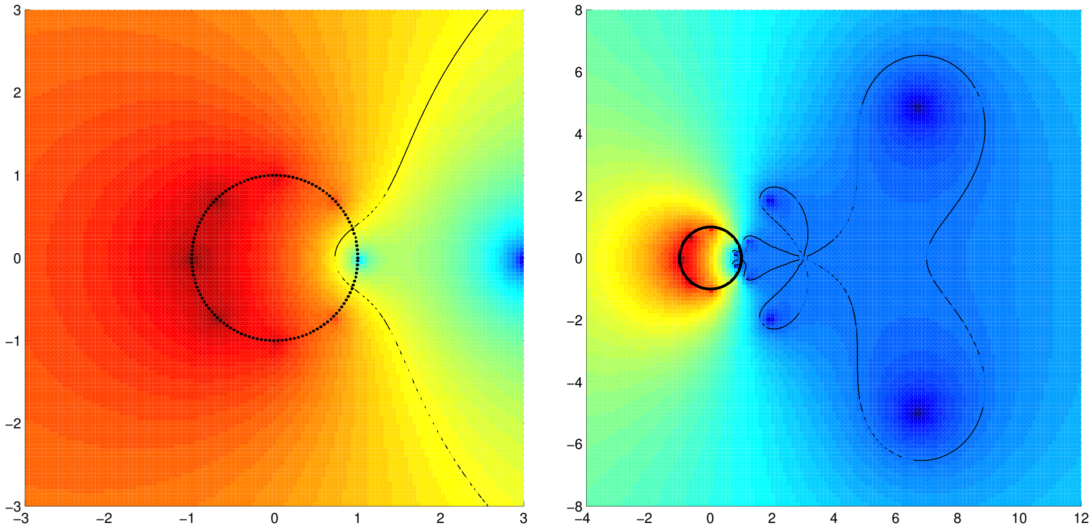
We examine in greater detail two striking examples of finite precision computation in the region around the critical points, with a deviation matrix E of rank 1.

7.2. *Venice in finite precision*

We study the matrix Venice with $E = ee_8^T$ of rank 1. The points of superconvergence are the 7 roots of $p_7(z) = \sum_{i=0}^{7} z^i$, that is the 7 roots of $z^8 = 1$ different from 1.

Case i) A, E Case ii) $A-E, E$

Map $z \longrightarrow \log\ r,\ r = 1/\rho$

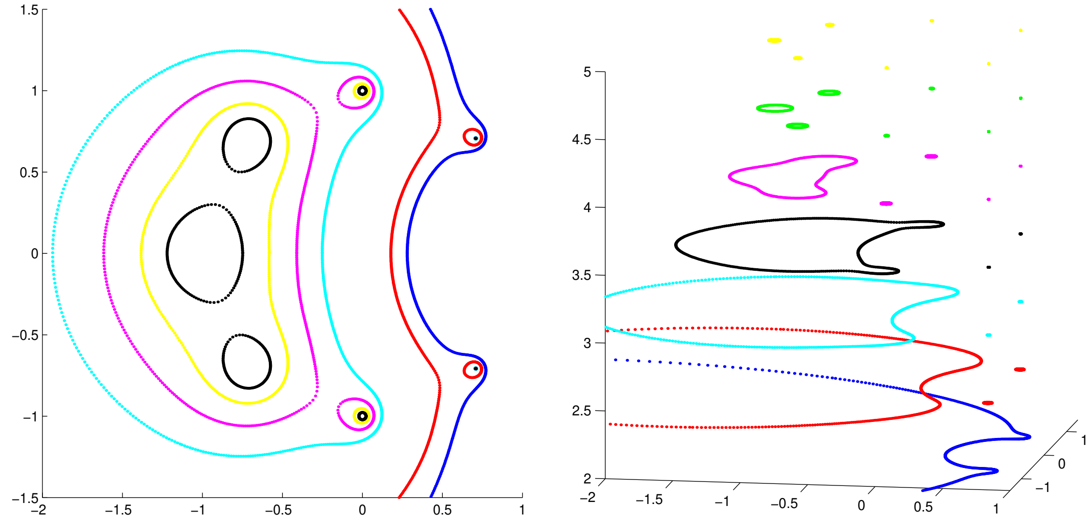


Unit circle in black

z	$e^{\pm i\frac{\pi}{4}}$	$\pm i$	$e^{\pm i\frac{3\pi}{4}}$	-1
$\tilde{\rho}$	1.3×10^{-10}	1.8×10^{-11}	3.8×10^{-12}	6.7×10^{-13}

Table 7.1. $\rho(E(A-zI)^{-1})$ computed at critical points

Case i) A, E



2D, $r = \{100, 200, 2000, 4000, 8000, 16000\}$ 3D, log $r = \{2, 2.5, 3, 3.5, 4, 4.5, 5\}$

The computed values $\tilde{\rho}(E(A-zI)^{-1})$ are listed in Table 7.1 for the 7 critical values of z.

The 7 critical points are on the unit circle: they are singular points for the map $\overline{\varphi}:\ z \to \log\ r$ with $r = 1/\rho$, so that $\overline{\varphi}$ is equivalent to the map $-\varphi:\ z \to -\log\ \rho$. The 2D displays of $\overline{\varphi}$ clearly show the influence of points on the unit circle, different from 1. The map $\overline{\varphi}$ is further displayed for case i) A, E for increasing values of r, both in 2D and 3D.

7.3. The family One(n) in finite precision

We consider the family of polynomials $p_n(x) = x^n$ and $q_n(x) = \sum_{i=0}^{n} x^i = p_n(x) + q_{n-1}(x)$, of degree $n = 1, 2, \ldots$, with $q_0(x) = 1$.
The companion matrix associated with $p_n(x)$ is J_n^T, where J_n is the Jordan block of order n:

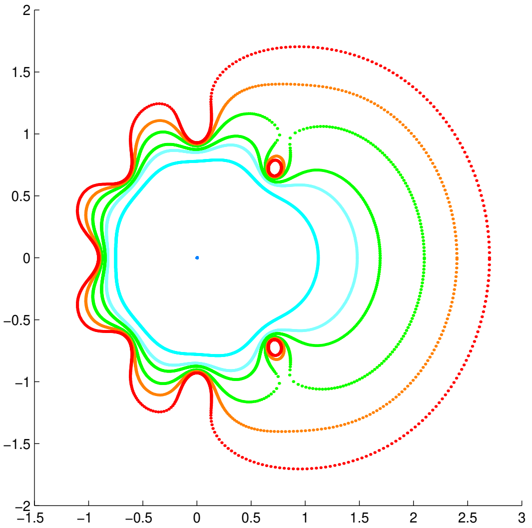
$$J_n = \begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix}.$$

Let $E_n = ee_n^T$, with $e = (1, \ldots, 1)^T$. The family **One**(n) is the family $A_n(t) = J_n^T + tE_n$ defined by (J_n^T, E_n). Clearly $A_n(-1) = J_n^T - E_n$ is the companion matrix associated with $q_n(x)$. The critical points of the unfolding are the $n-1$ roots of $q_{n-1}(x)$.

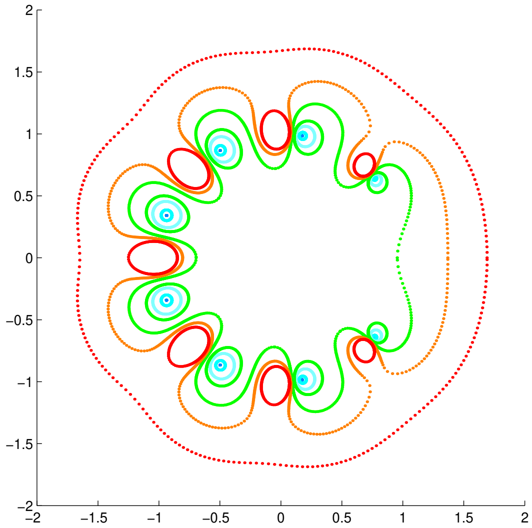
1. $n = 8$
   We give for case i) J_8^T, E_8 the 2D display of $\overline{\varphi}$ together with the 7 orbits $\Sigma(r)$ corresponding to

the values $r = \{0; 0.2; 0.5; 0.7; 1.1; 1.4; 1.7\}$. The local influence of the critical points is striking: around zero ($r \ll 1$) the orbits are almost perfect circles which get modified when r increases. For comparison, the reverse unfolding corresponding to the case ii) $J_8^T - E_8, E_8$ is also given. The 7 values correspond to $r = \{0; 0.2; 0.5; 0.7; 1.1; 1.4; 1.7\}$.

Case i) J_8^T, E_8



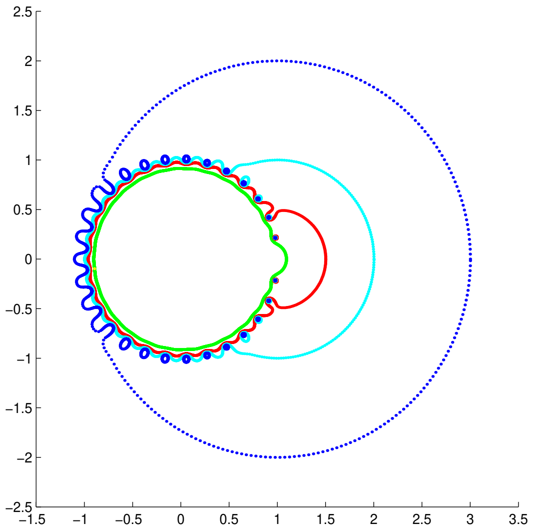
Case ii) $J_8^T - E_8, E_8$



$r = \{0; 0.2; 0.5; 0.7; 1.1; 1.4; 1.7\}$

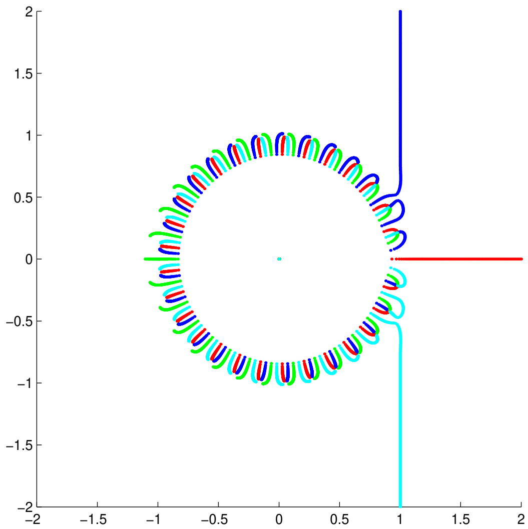
2. $n = 29$
   We consider case i) J_{29}^T, E_{29}. On the left is a display of $\overline{\varphi}$ with 4 orbits $\Sigma(r)$ corresponding to $r = 0.1,\ 0.5,\ 1$ and 2.
   On the right is a display of 4 rays $\Lambda(\theta)$ for $\theta = 0,\ \frac{\pi}{2},\ \pi,\ \frac{3\pi}{2}$. The local influence of the critical points is clearly visible: outside the unit circle the rays are just *straight lines*. Note that $\Lambda(\pi)$ does not extend horizontally much beyond $(-1, 0)$.

Case i) J_{29}^T, E_{29}



$r = \{0.1; 0.5; 1; 2\}$

Case ii) $J_{29}^T - E_{29}, E_{29}$



$\theta = \{0, \frac{\pi}{2}; \pi; \frac{3\pi}{2}\}$

8. Related work realized in the Qualitative Computing Group

As was shown in the introductory survey, the notion of homotopic perturbation is a very powerful tool to investigate theoretical questions related to the spectral approximation of matrices or linear operators in Banach spaces. It has therefore been a key notion, explicitly or implicitly, for all the works developed by the Qualitative Computing Group at CERFACS, during more than a decade. The list of topics, given below in chronological order, reflects this underlying influence.

1. The notion of *uniform radial convergence* ([15], p. 245) involves the spectral radius $r_\sigma[(T-T_n)(T-z)^{-1}]$ which is nothing but, in a different notation, $\rho(E(A-zI)^{-1})$ for z in a compact subset of the resolvent set of T (resp. A).

2. Computation in finite precision induces an unavoidable round-off error which has been analyzed by a backward error analysis ([6], chapter 5 and 6 in particular). Finite precision computation can be seen as *uncertain computation*, a particular case of inexact computation, where the deviation matrix E is not known, only its size $\|E\|$ is available as a measure of the backward error [6,10]. It is important to keep in mind that in finite precision, the norms have to be *scaled* because the notion of backward error is *relative to machine precision* [6,10].

3. The toolbox PRECISE (**PR**ecision **E**stimation and **C**ontrol **I**n **S**cientific and **E**ngineering computing) offers a variety of software tools developed under Matlab [6] and Fortran [21] to conduct computer experiments about stability. A detailed presentation can be found in [6], chapters 8 and 9, and in the user's guide of the Fortran version [21]. The key idea is to allow the size of the perturbation to vary in $]0,1]$ for $A(t)=A+tE$.

4. Finite precision computation of the Jordan form of a matrix is known to be unreliable because the exact Jordan structure is unstable, due to the underlying algebraic singularities. Homotopic perturbations has been successfully put at use in [20] to assess the most influent Jordan structure of a matrix [8].

5. When dealing with linear operators T in infinite dimensional spaces the question arises similarly whether the maps $z \to \rho(E(T-z)^{-1})$ and $z \to \| (T-z)^{-1} \|$ can be constant on an open domain of $\mathbb{C}\setminus\sigma(T)$. The question has been only partially answered so far [7,19] : the subharmonicity is true in a Hilbert space, and it is not yet known if it is true in general in a Banach space.

6. The incomplete Arnoldi decomposition can also be interpreted in the context of homotopic perturbation [10,24]. See also [23].This leads to a novel way of interpreting the stopping criterion in Krylov methods [4,10,14,11].

7. When the outer iteration is a Krylov-type method, inner-outer iterations exhibit a surprising robustness to perturbations due to an inexact resolution in the inner loop [1]. In this context too, the notion of homotopic perturbation is a big help to analyze this seemingly "magical" behavior [9].

8. In the homotopic viewpoint that we presented, the inexact pseudo-spectrum is structured by means of the eigenvalues of $A(t)$ of order n. Such eigenvalues are the roots of polynomials of degree n and, as such, have been considered as n complex numbers (i.e. 2D real vectors). However polynomials can have more roots, indeed infinitely many, if considered as quaternions (4D real vectors)

or octonions (8D real vectors) [3]. The importance of considering *hypercomplex* roots is discussed in [13].

9. Conclusion

This work has discussed the meaning which emerges from the linear relation between two matrices A and B, done via the complex parameter t such that $A(t) = A + t(B - A)$. Different meanings can be associated with different properties of the resolvent field $z \to (A(t) - zI)^{-1}$. The properties have to do with existence or analyticity of the resolvent. Any $z \in \mathbb{C}$ which is not an eigenvalue of $A(t)$ can be interpreted either as an eigenvalue of $A(t)$, or equally as a z such that $(A(t) - zI)^{-1}$ exists. The only instance where z cannot be interpreted as an eigenvalue of $A(t)$ is if it is a critical point of (A, E). In this case $(A(t) - zI)^{-1}$ has a finite representation in terms of t, rather than an asymptotic one.

The second idea which emerges from the theory of Inexact Computing that is presented is the importance of the duality of viewpoints which can be taken in mathematics with respect to singularities.

When continuity is the prevalent point of view, a finite set of singularities is *nongeneric* and can be ignored: this is the basis for the elimination of singularities for continuous subharmonic functions ([2], p. 303). This is a *global* point of view. However, *locally* the singularities do have an influence when computation is inexact, such as in finite precision. Their neighborhood is of *positive Lebesgue measure*. The homotopic unfolding of the resolvent field $z \to (A - zI)^{-1}$ is an extremely rich domain, where we are able to follow the subtle interplay between the continuous and the discrete, through the action of computation.

10. Annex

10.1. Convergence of the Neumann series of a matrix

We consider the matrix $M = I + N$ and the condition of its invertibility. Let ν be any eigenvalue of N. It is clear that M is invertible iff $\nu \neq -1$ for all eigenvalues ν of N.
The condition $\rho(N) < 1$ implies that $|\nu| < 1$. Therefore M is invertible and the Neumann series

$$(I+N)^{-1} = I + \sum_{k=1}^{\infty} (-N)^k$$

converges.
When $\rho(N) \geq 1$, M is still invertible iff $\nu \neq -1$. But the Neumann series does not converge because $N^k \nrightarrow 0$ as $k \to \infty$. $\rho(N) = 1$ is the radius of convergence of the Neumann series (See [15], theorem 2.21, p.101-102). The Neumann series expansion is the matrix equivalent to the geometric series expansion for a scalar of modulus less than 1: let $m = n+1$, then $m^{-1} = \frac{1}{1+n} = \sum_{k=0}^{\infty}(-n)^k$ iff $|n| < 1$.

There is a significant difference between the scalar and matrix case. For a scalar, $|n| = 0 \Leftrightarrow n = 0$, but for a matrix, $\rho(N) = 0$ does not implies $N = 0$: any nonzero **nilpotent** matrix satisfies:

$$\rho(N) = 0 \Leftrightarrow \exists l > 1 \text{ such that } N^l = 0.$$

Therefore the Neumann series expansion becomes **finite**:

$$(I+N)^{-1} = I + \sum_{1}^{l-1} (-N)^k.$$

References

1. A. Bouras, F. Chaitin-Chatelin, and V. Frayssé. Solveurs itératifs imbriqués. Contract Report FR/PA/00/19, CERFACS, 2000.
2. B. Chabat. *Introduction à l'analyse complexe, Tome 1.* Mir, Moscow, 1990.
3. F. Chaitin-Chatelin. The computing power of Geometry. In D. F. Griffiths and G. A. Watson, editors, *Numerical Analysis 1999, p. 83–92*. CRC Press LLC, 1999.
4. F. Chaitin-Chatelin. Comprendre les méthodes de Krylov en précision finie : le programme du Groupe Qualitative Computing au CERFACS. Technical Report TR/PA/00/11, CERFACS, 2000.
5. F. Chaitin-Chatelin. Unfolding algebraic singularities : the eigenvalue problem, June 2-6, 1999. Invited talk, presented by A. Ilahi, Second workshop on Large-Scale scientific computations, Sozopol, Bulgaria.
6. F. Chaitin-Chatelin and V. Frayssé. *Lectures on Finite Precision Computations*. SIAM, Philadelphia, 1996.
7. F. Chaitin-Chatelin and A. Harrabi. About definitions of pseudospectra of closed operators in Banach spaces. Technical Report TR/PA/98/08, CERFACS, 1998.
8. F. Chaitin-Chatelin, A. Harrabi, and A. Ilahi. About Hölder condition numbers and the stratification diagram for defective eigenvalues. *Mathematics and Computers in Simulation*, 54:397–402, 2000.
9. F. Chaitin-Chatelin and T. Meškauskas. Inner-outer iterations for mode solvers in structural mechanics: application to the Code-aster. Contract Report FR/PA/01/85, CERFACS, 2001.
10. F. Chaitin-Chatelin, V. Toumazou, and E. Traviesas. Accuracy assessment for eigencomputations : variety of backward errors and pseudospectra. *J. Linear Algebra Appl.*, 309:73–83, 2000.

11. F. Chaitin-Chatelin and E. Traviesas. The Arnoldi method in the light of homotopic perturbation. Work in progress.

12. F. Chaitin-Chatelin and E. Traviesas. On the superconvergence of geometric series expansions of matrices. Work in progress.

13. F. Chaitin-Chatelin and E. Traviesas. Qualitative computing. Technical Report TR/PA/02/??, CERFACS, 2002. To appear in the "Handbook on Accuracy and Reliability in Scientific Computing" published by IFIP WG 2.5, as a SIAM book in the series "Software, Environments, and Tools".

14. F. Chaitin-Chatelin, E. Traviesas, and L. Plantié. Understanding Krylov methods in finite precision. Technical Report TR/PA/00/40, CERFACS, 2000. Also in Proceedings of the Second Conference on Numerical Analysis and Applications, Rousse, Bulgaria, Springer, pp. 187–197.

15. F. Chatelin. *Spectral approximation of linear operators*. Academic Press, New York, 1983.

16. F. Chatelin. *Valeurs propres de matrices*. Masson, Paris, 1988.

17. F. Chatelin. *Eigenvalues of matrices*. Wiley, Chichester, 1993. Enlarged Translation of the French Publication with Masson.

18. E. Gallestey. *Theory and numerics of spectral value sets*. PhD thesis, University of Bremen, 1998.

19. A. Harrabi. Pseudospectre d'une suite d'operateurs bornés. *Mathematical Modelling and Numerical Analysis*, 32(6):671–680, 1998.

20. A. Ilahi. *Validation du calcul sur ordinateur: application de la théorie des singularités algébriques*. Ph.D. dissertation, Université Toulouse I, June 1998. TH/PA/98/31, CERFACS.

21. R. A. McCoy and V. Toumazou. PRECISE User's Guide - Version 1.0. Technical Report TR/PA/97/38, CERFACS, 1997.

22. B. N. Parlett. Private communication, May 2000.

23. V. Simoncini. Remarks on non-linear spectral perturbation. *BIT*, 39:350–365, 1999.

24. E. Traviesas. *Sur le déploiement du champ spectral d'une matrice*. Ph.D. dissertation, Université Toulouse I, TH/PA/00/30, CERFACS, Mai 2000.

25. L. N. Trefethen. Pseudospectra of linear operators. *SIAM Review*, 39:383–406, 1997.

26. L. N. Trefethen. Computation of Pseudospectra. *Acta numerica*, 8:247–295, 1999.