Vancouver, 8 August 2013

Differential equations for the approximation of the distance to the closest defective matrix.

N. Guglielmi (L'Aquila)

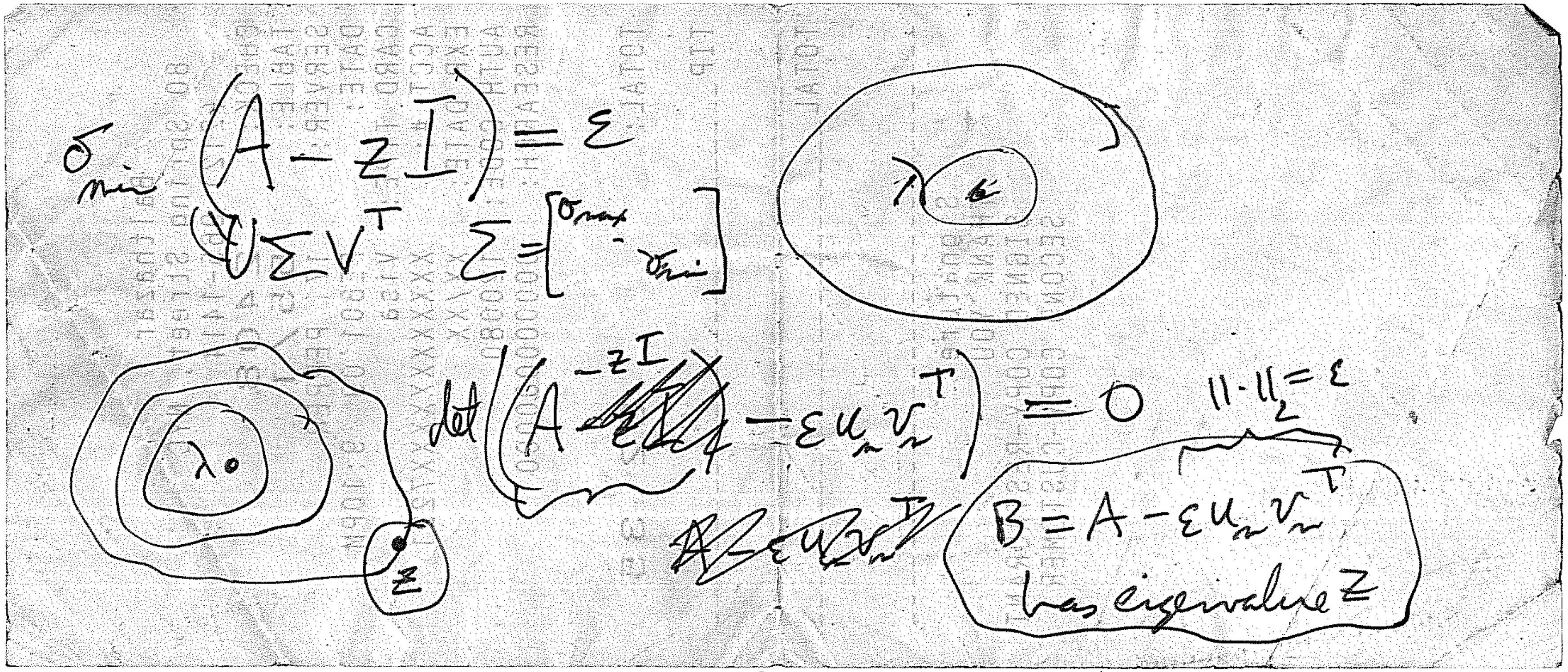
and

M. Manetta (L'Aquila), P. Buttà, S. Noschese (Roma)

Dedicated to Michael Overton.

Preamble

In Winter 2009 I visited Michael; during a party at Courant, I asked Michael how to obtain extremal perturbations associated to a boundary point in the ε-pseudospectrum . . .



This is his answer on a receipt of Whole Foods.

Summary

- Problem and literature.
- Low-rank odes and extremal pseudo-eigenvalues.
- Theoretical properties and examples.
- Extension to structured problems.

Problem

Framework: Let $A \in \mathbb{K}^{n,n}$ ($\mathbb{K} = \mathbb{C}$ or $\mathbb{K} = \mathbb{R}$) a matrix with all distinct eigenvalues. We denote by $\Lambda(A)$ the spectrum of A.

Problem

Framework: Let $A \in \mathbb{K}^{n,n}$ ($\mathbb{K} = \mathbb{C}$ or $\mathbb{K} = \mathbb{R}$) a matrix with all distinct eigenvalues. We denote by $\Lambda(A)$ the spectrum of A.

The distance to defectivity is defined as

$$w_{\mathbb{K}}(A) \; = \; \inf\left\{ \|A - B\| : \, B \in \mathbb{K}^{n,n} \text{ is defective} \right\}$$

where, in this talk, $\| \cdot \|$ denotes here the Frobenius norm.
If $\mathbb{K} = \mathbb{C}$ the 2-norm is equivalent, that means $w_{\mathbb{K}}(A)$ is the same number; but this not true in general for $\mathbb{K} = \mathbb{R}$.

Problem

Framework: Let $A \in \mathbb{K}^{n,n}$ ($\mathbb{K} = \mathbb{C}$ or $\mathbb{K} = \mathbb{R}$) a matrix with all distinct eigenvalues. We denote by $\Lambda(A)$ the spectrum of A.

The distance to defectivity is defined as

$$w_{\mathbb{K}}(A) \;=\; \inf\left\{ \|A - B\| : \, B \in \mathbb{K}^{n,n} \text{ is defective} \right\}$$

where, in this talk, $\| \cdot \|$ denotes here the Frobenius norm.
If $\mathbb{K} = \mathbb{C}$ the 2-norm is equivalent, that means $w_{\mathbb{K}}(A)$ is the same number; but this not true in general for $\mathbb{K} = \mathbb{R}$.

Generically we expect that an extremizer $B_{\text{opt}} \in \mathbb{K}^{n,n}$ (if exists) has a coalescent defective pair of eigenvalues.

Some literature

First $w_{\mathbb{C}}(A)$ was introduced by **Demmel** (1983) in his very well-known PhD thesis under the name $diss(A, path)$, $path$ referring to the path traveled by the eigenvalues in the complex plane under a smoothly varying perturbation to A.

Some literature

First $w_{\mathbb{C}}(A)$ was introduced by **Demmel** (1983) in his very well-known PhD thesis under the name $diss(A, path)$, $path$ referring to the path traveled by the eigenvalues in the complex plane under a smoothly varying perturbation to A.

The problem was investigated at the same time by **Wilkinson** (1984), who referred to it as a problem of primary interest.

Some literature

First $w_{\mathbb{C}}(A)$ was introduced by **Demmel** (1983) in his very well-known PhD thesis under the name $diss(A, path)$, $path$ referring to the path traveled by the eigenvalues in the complex plane under a smoothly varying perturbation to A.

The problem was investigated at the same time by **Wilkinson** (1984), who referred to it as a problem of primary interest.

The very interesting recent article by **Alam, Byers, Bora & Overton** (2011) shows that for $\mathbb{K} = \mathbb{C}$ the infimum is indeed a minimum. For approximating $w_{\mathbb{C}}(A)$, they also proposed an algorithm which is well-suited to problems of moderate size.

Some literature

First $w_{\mathbb{C}}(A)$ was introduced by **Demmel** (1983) in his very well-known PhD thesis under the name $diss(A, path)$, $path$ referring to the path traveled by the eigenvalues in the complex plane under a smoothly varying perturbation to A.

The problem was investigated at the same time by **Wilkinson** (1984), who referred to it as a problem of primary interest.

The very interesting recent article by **Alam, Byers, Bora & Overton** (2011) shows that for $\mathbb{K} = \mathbb{C}$ the infimum is indeed a minimum. For approximating $w_{\mathbb{C}}(A)$, they also proposed an algorithm which is well-suited to problems of moderate size.

Apparently the case $\mathbb{K} = \mathbb{R}$ is unexplored. Similarly there seem to be no methods to approximate any structured distance.

Methodology: two steps

(i) For a given ε we aim to approximate the quantity

$$r(\varepsilon) = \min\Big\{ y^* x : y \text{ and } x \text{ left/right eigenvectors to } \lambda \in \Lambda\left(A + \varepsilon E\right) \text{ for some } E : \ \|E\| \leq 1 \Big\},$$

with x and y normalized as: $\|x\| = \|y\| = 1, y^* x \geq 0$.

Connection: ε-pseudospectrum (**Trefethen & Embree** (2005))

Methodology: two steps

(i) For a given ε we aim to approximate the quantity

$$r(\varepsilon) = \min\Big\{ y^* x : y \text{ and } x \text{ left/right eigenvectors to } \lambda \in \Lambda\left(A + \varepsilon E\right) \text{ for some } E : \ \|E\| \le 1 \Big\},$$

with x and y normalized as: $\|x\| = \|y\| = 1, y^* x \ge 0$.

Connection: ε-pseudospectrum (**Trefethen & Embree** (2005))

(ii) In order to approximate

$$w_{\mathbb{K}}(A) = \varepsilon^* = \min\{\varepsilon : \ r(\varepsilon) = 0\},$$

we look for *locally minimal* solutions ε to $r(\varepsilon) = 0$.

Methodology: two steps

(i) For a given ε we aim to approximate the quantity

$$r(\varepsilon) = \min\Big\{ y^* x : y \text{ and } x \text{ left/right eigenvectors to } \lambda \in \Lambda\,(A + \varepsilon E) \text{ for some } E : \ \|E\| \le 1 \Big\},$$

with x and y normalized as: $\|x\| = \|y\| = 1, y^* x \geq 0$.

Connection: ε-pseudospectrum (**Trefethen & Embree** (2005))

(ii) In order to approximate

$$w_{\mathbb{K}}(A) = \varepsilon^* = \min\{\varepsilon : \ r(\varepsilon) = 0\},$$

we look for *locally minimal* solutions ε to $r(\varepsilon) = 0$.

Meaning. If $\mathbb{K} = \mathbb{C}$ at a *locally minimal* solution two discs in ε-pseudospectrum have a contact point (**Alam & Bora** (2005))

Also interesting to consider $r(\varepsilon) = \delta$ for a small threshold δ.

Constructing a path for the eigenvalues

Part (i): we construct a smooth matrix valued function

$$A + \varepsilon E(t) \qquad \text{where} \quad \|E(t)\| = 1.$$

Normalization: any selected pair of left/right eigenvectors of $A + \varepsilon E(t)$ is such that $\|x(t)\| = \|y(t)\| = 1,\ y(t)^* x(t) > 0$.

Constructing a path for the eigenvalues

Part (i): we construct a smooth matrix valued function

$$A + \varepsilon E(t) \qquad \text{where} \quad \|E(t)\| = 1.$$

Normalization: any selected pair of left/right eigenvectors of $A + \varepsilon E(t)$ is such that $\|x(t)\| = \|y(t)\| = 1,\ y(t)^* x(t) > 0$.

Desired properties

(a) the function $y(t)^* x(t)$ is decreasing:

(b) $\lim_{t \to \infty} E(t) = E_\infty$

(c) $y_\infty^* x_\infty$ *local* minimum of the function $y^* x(E) : \mathbb{K}^{n,n} \to \mathbb{R}^+$

Constructing a path for the eigenvalues

Part (i): we construct a smooth matrix valued function

$$A + \varepsilon E(t) \qquad \text{where} \quad \|E(t)\| = 1.$$

Normalization: any selected pair of left/right eigenvectors of $A + \varepsilon E(t)$ is such that $\|x(t)\| = \|y(t)\| = 1,\ y(t)^* x(t) > 0$.

Desired properties

(a) the function $y(t)^* x(t)$ is decreasing:

(b) $\lim_{t \to \infty} E(t) = E_\infty$

(c) $y_\infty^* x_\infty$ *local* minimum of the function $y^* x(E) : \mathbb{K}^{n,n} \to \mathbb{R}^+$

Idea: look for steepest descent direction $\dot{E}$ for $y(t)^* x(t)$, using

$$\frac{d}{dt}\left(y(t)^* x(t)\right) = \dot{y}(t)^* x(t) + y(t)^* \dot{x}(t).$$

Derivatives of eigenvectors

Proposition (**Meyer & Stewart** (1988))

Let the matrix $M(t)$ be smooth w.r.t. $t \in \mathbb{R}$, $\lambda(t)$ a simple eigenvalue with normalized left/right eigenvectors $y(t), x(t)$.

Derivatives of eigenvectors

Proposition (**Meyer & Stewart** (1988))

Let the matrix $M(t)$ be smooth w.r.t. $t \in \mathbb{R}$, $\lambda(t)$ a simple eigenvalue with normalized left/right eigenvectors $y(t), x(t)$.

Let $G(t)$ be the group-inverse of $M(t) - \lambda(t)I$, i.e. the inverse of $M(t) - \lambda(t)I$ in the maximal multiplicative subgroup containing $M(t) - \lambda(t)I$.

Derivatives of eigenvectors

Proposition (**Meyer & Stewart** (1988))

Let the matrix $M(t)$ be smooth w.r.t. $t \in \mathbb{R}$, $\lambda(t)$ a simple eigenvalue with normalized left/right eigenvectors $y(t), x(t)$.

Let $G(t)$ be the group-inverse of $M(t) - \lambda(t)I$, i.e. the inverse of $M(t) - \lambda(t)I$ in the maximal multiplicative subgroup containing $M(t) - \lambda(t)I$.

Then the following hold:

$$\dot{x} = x^* G \dot{M} x x - G \dot{M} x$$

$$\dot{y}^* = y^* \dot{M} G y y^* - y^* \dot{M} G$$

where we omit the explicit dependence on t.

Steepest descent direction lemma

Let y and x left and right eigenvectors of $A + \varepsilon E$ associated to λ and G the group-inverse of $A + \varepsilon E - \lambda I$. Then set

$$S = yy^* G^* + G^* xx^* \ .$$

Let $\mathcal{B}$ the unit ball of the Frobenius norm.

Steepest descent direction lemma

Let y and x left and right eigenvectors of $A+\varepsilon E$ associated to λ and G the group-inverse of $A+\varepsilon E-\lambda I$. Then set

$$S = yy^*G^* + G^*xx^* \ .$$

Let $\mathcal{B}$ the unit ball of the Frobenius norm.
Then (1) for any smooth path $E(t) \in \mathcal{B}$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\,(y^*x) \;\; = \;\; \varepsilon\, y^*x\,\mathrm{Re}\;\langle \dot{E}, S\rangle \ .$$

where $\langle A, B\rangle = \mathrm{trace}\,(A^*B)$ is the Frobenius inner product.

Steepest descent direction lemma

Let y and x left and right eigenvectors of $A + \varepsilon E$ associated to λ and G the group-inverse of $A + \varepsilon E - \lambda I$. Then set

$$S = yy^* G^* + G^* xx^* \ .$$

Let $\mathcal{B}$ the unit ball of the Frobenius norm.
Then (1) for any smooth path $E(t) \in \mathcal{B}$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\,(y^* x) \;\; = \;\; \varepsilon\, y^* x\, \mathrm{Re}\, \langle \dot{E}, S \rangle \ .$$

where $\langle A, B \rangle = \mathrm{trace}\,(A^* B)$ is the Frobenius inner product.

Moreover (2) the steepest descent direction for $y^* x$ in the tangent hyperplane $T_E \mathcal{B}$ is given by

$$\dot{E} = D = -\mu \big(S - \mathrm{Re}\, \langle E, S \rangle E \big) \quad \text{with } \mu \text{ normalizing factor.}$$

Steepest descent ode

We consider the ODE

$$\dot{E} = -\big(S - \mathrm{Re}\,\langle E, S\rangle E\big), \qquad E(0) \in \mathcal{B}\ .$$

Let $c(t) = y(t)^* x(t)$, $y(t), x(t)$ being the normalized left/right eigenvectors associated to an eigenvalue $\lambda(t)$ of $A + \varepsilon E(t)$.

Steepest descent ode

We consider the ODE

$$\dot{E} = -\big(S - \text{Re}\,\langle E, S\rangle E\big), \qquad E(0) \in \mathcal{B}\ .$$

Let $c(t) = y(t)^* x(t)$, $y(t), x(t)$ being the normalized left/right eigenvectors associated to an eigenvalue $\lambda(t)$ of $A + \varepsilon E(t)$.

Properties of ODE

(1) Norm conservation: $\|E(t)\| = 1$ for all t;

(2) Monotonicity: $c(t)$ decreasing along solutions of ODE;

(3) Stationary points: the matrix S does never vanish and the following statements are equivalent:

$$\dot{c} = 0 \quad \Longleftrightarrow \quad \dot{E} = 0 \quad \Longleftrightarrow \quad E \text{ is real multiple of } S.$$

Steepest descent ode

We consider the ODE

$$\dot{E} = -\big(S - \text{Re}\,\langle E, S\rangle E\big), \qquad E(0) \in \mathcal{B}\,.$$

Let $c(t) = y(t)^* x(t)$, $y(t), x(t)$ being the normalized left/right eigenvectors associated to an eigenvalue $\lambda(t)$ of $A + \varepsilon E(t)$.

Properties of ODE

(1) Norm conservation: $\|E(t)\| = 1$ for all t;

(2) Monotonicity: $c(t)$ decreasing along solutions of ODE;

(3) Stationary points: the matrix S does never vanish and the following statements are equivalent:

$$\dot{c} = 0 \quad \Longleftrightarrow \quad \dot{E} = 0 \quad \Longleftrightarrow \quad E \text{ is real multiple of } S.$$

The associated λ represents an extremal ε-pseudo-eigenvalue.

Projection onto the tangent space of $\mathcal{M}_2$

Key property: stationary points have rank-2.

Consider a new ODE on the manifold $\mathcal{M}_2$ of rank-2 matrices by F-orthogonal projection $\mathbf{P}_E$ to tangent space $T_E\mathcal{M}_2$:

$$\dot{E} = -\mathbf{P}_E\Big(S - \mathrm{Re}\,\langle E, S\rangle E\Big).$$

Projection onto the tangent space of $\mathcal{M}_2$

Key property: stationary points have rank-2.

Consider a new ODE on the manifold $\mathcal{M}_2$ of rank-2 matrices by F-orthogonal projection $\mathbf{P}_E$ to tangent space $T_E\mathcal{M}_2$:

$$\dot{E} = -\mathbf{P}_E\Big(S - \mathrm{Re}\,\langle E, S\rangle E\Big).$$

Properties of projected ODE:

(1) Monotonicity: $\dot{c} \leq 0$;

(2) Stationary points: same as unprojected ODE.

Projection onto the tangent space of $\mathcal{M}_2$

Key property: stationary points have rank-2.

Consider a new ODE on the manifold $\mathcal{M}_2$ of rank-2 matrices by F-orthogonal projection $\mathbf{P}_E$ to tangent space $T_E\mathcal{M}_2$:

$$\dot{E} = -\mathbf{P}_E\Big(S - \mathrm{Re}\,\langle E, S\rangle E\Big).$$

Properties of projected ODE:

(1) Monotonicity: $\dot{c} \leq 0$;

(2) Stationary points: same as unprojected ODE.

Writing

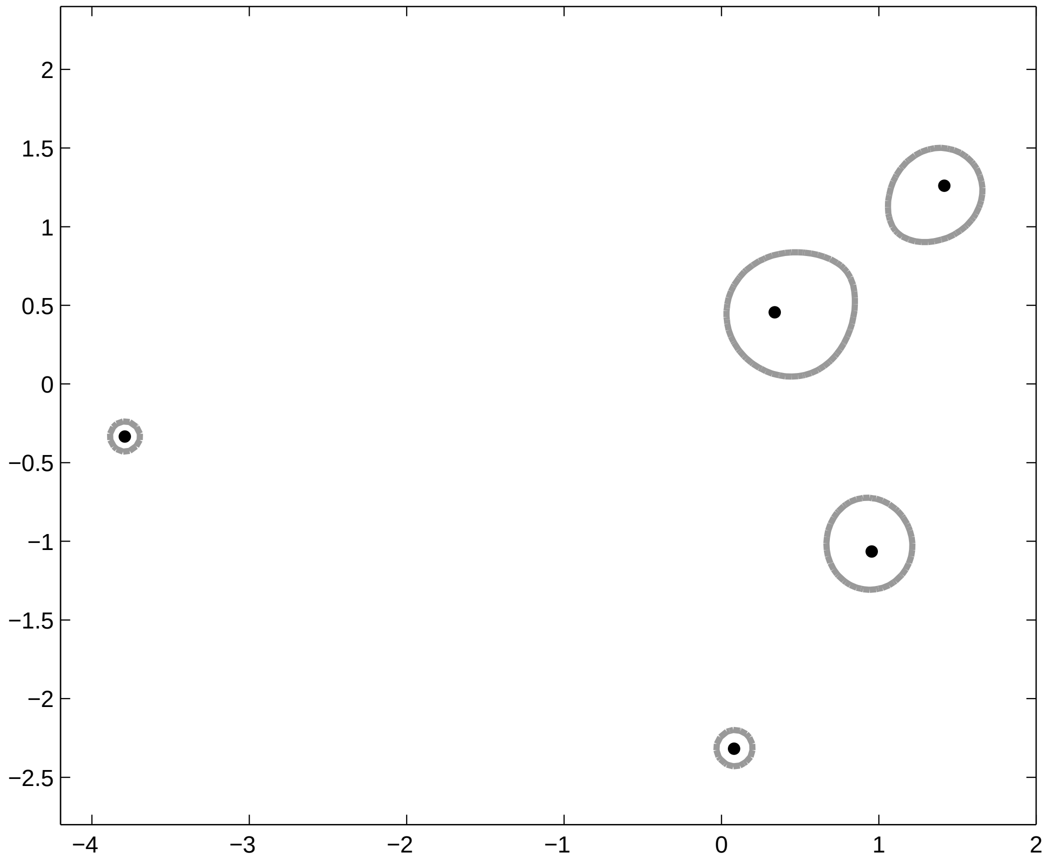
$$E = UTV^*$$

where $U, V \in \mathbb{C}^{n\times 2}$ have orthonormal columns and $T \in \mathbb{C}^{2\times 2}$ invertible, we are able to write a system of ODEs for U, V, T.

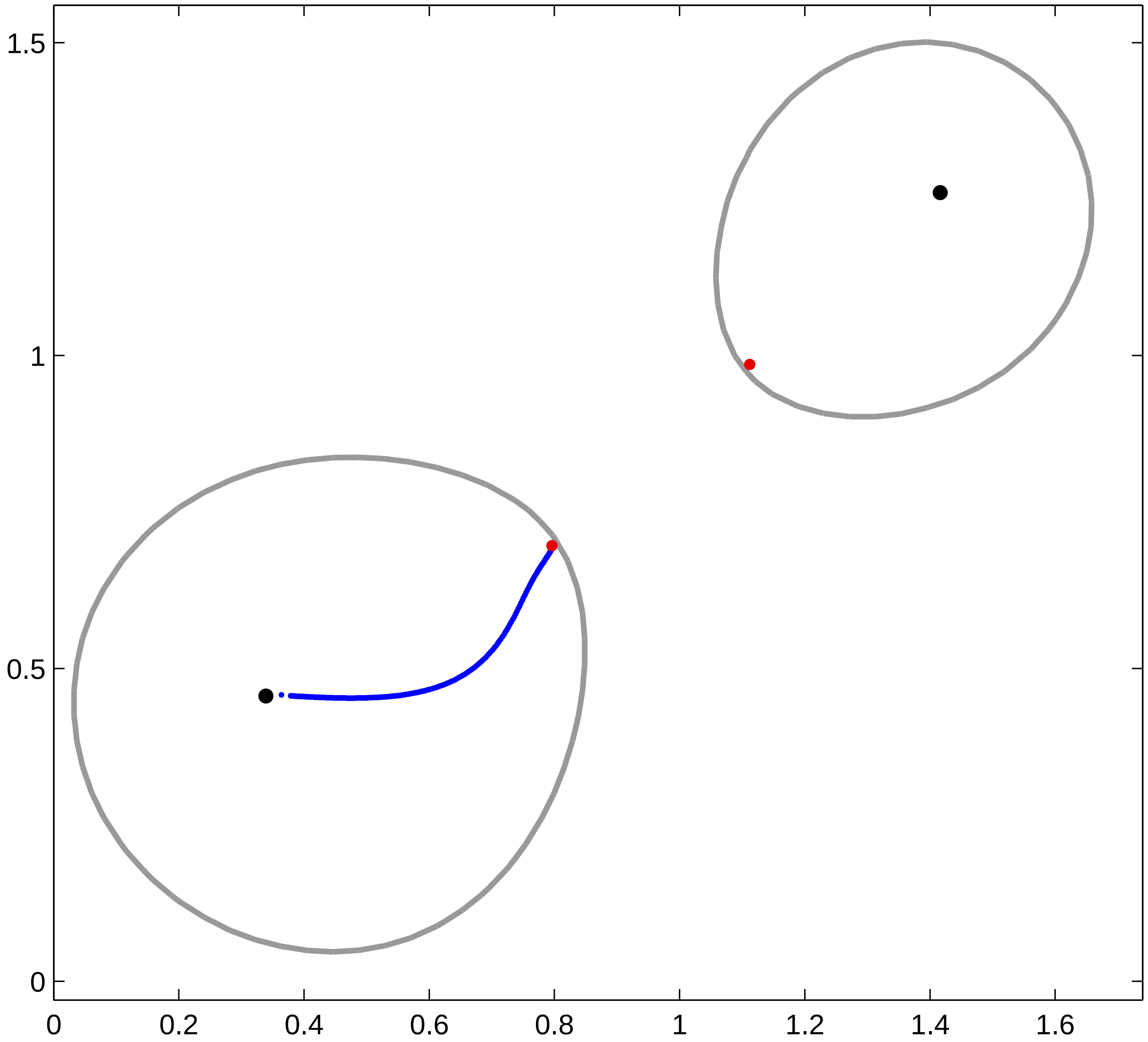
Example 1

$$A = \begin{pmatrix} 0 & 1+\mathbf{i} & 2+\mathbf{i} & 1+2\mathbf{i} & 1 \\ -1 & -1-\mathbf{i} & 1-\mathbf{i} & -\mathbf{i} & 0 \\ 1-\mathbf{i} & -1-2\mathbf{i} & 1+2\mathbf{i} & -2\mathbf{i} & 0 \\ 1-2\mathbf{i} & 1-\mathbf{i} & -1+2\mathbf{i} & -1-\mathbf{i} & 0 \\ 1 & -1-\mathbf{i} & 2\mathbf{i} & -1-\mathbf{i} & -2\mathbf{i} \end{pmatrix}$$

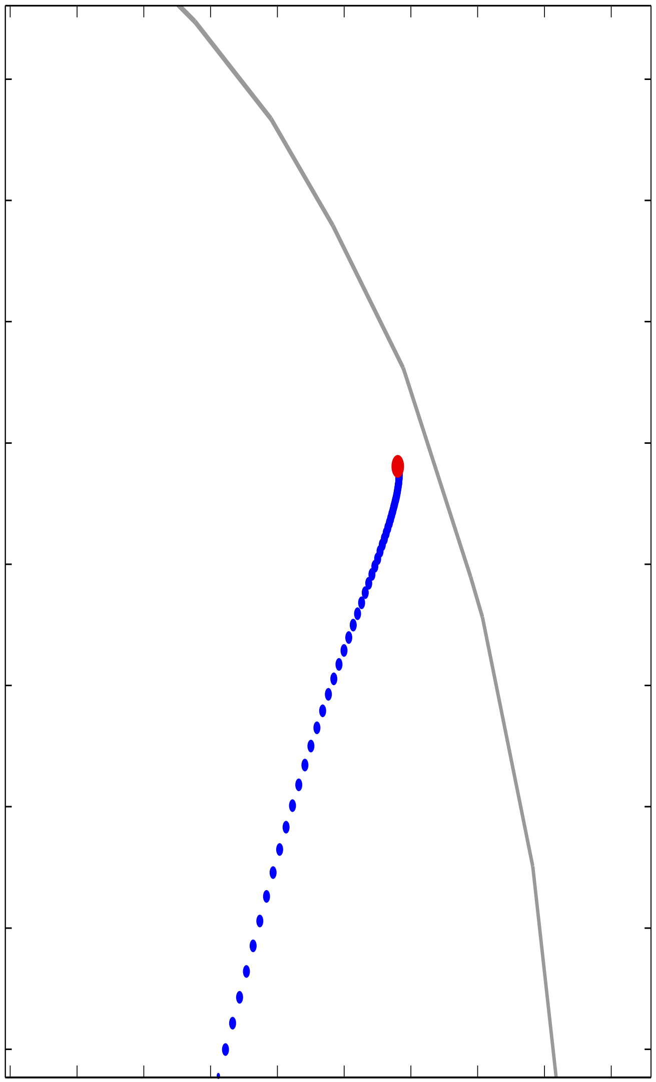
Pseudospectrum
($\varepsilon = 0.075$)



Trajectory of the ODE



Trajectory
in the ε-pseudospectrum



Zoom close
to boundary

Approximating the distance to defectivity

Part (ii). Let $\delta \geq 0$. In order to find an approximate solution of the minimization problem (slight generalization of $\delta = 0$)

$$\varepsilon^{\delta,*} \longrightarrow \min\{\varepsilon : \; r(\varepsilon) = \delta\}$$

we look for locally minimal solutions ε^{δ} of equation $r(\varepsilon) = \delta$.

Approximating the distance to defectivity

Part (ii). Let $\delta \geq 0$. In order to find an approximate solution of the minimization problem (slight generalization of $\delta = 0$)

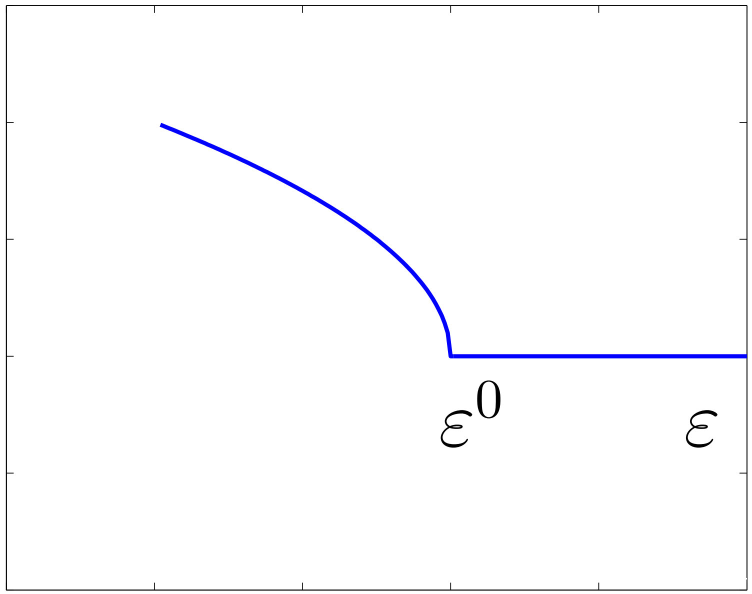
$$\varepsilon^{\delta,*} \longrightarrow \min\{\varepsilon : \ r(\varepsilon) = \delta\}$$

we look for locally minimal solutions ε^{δ} of equation $r(\varepsilon) = \delta$.

Modeling $r(\varepsilon)$

Under generic assumptions we get the expansion for $\varepsilon \leq \varepsilon^0$,

$$\begin{aligned} r(\varepsilon) &= \gamma\sqrt{\varepsilon^0 - \varepsilon} \\ &+ \mathcal{O}\big((\varepsilon^0 - \varepsilon)^{3/2}\big). \end{aligned}$$



Approximating the distance to defectivity

First order expansion

$$r(\varepsilon) = \gamma\sqrt{\varepsilon^0 - \varepsilon} + \ldots, \qquad \text{aim to solve } r(\varepsilon) = \delta$$

Approximating the distance to defectivity

First order expansion

$$r(\varepsilon) = \gamma\sqrt{\varepsilon^0 - \varepsilon} + \ldots, \qquad \text{aim to solve } r(\varepsilon) = \delta$$

Algorithm

Compute $r(\varepsilon)$ by solving the ODE and $dr(\varepsilon)/d\varepsilon$ by an exact inexpensive formula. Estimate γ and ε^0 and solve $r(\varepsilon) = \delta$. This yields a quadratically convergent method to ε^δ.

Approximating the distance to defectivity

First order expansion

$$r(\varepsilon) = \gamma\sqrt{\varepsilon^0 - \varepsilon} + \ldots, \qquad \text{aim to solve } r(\varepsilon) = \delta$$

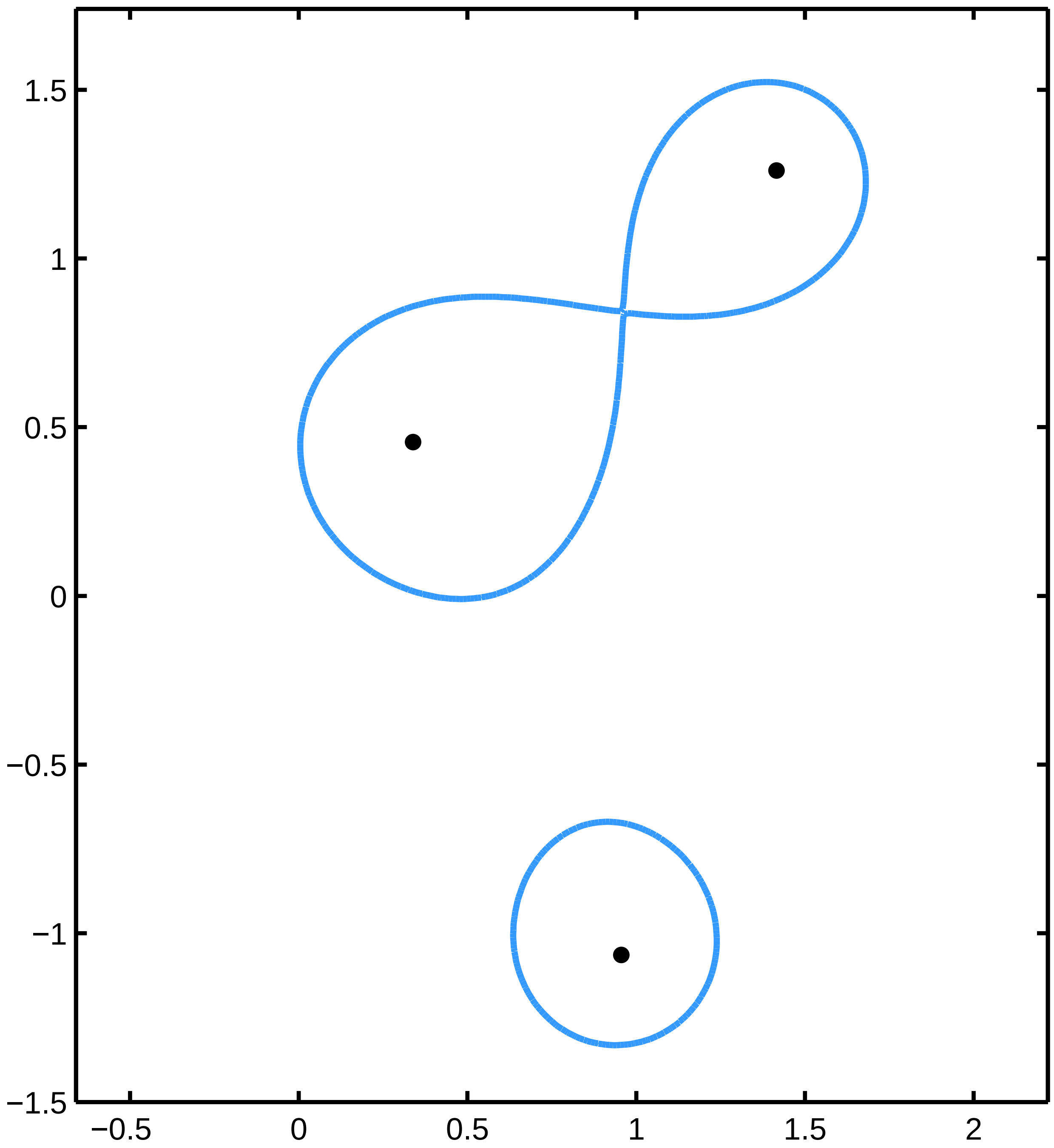
Algorithm

Compute $r(\varepsilon)$ by solving the ODE and $dr(\varepsilon)/d\varepsilon$ by an exact inexpensive formula. Estimate γ and ε^0 and solve $r(\varepsilon) = \delta$. This yields a quadratically convergent method to ε^δ.

Example 1 ($\delta = 10^{-4}$)

k	ε_k^δ	$r(\varepsilon_k^\delta)$
7	0.082876946962636	0.000**9**10106101987
8	0.082876706789675	0.000**9999**89689847
9	0.082876706760826	0.000**9999999999**761

Example 1: ε^0-pseudospectrum



Real-structured distance

Step (ii) is unaltered. Step (i): the modified ODE

It is sufficient to replace S by $\mathrm{Re}\,(S)$ in the complex ODE and observe that stationary points are now real rank-4 matrices. We also prove that $\mathrm{Re}\,(S)$ does never vanish if A is not normal.

Real-structured distance

Step (ii) is unaltered. Step (i): the modified ODE

It is sufficient to replace S by Re (S) in the complex ODE and observe that stationary points are now real rank-4 matrices. We also prove that Re (S) does never vanish if A is not normal.

Projected ODE

By F-orthogonal projection $\widetilde{\mathbf{P}}_E$ to tangent space $T_E\mathcal{M}_4$ of the manifold of real 4×4-matrices, we get

$$\dot{E} = -\widetilde{\mathbf{P}}_E\Big(\mathrm{Re}\,(S) - \mathrm{Re}\,\langle E, \mathrm{Re}\,(S)\rangle E\Big).$$

Real-structured distance

Step (ii) is unaltered. Step (i): the modified ODE

It is sufficient to replace S by Re (S) in the complex ODE and observe that stationary points are now real rank-4 matrices. We also prove that Re (S) does never vanish if A is not normal.

Projected ODE

By F-orthogonal projection $\widetilde{\mathbf{P}}_E$ to tangent space $T_E\mathcal{M}_4$ of the manifold of real 4×4-matrices, we get

$$\dot{E} = -\widetilde{\mathbf{P}}_E\Big(\mathrm{Re}\,(S) - \mathrm{Re}\,\langle E, \mathrm{Re}\,(S)\rangle E\Big).$$

Properties

(1) Monotonicity: $\dot{c} \leq 0$;

(2) Stationary points: same as unprojected ODE: $E \propto \mathrm{Re}\,(S)$.

Sparsity pattern ($\mathcal{P}$) structure

The sparsity preserving ODE

Is sufficient an F-orthogonal projecton of S onto $\mathcal{P}$ i.e. setting to zero all elements of S corresponding to zero elements of $\mathcal{P}$.

Sparsity pattern ($\mathcal{P}$) structure

The sparsity preserving ODE

Is sufficient an F-orthogonal projecton of S onto $\mathcal{P}$ i.e. setting to zero all elements of S corresponding to zero elements of $\mathcal{P}$.

Example 2 (Grcar matrix)

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ -1 & 1 & 1 & 1 & 1 & 0 \\ 0 & -1 & 1 & 1 & 1 & 1 \\ 0 & 0 & -1 & 1 & 1 & 1 \\ 0 & 0 & 0 & -1 & 1 & 1 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

Distances

$$w_{\mathbb{C}}(A) \approx 0.2151857$$

$$w_{\mathbb{R}}(A) \approx 0.3007253$$

$$w_{\mathbb{C},\mathcal{P}}(A) \approx 0.6845324$$

$$w_{\mathbb{R},\mathcal{P}}(A) \approx 0.9423366$$

Sparsity pattern ($\mathcal{P}$) structure

The sparsity preserving ODE
Is sufficient an F-orthogonal projecton of S onto $\mathcal{P}$ i.e. setting to zero all elements of S corresponding to zero elements of $\mathcal{P}$.

Example 2 (Grcar matrix)

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ -1 & 1 & 1 & 1 & 1 & 0 \\ 0 & -1 & 1 & 1 & 1 & 1 \\ 0 & 0 & -1 & 1 & 1 & 1 \\ 0 & 0 & 0 & -1 & 1 & 1 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

Distances

$$w_{\mathbb{C}}(A) \approx 0.2151857$$

$$w_{\mathbb{R}}(A) \approx 0.3007253$$

$$w_{\mathbb{C},\mathcal{P}}(A) \approx 0.6845324$$

$$w_{\mathbb{R},\mathcal{P}}(A) \approx 0.9423366$$

Large sparse problems may exploit the low rank-structure and computing efficiently the group-inverse (project with Michael).