Numerical Optimization of Eigenvalues of Hermitian Matrix Functions

Emre Mengi

Department of Mathematics
Koç University
İstanbul, Turkey

MOPNET 6, University of Bath
April 3rd, 2012

joint with Daniel Kressner (EPF-Lausanne), Ivica Nakić (Univ of Zagreb), Ninoslav Truhar (Univ of Osijek), Mustafa Kılıç (Koç) and E. Alper Yıldırım (Koç)

Supported in part by Marie Curie IRG Grant EC268355 and TÜBİTAK Career Grant 109T660

Outline

Wilkinson Distance

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \hookrightarrow \begin{bmatrix} 0 & 1 \\ \epsilon & 0 \end{bmatrix}$$

$$\lambda_{\pm}(\epsilon) = \pm\sqrt{\epsilon}$$

$$\frac{\lambda_{\pm}(\epsilon) - \lambda(0)}{\epsilon} = \pm\frac{1}{\sqrt{\epsilon}}$$

Eigenvalues associated with a Jordan block are very sensitive to perturbations.

Wilkinson Distance

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \hookrightarrow \begin{bmatrix} 0 & 1 \\ \epsilon & 0 \end{bmatrix}$$

$$\lambda_{\pm}(\epsilon) = \pm\sqrt{\epsilon}$$

$$\frac{\lambda_{\pm}(\epsilon) - \lambda(0)}{\epsilon} = \pm\frac{1}{\sqrt{\epsilon}}$$

Eigenvalues associated with a Jordan block are very sensitive to perturbations.

Wilkinson Distance

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \hookrightarrow \begin{bmatrix} 0 & 1 \\ \epsilon & 0 \end{bmatrix}$$

$$\lambda_{\pm}(\epsilon) = \pm\sqrt{\epsilon}$$

$$\frac{\lambda_{\pm}(\epsilon)-\lambda(0)}{\epsilon} = \pm\frac{1}{\sqrt{\epsilon}}$$

Eigenvalues associated with a Jordan block are very sensitive to perturbations.

Wilkinson Distance

$$\begin{bmatrix} 0 & 1 \\ 0 & \beta \end{bmatrix} \hookrightarrow \begin{bmatrix} 0 & 1 \\ \epsilon & \beta \end{bmatrix}$$

$$\lambda_{\pm}(\epsilon) = \frac{\beta \pm \sqrt{\beta^2 + 4\epsilon}}{2}$$

$$\frac{\lambda_-(\epsilon) - \lambda_-(0)}{\epsilon} = \frac{1}{\beta} + O(\epsilon)$$

Matrices close to defective matrices have also very sensitive eigenvalues.

Wilkinson Distance

$$\begin{bmatrix} 0 & 1 \\ 0 & \beta \end{bmatrix} \hookrightarrow \begin{bmatrix} 0 & 1 \\ \epsilon & \beta \end{bmatrix}$$

$$\lambda_{\pm}(\epsilon) = \frac{\beta \pm \sqrt{\beta^2 + 4\epsilon}}{2}$$

$$\frac{\lambda_{-}(\epsilon) - \lambda_{-}(0)}{\epsilon} = \frac{1}{\beta} + O(\epsilon)$$

Matrices close to defective matrices have also very sensitive eigenvalues.

Wilkinson Distance

$$\begin{bmatrix} 0 & 1 \\ 0 & \beta \end{bmatrix} \hookrightarrow \begin{bmatrix} 0 & 1 \\ \epsilon & \beta \end{bmatrix}$$

$$\lambda_{\pm}(\epsilon) = \frac{\beta \pm \sqrt{\beta^2 + 4\epsilon}}{2}$$

$$\frac{\lambda_{-}(\epsilon) - \lambda_{-}(0)}{\epsilon} = \frac{1}{\beta} + O(\epsilon)$$

Matrices close to defective matrices have also very sensitive eigenvalues.

Wilkinson Distance

Absolute condition number of an eigenvalue λ

$$\kappa(\lambda) = \lim_{\delta \to 0^+} \sup_{\|\delta A\| \leq \delta} \frac{|\lambda(\delta A) - \lambda|}{\|\delta A\|} = \frac{1}{y^* x}$$

where

$y, x \in \mathbb{C}^n$: unit left and right eigenvectors associated with λ.

- Jordan blocks: $y^* x = 0$
- Best conditioned - Normal matrices: $y = x \iff y^* x = 1$

Wilkinson Distance

Absolute condition number of an eigenvalue λ

$$\kappa(\lambda) = \lim_{\delta \to 0^+} \sup_{\|\delta A\| \leq \delta} \frac{|\lambda(\delta A) - \lambda|}{\|\delta A\|} = \frac{1}{y^* x}$$

where

$y, x \in \mathbb{C}^n$: unit left and right eigenvectors associated with λ.

- Jordan blocks: $y^* x = 0$
- Best conditioned - Normal matrices: $y = x \iff y^* x = 1$

Wilkinson Distance

Absolute condition number of an eigenvalue λ

$$\kappa(\lambda) = \lim_{\delta \to 0^+} \sup_{\|\delta A\| \leq \delta} \frac{|\lambda(\delta A) - \lambda|}{\|\delta A\|} = \frac{1}{y^* x}$$

where

$y, x \in \mathbb{C}^n$: unit left and right eigenvectors associated with λ.

- Jordan blocks: $y^* x = 0$
- Best conditioned - Normal matrices: $y = x \iff y^* x = 1$

Wilkinson Distance

Absolute condition number of an eigenvalue λ

$$\kappa(\lambda) = \lim_{\delta \to 0^+} \sup_{\|\delta A\| \leq \delta} \frac{|\lambda(\delta A) - \lambda|}{\|\delta A\|} = \frac{1}{y^* x}$$

where

$y, x \in \mathbb{C}^n$: unit left and right eigenvectors associated with λ.

- Jordan blocks: $y^* x = 0$
- Best conditioned - Normal matrices: $y = x \iff y^* x = 1$

Wilkinson Distance

J.H. Wilkinson, The Algebraic Eigenvalue Problem

The eigenvalues corresponding to non-linear elementary divisors must, in general, be regarded as ill-conditioned ... However, we must not be misled into thinking that this is the main form of ill-conditioning. Even if the eigenvalues are distinct and well separated they may still be very ill-conditioned.

Wilkinson Distance

$$W(\epsilon) = \begin{bmatrix} 20 & 20 & & & & \\ & 19 & 20 & & & \\ & & 18 & 20 & & \\ & & & \ddots & \ddots & \\ & & & & 2 & 20 \\ \epsilon & & & & & 1 \end{bmatrix}$$

- Condition numbers for $W(0)$,

  $$\kappa(\lambda) = \frac{20^{19}}{10!\ 9!} \quad \text{for } \lambda = 10, 11$$

- $W(\epsilon)$ has $\lambda = 10.5$ as a multiple eigenvalue for $\epsilon \approx 8 \times 10^{-14}$.

Wilkinson Distance

$$W(\epsilon) = \begin{bmatrix} 20 & 20 & & & & \\ & 19 & 20 & & & \\ & & 18 & 20 & & \\ & & & \ddots & \ddots & \\ & & & & 2 & 20 \\ \epsilon & & & & & 1 \end{bmatrix}$$

- Condition numbers for $W(0)$,

$$\kappa(\lambda) = \frac{20^{19}}{10!\ 9!} \quad \text{for } \lambda = 10, 11$$

- $W(\epsilon)$ has $\lambda = 10.5$ as a multiple eigenvalue for $\epsilon \approx 8 \times 10^{-14}$.

Wilkinson Distance

$$W(\epsilon) = \begin{bmatrix} 20 & 20 & & & & \\ & 19 & 20 & & & \\ & & 18 & 20 & & \\ & & & \ddots & \ddots & \\ & & & & 2 & 20 \\ \epsilon & & & & & 1 \end{bmatrix}$$

- Condition numbers for $W(0)$,

$$\kappa(\lambda) = \frac{20^{19}}{10!\,9!} \quad \text{for } \lambda = 10, 11$$

- $W(\epsilon)$ has $\lambda = 10.5$ as a multiple eigenvalue for $\epsilon \approx 8 \times 10^{-14}$.

Wilkinson Distance

Definition (Wilkinson Distance)

The distance in 2-norm from $A \in \mathbb{C}^{n \times n}$ to the nearest matrix with a multiple eigenvalue

$$\mathcal{W}(A) = \inf\{\|\delta A\|_2 : \exists \lambda \ \ (A + \delta A) \text{ has } \lambda \text{ as a multiple eigenvalue}\}$$

is called the *Wilkinson distance* of A.

Wilkinson's bound

$$\mathcal{W}(A) \leq \|A\|_2 / \sqrt{\kappa(\lambda)^2 - 1}$$

Wilkinson Distance

Definition (Wilkinson Distance)

The distance in 2-norm from $A \in \mathbb{C}^{n \times n}$ to the nearest matrix with a multiple eigenvalue

$$\mathcal{W}(A) = \inf\{\|\delta A\|_2 : \exists \lambda \;\; (A + \delta A) \text{ has } \lambda \text{ as a multiple eigenvalue}\}$$

is called the *Wilkinson distance* of A.

Wilkinson's bound

$$\mathcal{W}(A) \leq \|A\|_2 / \sqrt{\kappa(\lambda)^2 - 1}$$

Wilkinson Distance

Singular Value Characterization

Define also

$$\mathcal{W}(A, \lambda) := \inf\{\|\delta A\|_2 : (A + \delta A) \text{ has } \lambda \text{ as a multiple eigenvalue}\}.$$

Theorem (Malyshev, 1999)

(i) *Then for all* $\lambda \in \mathbb{C}$

$$\mathcal{W}(A, \lambda) = \sup_{\gamma \in \mathbb{R}^+} \sigma_{2n-1}\left(\begin{bmatrix} A - \lambda I & \gamma I \\ 0 & A - \lambda I \end{bmatrix}\right).$$

(ii) *Consequently*

$$\mathcal{W}(A) = \inf_{\lambda \in \mathbb{C}} \sup_{\gamma \in \mathbb{R}^+} \sigma_{2n-1}\left(\begin{bmatrix} A - \lambda I & \gamma I \\ 0 & A - \lambda I \end{bmatrix}\right).$$

Wilkinson Distance
Singular Value Characterization

Define also

$$\mathcal{W}(A, \lambda) := \inf\{\|\delta A\|_2 : (A + \delta A) \text{ has } \lambda \text{ as a multiple eigenvalue}\}.$$

Theorem (Malyshev, 1999)

(i) *Then for all* $\lambda \in \mathbb{C}$

$$\mathcal{W}(A, \lambda) = \sup_{\gamma \in \mathbb{R}^+} \sigma_{2n-1}\left(\begin{bmatrix} A - \lambda I & \gamma I \\ 0 & A - \lambda I \end{bmatrix}\right).$$

(ii) *Consequently*

$$\mathcal{W}(A) = \inf_{\lambda \in \mathbb{C}} \sup_{\gamma \in \mathbb{R}^+} \sigma_{2n-1}\left(\begin{bmatrix} A - \lambda I & \gamma I \\ 0 & A - \lambda I \end{bmatrix}\right).$$

Wilkinson Distance
Singular Value Characterization

Define also

$$\mathcal{W}(A, \lambda) := \inf\{\|\delta A\|_2 : (A + \delta A) \text{ has } \lambda \text{ as a multiple eigenvalue}\}.$$

Theorem (Malyshev, 1999)

(i) *Then for all* $\lambda \in \mathbb{C}$

$$\mathcal{W}(A, \lambda) = \sup_{\gamma \in \mathbb{R}^+} \sigma_{2n-1}\left(\begin{bmatrix} A - \lambda I & \gamma I \\ 0 & A - \lambda I \end{bmatrix}\right).$$

(ii) *Consequently*

$$\mathcal{W}(A) = \inf_{\lambda \in \mathbb{C}} \sup_{\gamma \in \mathbb{R}^+} \sigma_{2n-1}\left(\begin{bmatrix} A - \lambda I & \gamma I \\ 0 & A - \lambda I \end{bmatrix}\right).$$

Wilkinson Distance
Singular Value Characterization

For a matrix pencil $L(\mu) := A + \mu B$ with $A, B \in \mathbb{C}^{n \times n}$

$$\mathcal{W}(L, \lambda) := \inf\{\|\delta A\|_2 : (A+\delta A)+\mu B \text{ has } \lambda \text{ as a multiple eigenvalue}\}.$$

Theorem (Kressner, M, Nakic, Truhar 2011)

(i) *For all* $\lambda \in \mathbb{C}$

$$\mathcal{W}(L, \lambda) = \sup_{\gamma \in \mathbb{R}^+} \sigma_{2n-1}\left(\begin{bmatrix} A + \lambda B & \gamma B \\ 0 & A + \lambda B \end{bmatrix}\right).$$

(ii) *Consequently*

$$\mathcal{W}(L) = \inf_{\lambda \in \mathbb{C}} \sup_{\gamma \in \mathbb{R}^+} \sigma_{2n-1}\left(\begin{bmatrix} A + \lambda B & \gamma B \\ 0 & A + \lambda B \end{bmatrix}\right).$$

Wilkinson Distance
Singular Value Characterization

For a matrix pencil $L(\mu) := A + \mu B$ with $A, B \in \mathbb{C}^{n\times n}$

$\mathcal{W}(L,\lambda) := \inf\{\|\delta A\|_2 : (A+\delta A)+\mu B \text{ has } \lambda \text{ as a multiple eigenvalue}\}.$

Theorem (Kressner, M, Nakic, Truhar 2011)

(i) *For all* $\lambda \in \mathbb{C}$

$$\mathcal{W}(L,\lambda) = \sup_{\gamma\in\mathbb{R}^+} \sigma_{2n-1}\left(\begin{bmatrix} A+\lambda B & \gamma B \\ 0 & A+\lambda B \end{bmatrix}\right).$$

(ii) *Consequently*

$$\mathcal{W}(L) = \inf_{\lambda\in\mathbb{C}} \sup_{\gamma\in\mathbb{R}^+} \sigma_{2n-1}\left(\begin{bmatrix} A+\lambda B & \gamma B \\ 0 & A+\lambda B \end{bmatrix}\right).$$

Wilkinson Distance
Singular Value Characterization

For a matrix pencil $L(\mu) := A + \mu B$ with $A, B \in \mathbb{C}^{n \times n}$

$\mathcal{W}(L, \lambda) := \inf\{\|\delta A\|_2 : (A+\delta A)+\mu B$ has λ as a multiple eigenvalue$\}$.

Theorem (Kressner, M, Nakic, Truhar 2011)

(i) *For all* $\lambda \in \mathbb{C}$

$$\mathcal{W}(L, \lambda) = \sup_{\gamma \in \mathbb{R}^+} \sigma_{2n-1}\left(\begin{bmatrix} A + \lambda B & \gamma B \\ 0 & A + \lambda B \end{bmatrix}\right).$$

(ii) *Consequently*

$$\mathcal{W}(L) = \inf_{\lambda \in \mathbb{C}} \sup_{\gamma \in \mathbb{R}^+} \sigma_{2n-1}\left(\begin{bmatrix} A + \lambda B & \gamma B \\ 0 & A + \lambda B \end{bmatrix}\right).$$

Wilkinson Distance

Singular Value Characterization

For a matrix polynomial $P(\mu) := \sum_{j=0}^{k} \mu^j A_j$ with $A_j \in \mathbb{C}^{n \times n}$

$$\mathcal{W}(P, \lambda) := \inf\{\|\delta A\|_2 : \sum_{j=0}^{k} \mu^j A_j + \delta A \text{ has } \lambda \text{ as a multiple eigenvalue}\}.$$

Theorem (Karow, M, Pelen 2012)

(i) *For all* $\lambda \in \mathbb{C}$

$$\mathcal{W}(P, \lambda) = \sup_{\gamma \in \mathbb{R}^+} \sigma_{2n-1}\left(\begin{bmatrix} P(\lambda) & \gamma P'(\lambda) \\ 0 & P(\lambda) \end{bmatrix}\right).$$

(ii) *Consequently*

$$\mathcal{W}(P) = \inf_{\lambda \in \mathbb{C}} \sup_{\gamma \in \mathbb{R}^+} \sigma_{2n-1}\left(\begin{bmatrix} P(\lambda) & \gamma P'(\lambda) \\ 0 & P(\lambda) \end{bmatrix}\right).$$

Wilkinson Distance
Singular Value Characterization

For a matrix polynomial $P(\mu) := \sum_{j=0}^{k} \mu^j A_j$ with $A_j \in \mathbb{C}^{n\times n}$

$$\mathcal{W}(P, \lambda) := \inf\{\|\delta A\|_2 : \sum_{j=0}^{k} \mu^j A_j + \delta A \text{ has } \lambda \text{ as a multiple eigenvalue}\}.$$

Theorem (Karow, M, Pelen 2012)

(i) *For all* $\lambda \in \mathbb{C}$

$$\mathcal{W}(P, \lambda) = \sup_{\gamma \in \mathbb{R}^+} \sigma_{2n-1}\left(\begin{bmatrix} P(\lambda) & \gamma P'(\lambda) \\ 0 & P(\lambda) \end{bmatrix}\right).$$

(ii) *Consequently*

$$\mathcal{W}(P) = \inf_{\lambda \in \mathbb{C}} \sup_{\gamma \in \mathbb{R}^+} \sigma_{2n-1}\left(\begin{bmatrix} P(\lambda) & \gamma P'(\lambda) \\ 0 & P(\lambda) \end{bmatrix}\right).$$

Wilkinson Distance
Singular Value Characterization

For a matrix polynomial $P(\mu) := \sum_{j=0}^{k} \mu^j A_j$ with $A_j \in \mathbb{C}^{n \times n}$

$$\mathcal{W}(P, \lambda) := \inf\{\|\delta A\|_2 : \sum_{j=0}^{k} \mu^j A_j + \delta A \text{ has } \lambda \text{ as a multiple eigenvalue}\}.$$

Theorem (Karow, M, Pelen 2012)

(i) *For all* $\lambda \in \mathbb{C}$

$$\mathcal{W}(P, \lambda) = \sup_{\gamma \in \mathbb{R}^+} \sigma_{2n-1}\left(\begin{bmatrix} P(\lambda) & \gamma P'(\lambda) \\ 0 & P(\lambda) \end{bmatrix}\right).$$

(ii) *Consequently*

$$\mathcal{W}(P) = \inf_{\lambda \in \mathbb{C}} \sup_{\gamma \in \mathbb{R}^+} \sigma_{2n-1}\left(\begin{bmatrix} P(\lambda) & \gamma P'(\lambda) \\ 0 & P(\lambda) \end{bmatrix}\right).$$

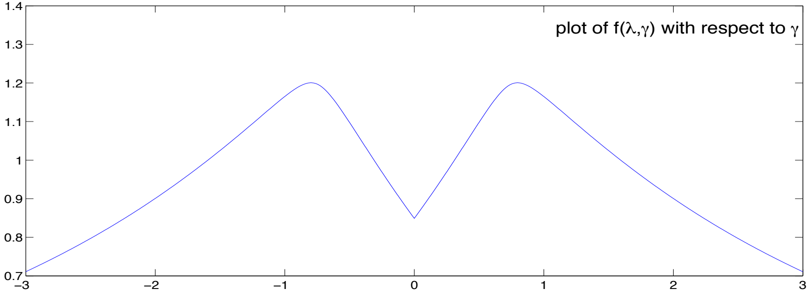
Wilkinson Distance
Numerical Computation

Inner maximization (Secant or Quasi-Newton)

Any stationary point $\gamma_* \in \mathbb{R}^+$ of the inner function

$$f(\lambda,\gamma) = \sigma_{2n-1}\left(\begin{bmatrix} A-\lambda I & \gamma I \\ 0 & A-\lambda I \end{bmatrix}\right)$$

is a global maximizer.



Wilkinson Distance
Numerical Computation

Outer minimization (Lipschitzness and Analyticity)

$f(\lambda, \gamma)$ is Lipschitz continuous w.r.t. λ and γ (from the Weyl's theorem).

Theorem (Weyl)

Let A and E be Hermitian $n \times n$ matrices and $\lambda_j(\cdot)$ denote the jth largest eigenvalue of its matrix argument. Then $|\lambda_j(A) - \lambda_j(A+E)| \leq \|E\|_2$.

- $\mathcal{W}(A, \lambda) = \sup_\gamma f(\lambda, \gamma)$ is Lipschitz continuous w.r.t. λ.
  Lipschitzness solely yields slow-converging algorithms.
- $\mathcal{W}(A, \lambda) = \sup_\gamma f(\lambda, \gamma)$ is also piece-wise analytic w.r.t. λ.
  Analyticity may yield faster algorithms.

Wilkinson Distance
Numerical Computation

Outer minimization (Lipschitzness and Analyticity)

$f(\lambda, \gamma)$ is Lipschitz continuous w.r.t. λ and γ (from the Weyl's theorem).

Theorem (Weyl)

Let A and E be Hermitian $n \times n$ matrices and $\lambda_j(\cdot)$ denote the jth largest eigenvalue of its matrix argument. Then $|\lambda_j(A) - \lambda_j(A + E)| \leq \|E\|_2$.

- $\mathcal{W}(A, \lambda) = \sup_\gamma f(\lambda, \gamma)$ is Lipschitz continuous w.r.t. λ.
  Lipschitzness solely yields slow-converging algorithms.
- $\mathcal{W}(A, \lambda) = \sup_\gamma f(\lambda, \gamma)$ is also piece-wise analytic w.r.t. λ.
  Analyticity may yield faster algorithms.

Wilkinson Distance
Numerical Computation

Outer minimization (Lipschitzness and Analyticity)

$f(\lambda, \gamma)$ is Lipschitz continuous w.r.t. λ and γ (from the Weyl's theorem).

Theorem (Weyl)

Let A and E be Hermitian $n \times n$ matrices and $\lambda_j(\cdot)$ denote the jth largest eigenvalue of its matrix argument. Then $|\lambda_j(A) - \lambda_j(A+E)| \leq \|E\|_2$.

- $\mathcal{W}(A, \lambda) = \sup_\gamma f(\lambda, \gamma)$ is Lipschitz continuous w.r.t. λ.
  Lipschitzness solely yields slow-converging algorithms.
- $\mathcal{W}(A, \lambda) = \sup_\gamma f(\lambda, \gamma)$ is also piece-wise analytic w.r.t. λ.
  Analyticity may yield faster algorithms.

Wilkinson Distance
Numerical Computation

Outer minimization (Lipschitzness and Analyticity)

$f(\lambda, \gamma)$ is Lipschitz continuous w.r.t. λ and γ (from the Weyl's theorem).

Theorem (Weyl)

Let A and E be Hermitian $n \times n$ matrices and $\lambda_j(\cdot)$ denote the jth largest eigenvalue of its matrix argument. Then $|\lambda_j(A) - \lambda_j(A+E)| \leq \|E\|_2$.

- $\mathcal{W}(A, \lambda) = \sup_\gamma f(\lambda, \gamma)$ is Lipschitz continuous w.r.t. λ.
  Lipschitzness solely yields slow-converging algorithms.
- $\mathcal{W}(A, \lambda) = \sup_\gamma f(\lambda, \gamma)$ is also piece-wise analytic w.r.t. λ.
  Analyticity may yield faster algorithms.

Outline

Analyticity Result

Theorem (Rellich)

Let $\mathcal{A}(\omega) : \mathbb{R} \to \mathbb{C}^{n \times n}$ be a Hermitian matrix function that depends on ω analytically.

(i) *The n roots of the characteristic polynomial of $\mathcal{A}(\omega)$ can be arranged so that each root $\tilde{\lambda}_j(\omega)$ for $j = 1, \dots, n$ is an analytic function of ω.*

(ii) *There exists an eigenvector $v_j(\omega)$ associated with $\tilde{\lambda}_j(\omega)$ for $j = 1, \dots, n$ satisfying*

(1) $\left(\tilde{\lambda}_j(\omega) I - \mathcal{A}(\omega)\right) v_j(\omega) = 0 \ \ \forall \omega \in \mathbb{R}$,

(2) $\|v_j(\omega)\|_2 = 1 \ \ \forall \omega \in \mathbb{R}$,

(3) $v_j^*(\omega) v_k(\omega) = 0 \ \ \forall \omega \in \mathbb{R}$ *for* $k \neq j$, *and*

(4) $v_j(\omega)$ *is an analytic function of* ω.

Analyticity Result

Theorem (Rellich)

Let $\mathcal{A}(\omega) : \mathbb{R} \to \mathbb{C}^{n\times n}$ be a Hermitian matrix function that depends on ω analytically.

(i) *The n roots of the characteristic polynomial of $\mathcal{A}(\omega)$ can be arranged so that each root $\tilde{\lambda}_j(\omega)$ for $j = 1, \dots, n$ is an analytic function of ω.*

(ii) *There exists an eigenvector $v_j(\omega)$ associated with $\tilde{\lambda}_j(\omega)$ for $j = 1, \dots, n$ satisfying*

(1) $\left(\tilde{\lambda}_j(\omega) I - \mathcal{A}(\omega)\right) v_j(\omega) = 0 \;\; \forall \omega \in \mathbb{R}$,

(2) $\|v_j(\omega)\|_2 = 1 \;\; \forall \omega \in \mathbb{R}$,

(3) $v_j^*(\omega) v_k(\omega) = 0 \;\; \forall \omega \in \mathbb{R}$ *for $k \neq j$, and*

(4) $v_j(\omega)$ *is an analytic function of ω.*

Analyticity Result

Theorem (Rellich)

Let $\mathcal{A}(\omega) : \mathbb{R} \to \mathbb{C}^{n \times n}$ be a Hermitian matrix function that depends on ω analytically.

(i) *The n roots of the characteristic polynomial of $\mathcal{A}(\omega)$ can be arranged so that each root $\tilde{\lambda}_j(\omega)$ for $j = 1, \ldots, n$ is an analytic function of ω.*

(ii) *There exists an eigenvector $v_j(\omega)$ associated with $\tilde{\lambda}_j(\omega)$ for $j = 1, \ldots, n$ satisfying*

(1) $\left(\tilde{\lambda}_j(\omega) I - \mathcal{A}(\omega)\right) v_j(\omega) = 0 \;\; \forall \omega \in \mathbb{R}$,

(2) $\|v_j(\omega)\|_2 = 1 \;\; \forall \omega \in \mathbb{R}$,

(3) $v_j^*(\omega) v_k(\omega) = 0 \;\; \forall \omega \in \mathbb{R}$ *for* $k \neq j$, *and*

(4) $v_j(\omega)$ *is an analytic function of* ω.

Analyticity Result

The eigenvalues $\lambda_1(\omega), \ldots, \lambda_n(\omega)$ ordered from largest to smallest of $\mathcal{A}(\omega)$ are continuous and piece-wise analytic.

e.g.

Let $\mathcal{A}(\omega) = \begin{bmatrix} 0 & \omega \\ \omega & 0 \end{bmatrix}$ with analytic eigenvalues

$$\tilde{\lambda}_1(\omega) = \omega \quad \text{and} \quad \tilde{\lambda}_2(\omega) = -\omega$$

Sorted continuous and piece-wise analytic eigenvalues

$$\lambda_1(\omega) = |\omega| \quad \text{and} \quad \lambda_2(\omega) = -|\omega|$$

Analyticity Result

The eigenvalues $\lambda_1(\omega), \ldots, \lambda_n(\omega)$ ordered from largest to smallest of $\mathcal{A}(\omega)$ are continuous and piece-wise analytic.

e.g.

Let $\mathcal{A}(\omega) = \begin{bmatrix} 0 & \omega \\ \omega & 0 \end{bmatrix}$ with analytic eigenvalues

$$\tilde{\lambda}_1(\omega) = \omega \ \text{ and } \ \tilde{\lambda}_2(\omega) = -\omega$$

Sorted continuous and piece-wise analytic eigenvalues

$$\lambda_1(\omega) = |\omega| \ \text{ and } \ \lambda_2(\omega) = -|\omega|$$

Analyticity Result

The eigenvalues $\lambda_1(\omega), \ldots, \lambda_n(\omega)$ ordered from largest to smallest of $\mathcal{A}(\omega)$ are continuous and piece-wise analytic.

e.g.

Let $\mathcal{A}(\omega) = \begin{bmatrix} 0 & \omega \\ \omega & 0 \end{bmatrix}$ with analytic eigenvalues

$$\tilde{\lambda}_1(\omega) = \omega \;\; \text{and} \;\; \tilde{\lambda}_2(\omega) = -\omega$$

Sorted continuous and piece-wise analytic eigenvalues

$$\lambda_1(\omega) = |\omega| \;\; \text{and} \;\; \lambda_2(\omega) = -|\omega|$$

Analyticity Result

The eigenvalues $\lambda_1(\omega), \ldots, \lambda_n(\omega)$ ordered from largest to smallest of $\mathcal{A}(\omega)$ are continuous and piece-wise analytic.

e.g.

Let $\mathcal{A}(\omega) = \begin{bmatrix} 0 & \omega \\ \omega & 0 \end{bmatrix}$ with analytic eigenvalues

$$\tilde{\lambda}_1(\omega) = \omega \text{ and } \tilde{\lambda}_2(\omega) = -\omega$$

Sorted continuous and piece-wise analytic eigenvalues

$$\lambda_1(\omega) = |\omega| \text{ and } \lambda_2(\omega) = -|\omega|$$

Analyticity Result

The analyticity result does not extend to non-Hermitian functions.

e.g.
the roots of the characteristic polynomial of

$$\mathcal{A}(\omega) = \begin{bmatrix} 0 & 1 \\ \omega & 0 \end{bmatrix}$$

are given by $\pm\sqrt{\omega}$ and not analytic.

Analyticity Result

The analyticity result does not extend to non-Hermitian functions.

e.g.
the roots of the characteristic polynomial of

$$\mathcal{A}(\omega) = \begin{bmatrix} 0 & 1 \\ \omega & 0 \end{bmatrix}$$

are given by $\pm\sqrt{\omega}$ and not analytic.

Derivatives of Eigenvalues

Let $\tilde{\lambda}(\omega)$ be one of the analytic eigenvalues with the assoc. unit eigenvector $v(\omega)$ (which also varies analytically w.r.t ω).

First Derivative

$$\tilde{\lambda}'(\omega) = v^*(\omega)\frac{d\mathcal{A}(\omega)}{d\omega}v(\omega)$$

Second Derivative

$$\tilde{\lambda}''(\omega) = v^*(\omega)\frac{d^2\mathcal{A}(\omega)}{d\omega^2}v(\omega) + 2\sum_{j,\tilde{\lambda}_j(\omega)\neq\tilde{\lambda}(\omega)}\frac{1}{\tilde{\lambda}(\omega)-\tilde{\lambda}_j(\omega)}\left|v^*(\omega)\frac{d\mathcal{A}(\omega)}{d\omega}v_j(\omega)\right|^2$$

Derivatives of Eigenvalues

Let $\tilde{\lambda}(\omega)$ be one of the analytic eigenvalues with the assoc. unit eigenvector $v(\omega)$ (which also varies analytically w.r.t ω).

First Derivative

$$\tilde{\lambda}'(\omega) = v^*(\omega)\frac{d\mathcal{A}(\omega)}{d\omega}v(\omega)$$

Second Derivative

$$\tilde{\lambda}''(\omega) = v^*(\omega)\frac{d^2\mathcal{A}(\omega)}{d\omega^2}v(\omega) + 2\sum_{j,\tilde{\lambda}_j(\omega)\neq\tilde{\lambda}(\omega)} \frac{1}{\tilde{\lambda}(\omega)-\tilde{\lambda}_j(\omega)}\left|v^*(\omega)\frac{d\mathcal{A}(\omega)}{d\omega}v_j(\omega)\right|^2$$

Derivatives of Eigenvalues

Let $\tilde{\lambda}(\omega)$ be one of the analytic eigenvalues with the assoc. unit eigenvector $v(\omega)$ (which also varies analytically w.r.t ω).

First Derivative

$$\tilde{\lambda}'(\omega) = v^*(\omega)\frac{d\mathcal{A}(\omega)}{d\omega}v(\omega)$$

Second Derivative

$$\tilde{\lambda}''(\omega) = v^*(\omega)\frac{d^2\mathcal{A}(\omega)}{d\omega^2}v(\omega) + 2\sum_{j,\tilde{\lambda}_j(\omega)\neq\tilde{\lambda}(\omega)} \frac{1}{\tilde{\lambda}(\omega) - \tilde{\lambda}_j(\omega)}\left|v^*(\omega)\frac{d\mathcal{A}(\omega)}{d\omega}v_j(\omega)\right|^2$$

Derivatives of Eigenvalues

Some observations helpful algorithmically

- Analyticity implies the boundedness of derivatives. In particular we will exploit the existence of a γ such that

$$\left|\tilde{\lambda}''(\omega)\right| \leq \gamma \quad \forall \omega.$$

- Once $\tilde{\lambda}(\omega)$ is computed, $\tilde{\lambda}'(\omega)$ is available for free.

Derivatives of Eigenvalues

Some observations helpful algorithmically

- Analyticity implies the boundedness of derivatives. In particular we will exploit the existence of a γ such that

$$\left|\tilde{\lambda}''(\omega)\right| \leq \gamma \quad \forall \omega.$$

- Once $\tilde{\lambda}(\omega)$ is computed, $\tilde{\lambda}'(\omega)$ is available for free.

Derivatives of Eigenvalues

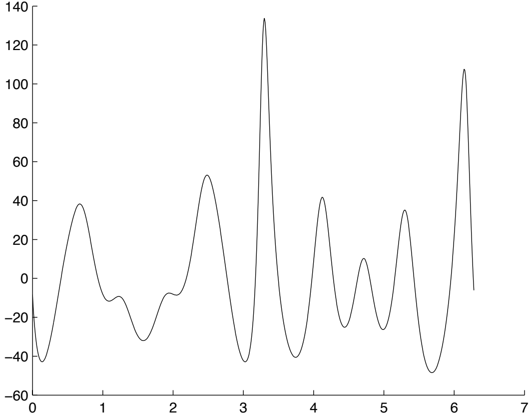
Some observations helpful algorithmically

- Analyticity implies the boundedness of derivatives. In particular we will exploit the existence of a γ such that

$$\left|\tilde{\lambda}''(\omega)\right| \leq \gamma \quad \forall \omega.$$

- Once $\tilde{\lambda}(\omega)$ is computed, $\tilde{\lambda}'(\omega)$ is available for free.

Derivatives of Eigenvalues



The plot of the second derivative of $f(\theta) := \lambda_1 \left(\frac{Ae^{i\theta} + A^* e^{-i\theta}}{2} \right)$ whose maximum over θ gives numerical radius.

Outline

Quadratic Models

The algorithm is based on quadratic models.

- Let $f : \mathbb{R} \to \mathbb{R}$ be a piece-wise analytic and continuous function in terms of analytic functions $f_1, \dots, f_n : \mathbb{R} \to \mathbb{R}$.
- The quadratic model $q_k(x)$ about $x_k \in \mathbb{R}$ satisfies
$$q_k(x_k) = f(x_k) \quad \text{and} \quad q_k'(x_k) = \underline{f}'(x_k) := \min_{j=1,n} f_j'(x_k).$$
- Furthermore for all $x \in \mathbb{R}$ the quadratic model satisfies
$$f(x) \geq q_k(x)$$

Quadratic Models

The algorithm is based on quadratic models.

- Let $f : \mathbb{R} \to \mathbb{R}$ be a piece-wise analytic and continuous function in terms of analytic functions $f_1, \ldots, f_n : \mathbb{R} \to \mathbb{R}$.
- The quadratic model $q_k(x)$ about $x_k \in \mathbb{R}$ satisfies
$$q_k(x_k) = f(x_k) \quad \text{and} \quad q_k'(x_k) = \underline{f}'(x_k) := \min_{j=1,n} f_j'(x_k).$$
- Furthermore for all $x \in \mathbb{R}$ the quadratic model satisfies
$$f(x) \geq q_k(x)$$

Quadratic Models

The algorithm is based on quadratic models.

- Let $f : \mathbb{R} \to \mathbb{R}$ be a piece-wise analytic and continuous function in terms of analytic functions $f_1, \dots, f_n : \mathbb{R} \to \mathbb{R}$.
- The quadratic model $q_k(x)$ about $x_k \in \mathbb{R}$ satisfies

$$q_k(x_k) = f(x_k) \quad \text{and} \quad q_k'(x_k) = \underline{f}'(x_k) := \min_{j=1,n} f_j'(x_k).$$

- Furthermore for all $x \in \mathbb{R}$ the quadratic model satisfies

$$f(x) \geq q_k(x)$$

Quadratic Models

The algorithm is based on quadratic models.

- Let $f : \mathbb{R} \to \mathbb{R}$ be a piece-wise analytic and continuous function in terms of analytic functions $f_1, \ldots, f_n : \mathbb{R} \to \mathbb{R}$.
- The quadratic model $q_k(x)$ about $x_k \in \mathbb{R}$ satisfies
$$q_k(x_k) = f(x_k) \quad \text{and} \quad q_k'(x_k) = \underline{f}'(x_k) := \min_{j=1,n} f_j'(x_k).$$
- Furthermore for all $x \in \mathbb{R}$ the quadratic model satisfies
$$f(x) \geq q_k(x)$$

Quadratic Models

Let γ be an upper bound on second derivatives (in abs value) of f_j and $x_{k,1}, \ldots, x_{k,m}$ be points in (x_k, x) where f is not analytic

Quadratic Models

Let γ be an upper bound on second derivatives (in abs value) of f_j and $x_{k,1}, \ldots, x_{k,m}$ be points in (x_k, x) where f is not analytic

$$f(x) = f(x_k) + \sum_{\ell=0}^{m} \int_{x_{k,\ell}}^{x_{k,\ell+1}} f'(t)dt$$

Note: $x_{k,0} = x_k$ and $x_{k,m+1} = x$

Quadratic Models

Let γ be an upper bound on second derivatives (in abs value) of f_j and $x_{k,1}, \ldots, x_{k,m}$ be points in (x_k, x) where f is not analytic

$$\begin{aligned} f(x) &= f(x_k) + \sum_{\ell=0}^{m} \int_{x_{k,\ell}}^{x_{k,\ell+1}} f'(t)dt \\ &\geq f(x_k) + \sum_{\ell=0}^{m} \int_{x_{k,\ell}}^{x_{k,\ell+1}} \underline{f}'(x_k) - \gamma(t - x_k)dt \end{aligned}$$

Note: $f'(t) \geq \underline{f}'(x_k) - \gamma(t - x_k) \;\; \forall t \in (x_k, x) \setminus \{x_{k,1}, \ldots, x_{k,m}\}$

Quadratic Models

Let γ be an upper bound on second derivatives (in abs value) of f_j and $x_{k,1}, \ldots, x_{k,m}$ be points in (x_k, x) where f is not analytic

$$
\begin{aligned}
f(x) &= f(x_k) + \sum_{\ell=0}^{m} \int_{x_{k,\ell}}^{x_{k,\ell+1}} f'(t)dt \\
&\geq f(x_k) + \int_{x_k}^{x} \underline{f}'(x_k) - \gamma(t - x_k)dt
\end{aligned}
$$

Quadratic Models

Let γ be an upper bound on second derivatives (in abs value) of f_j and $x_{k,1}, \ldots, x_{k,m}$ be points in (x_k, x) where f is not analytic

$$
\begin{aligned}
f(x) &= f(x_k) + \sum_{\ell=0}^{m} \int_{x_{k,\ell}}^{x_{k,\ell+1}} f'(t)dt \\
&\geq f(x_k) + \int_{x_k}^{x} \underline{f}'(x_k) - \gamma(t - x_k)dt \\
&= f(x_k) + \underline{f}'(x_k)(x - x_k) - \frac{\gamma}{2}(x - x_k)^2
\end{aligned}
$$

Quadratic Models

Quadratic Model about x_k

$$q_k(x) := f(x_k) + \underline{f}'(x_k)(x - x_k) - \frac{\gamma}{2}(x - x_k)^2$$

satisfies $f(x) \geq q_k(x)$ for all $x \in \mathbb{R}$.

The Algorithm

Task : locate a global minimizer of f on a given interval $[a, b]$.

1. Initially $x_0 = a$, $x_1 = b$ and $s = 1$. Evaluate $f(x_0)$, $f(x_1)$, $f'(x_0)$, and $f'(x_1)$.
2. Find the global minimizer x_* of $q(x)$ on $[a, b]$ where
$$q(x) = \max_{k=0,s} q_k(x).$$
3. Set $x_{s+1} = x_*$, evaluate $f(x_{s+1})$, $f'(x_{s+1})$.
4. Let $\ell = q(x_*)$ and $u = max_{k=0,s+1} f(x_k)$.
5. While $u - l > \epsilon$, increment s and repeat steps 2-4.

The Algorithm

Task : locate a global minimizer of f on a given interval $[a, b]$.

1. Initially $x_0 = a$, $x_1 = b$ and $s = 1$. Evaluate $f(x_0)$, $f(x_1)$, $f'(x_0)$, and $f'(x_1)$.
2. Find the global minimizer x_* of $q(x)$ on $[a, b]$ where
$$q(x) = \max_{k=0,s} q_k(x).$$
3. Set $x_{s+1} = x_*$, evaluate $f(x_{s+1})$, $f'(x_{s+1})$.
4. Let $\ell = q(x_*)$ and $u = max_{k=0,s+1} f(x_k)$.
5. While $u - l > \epsilon$, increment s and repeat steps 2-4.

The Algorithm

Task : locate a global minimizer of f on a given interval $[a, b]$.

1. Initially $x_0 = a$, $x_1 = b$ and $s = 1$. Evaluate $f(x_0)$, $f(x_1)$, $f'(x_0)$, and $f'(x_1)$.
2. Find the global minimizer x_* of $q(x)$ on $[a, b]$ where
$$q(x) = \max_{k=0,s} q_k(x).$$
3. Set $x_{s+1} = x_*$, evaluate $f(x_{s+1})$, $f'(x_{s+1})$.
4. Let $\ell = q(x_*)$ and $u = max_{k=0,s+1} f(x_k)$.
5. While $u - l > \epsilon$, increment s and repeat steps 2-4.

The Algorithm

Task : locate a global minimizer of f on a given interval $[a, b]$.

1. Initially $x_0 = a$, $x_1 = b$ and $s = 1$. Evaluate $f(x_0)$, $f(x_1)$, $f'(x_0)$, and $f'(x_1)$.
2. Find the global minimizer x_* of $q(x)$ on $[a, b]$ where
$$q(x) = \max_{k=0,s} q_k(x).$$
3. Set $x_{s+1} = x_*$, evaluate $f(x_{s+1})$, $f'(x_{s+1})$.
4. Let $\ell = q(x_*)$ and $u = max_{k=0,s+1} f(x_k)$.
5. While $u - l > \epsilon$, increment s and repeat steps 2-4.

The Algorithm

Task : locate a global minimizer of f on a given interval $[a, b]$.

1. Initially $x_0 = a$, $x_1 = b$ and $s = 1$. Evaluate $f(x_0)$, $f(x_1)$, $f'(x_0)$, and $f'(x_1)$.
2. Find the global minimizer x_* of $q(x)$ on $[a, b]$ where
$$q(x) = \max_{k=0,s} q_k(x).$$
3. Set $x_{s+1} = x_*$, evaluate $f(x_{s+1})$, $f'(x_{s+1})$.
4. Let $\ell = q(x_*)$ and $u = max_{k=0,s+1} f(x_k)$.
5. While $u - l > \epsilon$, increment s and repeat steps 2-4.

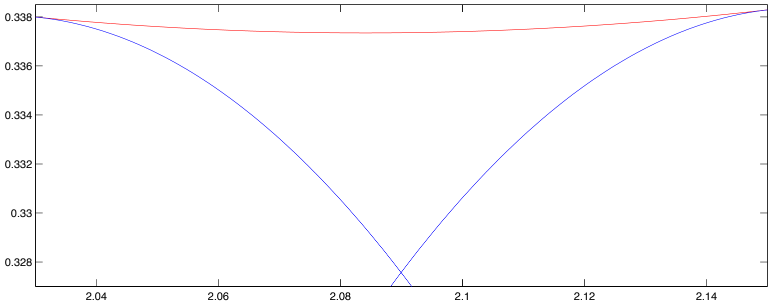
The Algorithm

Task : locate a global minimizer of f on a given interval $[a, b]$.

1. Initially $x_0 = a$, $x_1 = b$ and $s = 1$. Evaluate $f(x_0)$, $f(x_1)$, $f'(x_0)$, and $f'(x_1)$.
2. Find the global minimizer x_* of $q(x)$ on $[a, b]$ where
$$q(x) = \max_{k=0,s} q_k(x).$$
3. Set $x_{s+1} = x_*$, evaluate $f(x_{s+1})$, $f'(x_{s+1})$.
4. Let $\ell = q(x_*)$ and $u = max_{k=0,s+1} f(x_k)$.
5. While $u - l > \epsilon$, increment s and repeat steps 2-4.

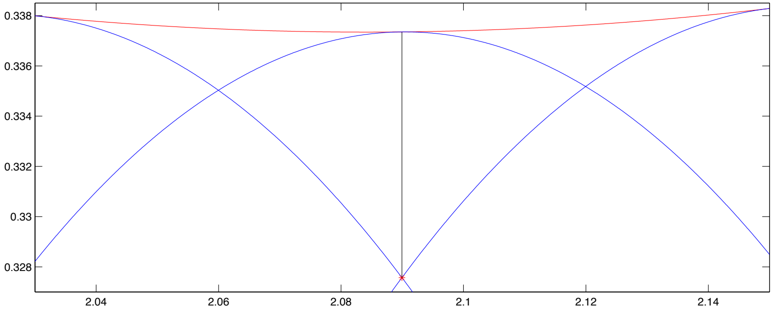
The Algorithm

Illustration of the algorithm on $\sigma_n(A - \omega iI)$ where σ_n denotes the smallest singular value.



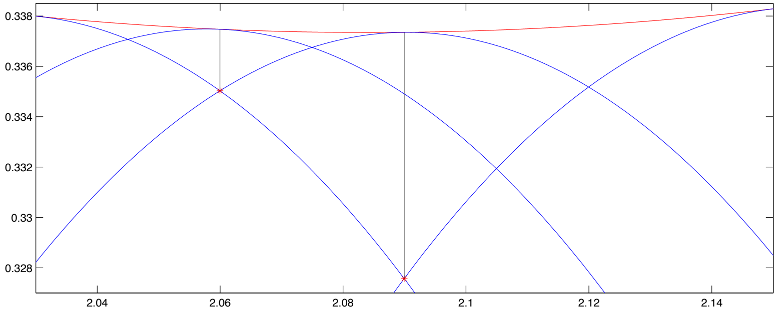
The Algorithm

Illustration of the algorithm on $\sigma_n(A - \omega i I)$ where σ_n denotes the smallest singular value.



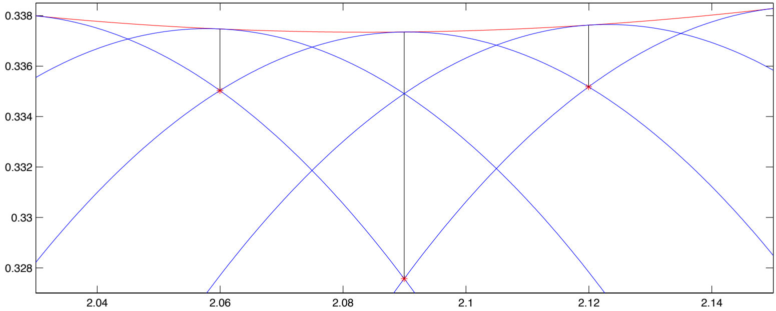
The Algorithm

Illustration of the algorithm on $\sigma_n(A - \omega i I)$ where σ_n denotes the smallest singular value.



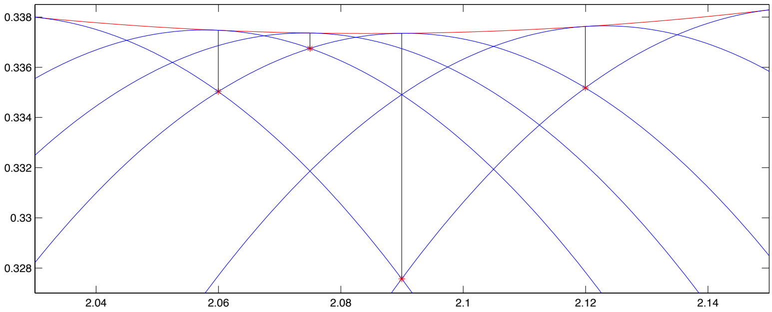
The Algorithm

Illustration of the algorithm on $\sigma_n(A - \omega i I)$ where σ_n denotes the smallest singular value.



The Algorithm

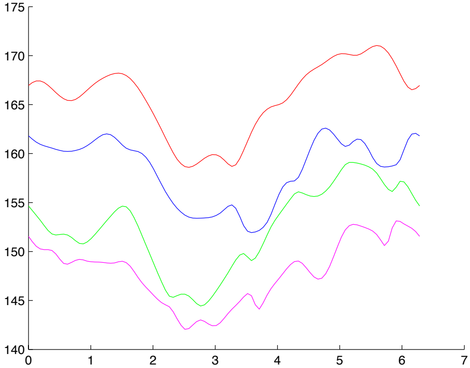
Illustration of the algorithm on $\sigma_n(A - \omega i I)$ where σ_n denotes the smallest singular value.



Generic Analyticity

Many eigenvalue functions of interest are generically analytic on a dense subset.

e.g. $f(\omega) := \lambda_1\left(\frac{Ae^{i\theta}+A^*e^{-i\theta}}{2}\right)$ is generically *analytic at all* θ.



The largest four eigenvalues of $\left(\frac{Ae^{i\theta}+A^*e^{-i\theta}}{2}\right)$

Case Study

Modulus of the outermost point in the field of values

$$F(A) = \{z^* A z \mid z \in \mathbb{C}^n \ \text{s.t.} \ \|z\|_2 = 1\}$$

is called the numerical radius.

Numerical Radius

$$r(A) := \max_{[0,2\pi)} \ \lambda_1 \left(\frac{A \cdot e^{i\theta} + A^* e^{-i\theta}}{2} \right)$$

- Field of values is a convex set.
- Contains the eigenvalues.
- Numerical radius is used to analyze the convergence of the classical iterative algorithms for linear systems.

Case Study

Modulus of the outermost point in the field of values

$$F(A) = \{z^* A z \mid z \in \mathbb{C}^n \ \text{s.t.} \ \|z\|_2 = 1\}$$

is called the numerical radius.

Numerical Radius

$$r(A) := \max_{[0,2\pi)} \ \lambda_1 \left(\frac{A \cdot e^{i\theta} + A^* e^{-i\theta}}{2} \right)$$

- Field of values is a convex set.
- Contains the eigenvalues.
- Numerical radius is used to analyze the convergence of the classical iterative algorithms for linear systems.

Case Study

Modulus of the outermost point in the field of values

$$F(A) = \{z^* A z \mid z \in \mathbb{C}^n \ \text{s.t.} \ \|z\|_2 = 1\}$$

is called the numerical radius.

Numerical Radius

$$r(A) := \max_{[0,2\pi)} \lambda_1 \left(\frac{A \cdot e^{i\theta} + A^* e^{-i\theta}}{2} \right)$$

- Field of values is a convex set.
- Contains the eigenvalues.
- Numerical radius is used to analyze the convergence of the classical iterative algorithms for linear systems.

Case Study

Modulus of the outermost point in the field of values

$$F(A) = \{z^* A z \mid z \in \mathbb{C}^n \ \text{s.t.} \ \|z\|_2 = 1\}$$

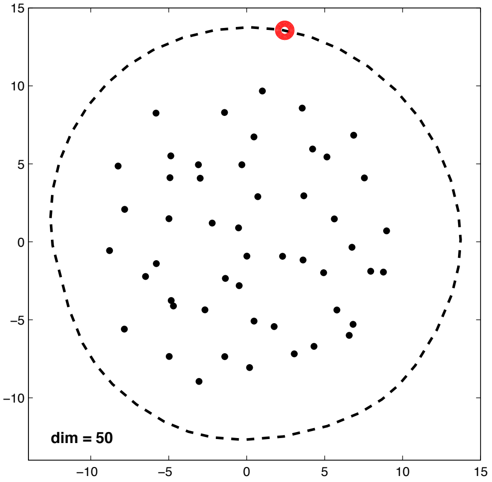
is called the numerical radius.

Numerical Radius

$$r(A) := \max_{[0,2\pi)} \ \lambda_1 \left(\frac{A \cdot e^{i\theta} + A^* e^{-i\theta}}{2} \right)$$

- Field of values is a convex set.
- Contains the eigenvalues.
- Numerical radius is used to analyze the convergence of the classical iterative algorithms for linear systems.

Case Study



Dotted-lines represent the boundary of the field of values, red circle marks the outermost point in the field of values

Case Study

Computation of numerical radius for matrices resulting from Poisson equation

\# of function evaluations

n / ϵ	10^{-4}	10^{-6}	10^{-8}	10^{-10}	10^{-12}
100	45	54	64	73	81
400	44	54	65	74	83
900	67	77	88	99	119

Case Study

Computation of numerical radius for matrices resulting from Poisson equation

cpu-times

n / ϵ	10^{-4}	10^{-6}	10^{-8}	10^{-10}	10^{-12}
100	1.0	1.2	1.4	1.6	1.9
400	9.0	10.9	12.9	14.6	17.5
900	156	177	201	225	267

Case Study

The transfer function for the linear system

$$x'(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t)$$

is given by $H(s) = C(sI - A)^{-1}B + D$.

H_∞ norm

$$\sup_{\omega \in \mathbb{R}} \sigma_1 \left(C(\omega i I - A)^{-1} B + D \right)$$

Matrices result from a discretization of the heat equation

\# of function evaluations

n / ϵ	10^{-4}	10^{-6}	10^{-8}	10^{-10}
100	23	32	39	47
200	22	29	36	44
400	18	24	29	34
800	16	19	22	27

Case Study

The transfer function for the linear system

$$x'(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t)$$

is given by $H(s) = C(sI - A)^{-1}B + D$.

H_∞ norm

$$\sup_{\omega \in \mathbb{R}} \sigma_1 \left(C(\omega iI - A)^{-1}B + D \right)$$

Matrices result from a discretization of the heat equation

# of function evaluations

n / ϵ	10^{-4}	10^{-6}	10^{-8}	10^{-10}
100	23	32	39	47
200	22	29	36	44
400	18	24	29	34
800	16	19	22	27

Case Study

The transfer function for the linear system

$$x'(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t)$$

is given by $H(s) = C(sI - A)^{-1}B + D$.

H_∞ norm

$$\sup_{\omega \in \mathbb{R}} \sigma_1 \left(C(\omega i I - A)^{-1} B + D \right)$$

Matrices result from a discretization of the heat equation

cpu-times

n / ϵ	10^{-4}	10^{-6}	10^{-8}	10^{-10}
100	0.3	0.5	0.5	0.6
200	1.5	1.9	2.3	2.8
400	8.3	10.8	12.9	17.6
800	53	63	73	92

Outline

Non-analyticity Result

- For a multivariate Hermitian function $\mathcal{A}(\omega) : \mathbb{R}^n \to \mathbb{C}^{n \times n}$ that depends on ω the eigenvalues $\tilde{\lambda}_j(\omega)$ are not analytic in general no matter how they are ordered.

  e.g. The roots of the characteristic polynomial of

$$\mathcal{A}(\omega) = \begin{bmatrix} \omega_1 & \frac{\omega_1+\omega_2}{2} \\ \frac{\omega_1+\omega_2}{2} & \omega_2 \end{bmatrix}$$

  are given by $\omega_1 + \omega_2 \pm \sqrt{2}\sqrt{\omega_1^2 + \omega_2^2}$ and not analytic.

- But there is an ordering such that $\tilde{\lambda}_j(\omega)$ for $j = 1, \ldots, n$ is analytic over any line in $\mathbb{R}^n$ (Rellich's result).

Non-analyticity Result

- For a multivariate Hermitian function $\mathcal{A}(\omega) : \mathbb{R}^n \to \mathbb{C}^{n \times n}$ that depends on ω the eigenvalues $\tilde{\lambda}_j(\omega)$ are not analytic in general no matter how they are ordered.

  e.g. The roots of the characteristic polynomial of

$$\mathcal{A}(\omega) = \begin{bmatrix} \omega_1 & \frac{\omega_1+\omega_2}{2} \\ \frac{\omega_1+\omega_2}{2} & \omega_2 \end{bmatrix}$$

  are given by $\omega_1 + \omega_2 \pm \sqrt{2}\sqrt{\omega_1^2 + \omega_2^2}$ and not analytic.

- But there is an ordering such that $\tilde{\lambda}_j(\omega)$ for $j = 1, \ldots, n$ is analytic over any line in $\mathbb{R}^n$ (Rellich's result).

Non-analyticity Result

- For a multivariate Hermitian function $\mathcal{A}(\omega) : \mathbb{R}^n \to \mathbb{C}^{n \times n}$ that depends on ω the eigenvalues $\tilde{\lambda}_j(\omega)$ are not analytic in general no matter how they are ordered.

  e.g. The roots of the characteristic polynomial of

$$\mathcal{A}(\omega) = \begin{bmatrix} \omega_1 & \frac{\omega_1+\omega_2}{2} \\ \frac{\omega_1+\omega_2}{2} & \omega_2 \end{bmatrix}$$

  are given by $\omega_1 + \omega_2 \pm \sqrt{2}\sqrt{\omega_1^2 + \omega_2^2}$ and not analytic.

- But there is an ordering such that $\tilde{\lambda}_j(\omega)$ for $j = 1, \ldots, n$ is analytic over any line in $\mathbb{R}^n$ (Rellich's result).

Model Functions

- Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be analytic over any line in $\mathbb{R}^n$, and
- The quadratic model $q_k(x)$ about $x_k \in \mathbb{R}^n$ satisfies
$$q_k(x_k) = f(x_k) \quad \text{and} \quad \nabla q_k(x_k) = \nabla f(x_k).$$
- Furthermore for all $x \in \mathbb{R}^n$ the quadratic model satisfies
$$f(x) \geq q_k(x).$$

Model Functions

- Let $f : \mathbb{R}^n \to \mathbb{R}$ be analytic over any line in $\mathbb{R}^n$, and
- The quadratic model $q_k(x)$ about $x_k \in \mathbb{R}^n$ satisfies

$$q_k(x_k) = f(x_k) \quad \text{and} \quad \nabla q_k(x_k) = \nabla f(x_k).$$

- Furthermore for all $x \in \mathbb{R}^n$ the quadratic model satisfies

$$f(x) \geq q_k(x).$$

Model Functions

- Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be analytic over any line in $\mathbb{R}^n$, and
- The quadratic model $q_k(x)$ about $x_k \in \mathbb{R}^n$ satisfies
$$q_k(x_k) = f(x_k) \quad \text{and} \quad \nabla q_k(x_k) = \nabla f(x_k).$$
- Furthermore for all $x \in \mathbb{R}^n$ the quadratic model satisfies
$$f(x) \geq q_k(x).$$

Quadratic Models

Let $\phi(\alpha) := f(x_k + \alpha p)$ where $p := (x - x_k)/\|x - x_k\|$ and γ be an upper bound on the second derivative (on any line in $\mathbb{R}^n$) of ϕ. Denote also points in the interval $[0, \|x - x_k\|]$ where $\phi(\alpha)$ is not differentiable by $\alpha^{(1)}, \ldots, \alpha^{(m)}$.

Quadratic Models

Let $\phi(\alpha) := f(x_k + \alpha p)$ where $p := (x - x_k)/\|x - x_k\|$ and γ be an upper bound on the second derivative (on any line in $\mathbb{R}^n$) of ϕ. Denote also points in the interval $[0, \|x - x_k\|]$ where $\phi(\alpha)$ is not differentiable by $\alpha^{(1)}, \ldots, \alpha^{(m)}$.

$$f(x) = f(x_k) + \sum_{\ell=0}^{m} \int_{\alpha^{(\ell)}}^{\alpha^{(\ell+1)}} \phi'(t)dt$$

Note: $\alpha^{(0)} := 0$ and $\alpha^{(m+1)} := \|x - x_k\|$.

Quadratic Models

Let $\phi(\alpha) := f(x_k + \alpha p)$ where $p := (x - x_k)/\|x - x_k\|$ and γ be an upper bound on the second derivative (on any line in $\mathbb{R}^n$) of ϕ. Denote also points in the interval $[0, \|x - x_k\|]$ where $\phi(\alpha)$ is not differentiable by $\alpha^{(1)}, \ldots, \alpha^{(m)}$.

$$\begin{aligned} f(x) &= f(x_k) + \sum_{\ell=0}^{m} \int_{\alpha^{(\ell)}}^{\alpha^{(\ell+1)}} \phi'(t)dt \\ &\geq f(x_k) + \sum_{\ell=0}^{m} \int_{\alpha^{(\ell)}}^{\alpha^{(\ell+1)}} \phi'(0) - \gamma t \, dt \end{aligned}$$

Note: $\phi'(t) = \phi'(0) - \gamma t = \nabla f(x_k)^T p - \gamma t$

Quadratic Models

Let $\phi(\alpha) := f(x_k + \alpha p)$ where $p := (x - x_k)/\|x - x_k\|$ and γ be an upper bound on the second derivative (on any line in $\mathbb{R}^n$) of ϕ. Denote also points in the interval $[0, \|x - x_k\|]$ where $\phi(\alpha)$ is not differentiable by $\alpha^{(1)}, \dots, \alpha^{(m)}$.

$$\begin{aligned} f(x) &= f(x_k) + \sum_{\ell=0}^{m} \int_{\alpha^{(\ell)}}^{\alpha^{(\ell+1)}} \phi'(t)dt \\ &\geq f(x_k) + \int_0^{\|x-x_k\|} \phi'(0) - \gamma t \, dt \end{aligned}$$

Quadratic Models

Let $\phi(\alpha) := f(x_k + \alpha p)$ where $p := (x - x_k)/\|x - x_k\|$ and γ be an upper bound on the second derivative (on any line in $\mathbb{R}^n$) of ϕ. Denote also points in the interval $[0, \|x - x_k\|]$ where $\phi(\alpha)$ is not differentiable by $\alpha^{(1)}, \ldots, \alpha^{(m)}$.

$$
\begin{aligned}
f(x) &= f(x_k) + \sum_{\ell=0}^{m} \int_{\alpha^{(\ell)}}^{\alpha^{(\ell+1)}} \phi'(t)dt \\
&\geq f(x_k) + \int_{0}^{\|x-x_k\|} \phi'(0) - \gamma t \, dt \\
&= f(x_k) + \nabla f(x_k)^T (x - x_k) - \frac{\gamma}{2}\|x - x_k\|^2
\end{aligned}
$$

Model Functions

Quadratic Model about x_k

$$q_k(x) := f(x_k) + \nabla f(x_k)^T (x - x_k) - \frac{\gamma}{2}(x - x_k)^T (x - x_k)$$

satisfies $f(x) \geq q_k(x)$ for all $x \in \mathbb{R}^n$.

The Algorithm

Task : locate a global minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ on a given box

$$\mathcal{B} := \{x \in \mathbb{R}^n \mid x_\ell \in [a_\ell, b_\ell] \text{ for } \ell = 1, \ldots, n\}$$

1. Initially let x_0 be the midpoint of the box and $s = 0$. Evaluate $f(x_0)$ and $f'(x_0)$.
2. Find the global minimizer x_* of $q(x)$ on $\mathcal{B}$ where
$$q(x) = \max_{k=0,s} q_k(x).$$
3. Set $x_{s+1} = x_*$, evaluate $f(x_{s+1}), f'(x_{s+1})$.
4. Let $\ell = q(x_*)$ and $u = max_{k=0,s+1} f(x_k)$.
5. While $u - l > \epsilon$, increment s and repeat steps 2-4.

The Algorithm

Task : locate a global minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ on a given box

$$\mathcal{B} := \{x \in \mathbb{R}^n \mid x_\ell \in [a_\ell, b_\ell] \text{ for } \ell = 1, \ldots, n\}$$

1. Initially let x_0 be the midpoint of the box and $s = 0$. Evaluate $f(x_0)$ and $f'(x_0)$.
2. Find the global minimizer x_* of $q(x)$ on $\mathcal{B}$ where
$$q(x) = \max_{k=0,s} q_k(x).$$
3. Set $x_{s+1} = x_*$, evaluate $f(x_{s+1}), f'(x_{s+1})$.
4. Let $\ell = q(x_*)$ and $u = max_{k=0,s+1} f(x_k)$.
5. While $u - l > \epsilon$, increment s and repeat steps 2-4.

The Algorithm

Task : locate a global minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ on a given box

$$\mathcal{B} := \{x \in \mathbb{R}^n \mid x_\ell \in [a_\ell, b_\ell] \text{ for } \ell = 1, \ldots, n\}$$

1. Initially let x_0 be the midpoint of the box and $s = 0$. Evaluate $f(x_0)$ and $f'(x_0)$.
2. Find the global minimizer x_* of $q(x)$ on $\mathcal{B}$ where
$$q(x) = \max_{k=0,s} q_k(x).$$
3. Set $x_{s+1} = x_*$, evaluate $f(x_{s+1}), f'(x_{s+1})$.
4. Let $\ell = q(x_*)$ and $u = max_{k=0,s+1} f(x_k)$.
5. While $u - l > \epsilon$, increment s and repeat steps 2-4.

The Algorithm

Task : locate a global minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ on a given box

$$\mathcal{B} := \{x \in \mathbb{R}^n \mid x_\ell \in [a_\ell, b_\ell] \text{ for } \ell = 1, \ldots, n\}$$

1. Initially let x_0 be the midpoint of the box and $s = 0$. Evaluate $f(x_0)$ and $f'(x_0)$.
2. Find the global minimizer x_* of $q(x)$ on $\mathcal{B}$ where
$$q(x) = \max_{k=0,s} q_k(x).$$
3. Set $x_{s+1} = x_*$, evaluate $f(x_{s+1}), f'(x_{s+1})$.
4. Let $\ell = q(x_*)$ and $u = max_{k=0,s+1} f(x_k)$.
5. While $u - l > \epsilon$, increment s and repeat steps 2-4.

The Algorithm

Task : locate a global minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ on a given box

$$\mathcal{B} := \{x \in \mathbb{R}^n \mid x_\ell \in [a_\ell, b_\ell] \text{ for } \ell = 1, \dots, n\}$$

1. Initially let x_0 be the midpoint of the box and $s = 0$. Evaluate $f(x_0)$ and $f'(x_0)$.
2. Find the global minimizer x_* of $q(x)$ on $\mathcal{B}$ where
$$q(x) = \max_{k=0,s} q_k(x).$$
3. Set $x_{s+1} = x_*$, evaluate $f(x_{s+1}), f'(x_{s+1})$.
4. Let $\ell = q(x_*)$ and $u = max_{k=0,s+1} f(x_k)$.
5. While $u - l > \epsilon$, increment s and repeat steps 2-4.

The Algorithm

Task : locate a global minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ on a given box

$$\mathcal{B} := \{x \in \mathbb{R}^n \mid x_\ell \in [a_\ell, b_\ell] \text{ for } \ell = 1, \ldots, n\}$$

1. Initially let x_0 be the midpoint of the box and $s = 0$. Evaluate $f(x_0)$ and $f'(x_0)$.
2. Find the global minimizer x_* of $q(x)$ on $\mathcal{B}$ where
$$q(x) = \max_{k=0,s} q_k(x).$$
3. Set $x_{s+1} = x_*$, evaluate $f(x_{s+1}), f'(x_{s+1})$.
4. Let $\ell = q(x_*)$ and $u = max_{k=0,s+1} f(x_k)$.
5. While $u - l > \epsilon$, increment s and repeat steps 2-4.

The Algorithm

The calculation of a global minimizer of

$$q(x) = \max_{k=0,s} q_k(x)$$

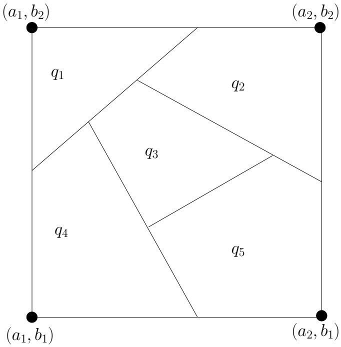
on the box $\mathcal{B}$ appears to be difficult computationally.

The Algorithm

The calculation of a global minimizer of

$$q(x) = \max_{k=0,s} q_k(x)$$

on the box $\mathcal{B}$ appears to be difficult computationally.



- Split the region where a global minimizer is known to lie into subregions.
- In subregion q_k the quadratic function $q_k(x) \geq q_j(x) \;\; \forall j \neq k$.

The Algorithm

Finding a global minimizer of $q(x) = \max_{k=0,s} q_k(x)$ **on** $\mathcal{B}$

Solve the quadratic program (QP) for $k = 0, \ldots, s$.

$$\begin{array}{ll} \text{minimize}_{x \in \mathbb{R}^n} & q_k(x) \\ \text{subject to} & q_k(x) \geq q_j(x), \quad j \neq k \\ & x_\ell \in [a_\ell, b_\ell] \quad \ell = 1, \ldots, n \end{array}$$

The Algorithm

Notes on the quadratic program

- The constraints $q_k(x) \geq q_j(x)$ are linear.
- The fact that $q_k(x)$ is negative definite makes the QP NP-hard.
- The solution will be attained at a vertex. There are at most $\binom{s+1}{n}$ vertices.
- In practice number of vertices is much smaller; for $n = 2$ typically each QP has 5-6 vertices regardless of s.
- For small n each QP can be solved efficiently.

The Algorithm

Notes on the quadratic program

- The constraints $q_k(x) \geq q_j(x)$ are linear.
- The fact that $q_k(x)$ is negative definite makes the QP NP-hard.
- The solution will be attained at a vertex. There are at most $\binom{s+1}{n}$ vertices.
- In practice number of vertices is much smaller; for $n = 2$ typically each QP has 5-6 vertices regardless of s.
- For small n each QP can be solved efficiently.

The Algorithm

Notes on the quadratic program

- The constraints $q_k(x) \geq q_j(x)$ are linear.
- The fact that $q_k(x)$ is negative definite makes the QP NP-hard.
- The solution will be attained at a vertex. There are at most $\binom{s+1}{n}$ vertices.
- In practice number of vertices is much smaller; for $n = 2$ typically each QP has 5-6 vertices regardless of s.
- For small n each QP can be solved efficiently.

The Algorithm

Notes on the quadratic program

- The constraints $q_k(x) \geq q_j(x)$ are linear.
- The fact that $q_k(x)$ is negative definite makes the QP NP-hard.
- The solution will be attained at a vertex. There are at most $\begin{pmatrix} s+1 \\ n \end{pmatrix}$ vertices.
- In practice number of vertices is much smaller; for $n = 2$ typically each QP has 5-6 vertices regardless of s.
- For small n each QP can be solved efficiently.

The Algorithm

Notes on the quadratic program

- The constraints $q_k(x) \geq q_j(x)$ are linear.
- The fact that $q_k(x)$ is negative definite makes the QP NP-hard.
- The solution will be attained at a vertex. There are at most $\begin{pmatrix} s+1 \\ n \end{pmatrix}$ vertices.
- In practice number of vertices is much smaller; for $n = 2$ typically each QP has 5-6 vertices regardless of s.
- For small n each QP can be solved efficiently.

The Algorithm

Notes on the quadratic program

- The constraints $q_k(x) \geq q_j(x)$ are linear.
- The fact that $q_k(x)$ is negative definite makes the QP NP-hard.
- The solution will be attained at a vertex. There are at most $\binom{s+1}{n}$ vertices.
- In practice number of vertices is much smaller; for $n = 2$ typically each QP has 5-6 vertices regardless of s.
- For small n each QP can be solved efficiently.

Convergence

Theorem (Convergence)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be an analytic function. Then every limit point of the sequence of iterates generated by the multi-dimensional algorithm is a global minimizer of f over the box

$$\mathcal{B} := \{x \in \mathbb{R}^n : x_\ell \in [a_\ell, b_\ell] \text{ for } \ell = 1, \ldots, n\}$$

In practice the rate of convergence appears *linear* due to the fact that when the estimates are close to global minimizers, the algorithm essentially becomes a steepest descent algorithm.

Convergence

Theorem (Convergence)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be an analytic function. Then every limit point of the sequence of iterates generated by the multi-dimensional algorithm is a global minimizer of f over the box

$$\mathcal{B} := \{x \in \mathbb{R}^n : x_\ell \in [a_\ell, b_\ell] \text{ for } \ell = 1, \ldots, n\}$$

In practice the rate of convergence appears *linear* due to the fact that when the estimates are close to global minimizers, the algorithm essentially becomes a steepest descent algorithm.

Case Study

Distance to Uncontrollability

$$\mathcal{U}(A,B) := \inf\{\|\begin{bmatrix} \Delta A & \Delta B \end{bmatrix}\|_2 \mid x'(t) = (A+\Delta A)x(t) + (B+\Delta B)u(t) \text{ is uncontrollable}\}$$
$$= \min_{z\in\mathbb{C}} \sigma_n\left(\begin{bmatrix} A - zI & B \end{bmatrix}\right)$$

Matrices resulting from heat equation

\# of function evaluations

n / ϵ	10^{-2}	10^{-4}	10^{-6}	10^{-8}
100	345	548	747	850
200	456	569	767	1066
400	615	734	849	1047

Case Study

Distance to Uncontrollability

$$\mathcal{U}(A,B) := \inf\{ \| [\ \Delta A \quad \Delta B\ ] \|_2 \mid x'(t) = (A+\Delta A)x(t) + (B+\Delta B)u(t) \text{ is uncontrollable} \}$$
$$= \min_{z \in \mathbb{C}} \sigma_n([\ A - zI \quad B\ ])$$

Matrices resulting from heat equation

\# of function evaluations

n / ϵ	10^{-2}	10^{-4}	10^{-6}	10^{-8}
100	345	548	747	850
200	456	569	767	1066
400	615	734	849	1047

Case Study

Distance to Uncontrollability

$$\mathcal{U}(A,B) := \inf\{ \| [\ \Delta A \quad \Delta B\ ] \|_2 \mid x'(t) = (A+\Delta A)x(t) + (B+\Delta B)u(t) \text{ is uncontrollable}\}$$
$$= \min_{z\in\mathbb{C}} \sigma_n \left([\ A - zI \quad B\ ] \right)$$

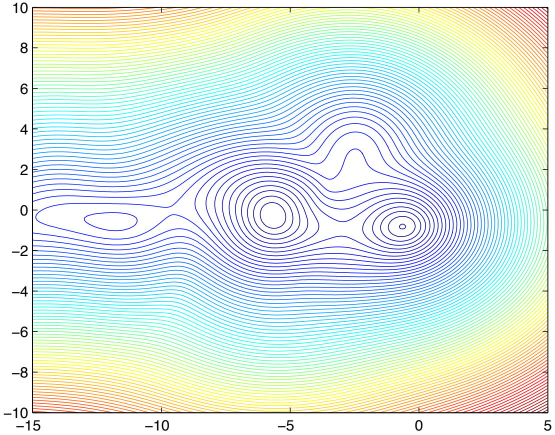
Matrices resulting from heat equation

cpu-times

n / ϵ	10^{-2}	10^{-4}	10^{-6}	10^{-8}
100	38	56	73	82
200	53	65	84	113
400	315	374	427	521

Case Study

Level sets of the function $g(z) = \sigma_n\left(\begin{bmatrix} A - zI & B \end{bmatrix}\right)$ on the complex plane.



Case Study

Wilkinson Distance

$$\mathcal{W}(A) := \inf\{\|\delta A\|_2 : \exists\lambda \;\; (A+\delta A) \text{ has } \lambda \text{ as a multiple eigenvalue}\}$$
$$= \inf_{\lambda\in\mathbb{C}} \sup_{\gamma\in\mathbb{C}} \sigma_{2n-1}\left(\begin{bmatrix} A-\lambda I & \gamma I \\ 0 & A-\lambda I \end{bmatrix}\right)$$

Random matrices

\# of function evaluations

n / ϵ	10^{-2}	10^{-3}	10^{-4}	10^{-5}
10	74	80	84	89
20	102	111	114	115
40	101	135	148	155

Case Study

Wilkinson Distance

$$\mathcal{W}(A) := \inf\{\|\delta A\|_2 : \exists\lambda \ \ (A+\delta A) \text{ has } \lambda \text{ as a multiple eigenvalue}\}$$
$$= \inf_{\lambda\in\mathbb{C}} \sup_{\gamma\in\mathbb{C}} \sigma_{2n-1}\left(\begin{bmatrix} A-\lambda I & \gamma I \\ 0 & A-\lambda I \end{bmatrix}\right)$$

Random matrices

$\#$ of function evaluations

n / ϵ	10^{-2}	10^{-3}	10^{-4}	10^{-5}
10	74	80	84	89
20	102	111	114	115
40	101	135	148	155

Case Study

Wilkinson Distance

$$\mathcal{W}(A) := \inf\{\|\delta A\|_2 : \exists\lambda \ (A+\delta A) \text{ has } \lambda \text{ as a multiple eigenvalue}\}$$
$$= \inf_{\lambda\in\mathbb{C}} \sup_{\gamma\in\mathbb{C}} \sigma_{2n-1}\left(\begin{bmatrix} A-\lambda I & \gamma I \\ 0 & A-\lambda I \end{bmatrix}\right)$$

Random matrices

cpu-times

n / ϵ	10^{-2}	10^{-3}	10^{-4}	10^{-5}
10	4.90	6.09	6.99	9.22
20	24.5	30.1	34.0	34.3
40	32.8	69.7	90.4	103.6

Outline

Problem Definition

- $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ be given scalars, and $\mathcal{S} = \{\lambda_1, \ldots, \lambda_k\}$
- r be a given positive integer
- $m_j(A, B)$: Algebraic multip of λ_j as an eigenvalue of $A - \lambda B$

Definition (Distance to Pencils with Specified Eigenvalues)

$$\tau_r(A, B, \mathcal{S}) = \inf \left\{ \|\delta A\|_2 : \sum_{j=1}^{k} m_j(A + \delta A, B) \geq r \right\}$$

Problem Definition

- $\lambda_1, \dots, \lambda_k \in \mathbb{C}$ be given scalars, and $\mathcal{S} = \{\lambda_1, \dots, \lambda_k\}$
- r be a given positive integer
- $m_j(A, B)$: Algebraic multip of λ_j as an eigenvalue of $A - \lambda B$

Definition (Distance to Pencils with Specified Eigenvalues)

$$\tau_r(A, B, \mathcal{S}) = \inf \left\{ \|\delta A\|_2 : \sum_{j=1}^{k} m_j(A + \delta A, B) \geq r \right\}$$

Problem Definition

- $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ be given scalars, and $\mathcal{S} = \{\lambda_1, \ldots, \lambda_k\}$
- r be a given positive integer
- $m_j(A, B)$: Algebraic multip of λ_j as an eigenvalue of $A - \lambda B$

Definition (Distance to Pencils with Specified Eigenvalues)

$$\tau_r(A, B, \mathcal{S}) = \inf \left\{ \|\delta A\|_2 : \sum_{j=1}^{k} m_j(A + \delta A, B) \geq r \right\}$$

Problem Definition

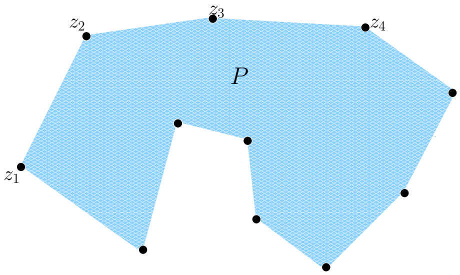
- $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ be given scalars, and $\mathcal{S} = \{\lambda_1, \ldots, \lambda_k\}$
- r be a given positive integer
- $m_j(A, B)$: Algebraic multip of λ_j as an eigenvalue of $A - \lambda B$

Definition (Distance to Pencils with Specified Eigenvalues)

$$\tau_r(A, B, \mathcal{S}) = \inf \left\{ \|\delta A\|_2 : \sum_{j=1}^{k} m_j(A + \delta A, B) \geq r \right\}$$

Shape Estimation from Moments

Estimating a polygon from moments
(Elad, Milanfar, Golub; 2004)



Given moments

$$\mathcal{M}_k = \int \int_P z^k \, dx \, dy$$

for $k = 1, \ldots, m$.

Estimate the vertices $z_j \in \mathbb{C}$ for $j = 1, \ldots, n$ of P.

Shape Estimation from Moments

- The vertices z_j are the eigenvalues of a pencil $T_0 - \lambda T_1$ where $T_0, T_1 \in \mathbb{C}^{m \times n}$ (with $m > n$) are Hankel matrices defined in terms of $\mathcal{M}_k$.
- Because of measurement errors the perturbed pencil $\widetilde{T}_0 - \lambda \widetilde{T}_1$ has generically no eigenvalues.
- Find a nearby pencil with the full set of eigenvalues

$$\inf_{\mathcal{S} \in \mathbb{C}^n} \tau_n(\widetilde{T}_0, \widetilde{T}_1, \mathcal{S}).$$

Shape Estimation from Moments

- The vertices z_j are the eigenvalues of a pencil $T_0 - \lambda T_1$ where $T_0, T_1 \in \mathbb{C}^{m \times n}$ (with $m > n$) are Hankel matrices defined in terms of $\mathcal{M}_k$.
- Because of measurement errors the perturbed pencil $\widetilde{T}_0 - \lambda \widetilde{T}_1$ has generically no eigenvalues.
- Find a nearby pencil with the full set of eigenvalues

$$\inf_{\mathcal{S} \in \mathbb{C}^n} \tau_n(\widetilde{T}_0, \widetilde{T}_1, \mathcal{S}).$$

Shape Estimation from Moments

- The vertices z_j are the eigenvalues of a pencil $T_0 - \lambda T_1$ where $T_0, T_1 \in \mathbb{C}^{m \times n}$ (with $m > n$) are Hankel matrices defined in terms of $\mathcal{M}_k$.
- Because of measurement errors the perturbed pencil $\widetilde{T}_0 - \lambda \widetilde{T}_1$ has generically no eigenvalues.
- Find a nearby pencil with the full set of eigenvalues

$$\inf_{\mathcal{S} \in \mathbb{C}^n} \ \tau_n(\widetilde{T}_0, \widetilde{T}_1, \mathcal{S}).$$

Rank Characterization

$$C(\mu, \Gamma) := \begin{bmatrix} \mu_1 & -\gamma_{21} & \dots & -\gamma_{r1} \\ 0 & \mu_2 & \dots & -\gamma_{r2} \\ & & \ddots & \\ 0 & & & \mu_r \end{bmatrix}$$

Theorem (Sylvester Characterization)

Let $A - \lambda B$ be a pencil with $A, B \in \mathbb{C}^{m \times n}$ such that $m \geq n$ and $\mathrm{rank}(B) = n$, $\mathcal{S} = \{\lambda_1, \dots, \lambda_k\}$ be a set consisting of distinct complex scalars and $r \in \mathbb{Z}^+$. Then the following two statements are equivalent.

1. $\sum_{j=1}^{k} m_j(A, B) \geq r$
2. *There exists a $\mu \in \mathcal{S}^r$ such that*

$$\dim\{X \in \mathbb{C}^{n \times r} : AX - BXC(\mu, \Gamma) = 0\} \geq r$$

for all $\Gamma \in \mathcal{G}(\mu) := \{\Gamma : C(\mu, \Gamma) \text{ has Jordan blocks of maximal size.}\}$.

Rank Characterization

$$C(\mu,\Gamma) := \begin{bmatrix} \mu_1 & -\gamma_{21} & \dots & -\gamma_{r1} \\ 0 & \mu_2 & \dots & -\gamma_{r2} \\ & & \ddots & \\ 0 & & & \mu_r \end{bmatrix}$$

Theorem (Sylvester Characterization)

Let $A - \lambda B$ be a pencil with $A, B \in \mathbb{C}^{m\times n}$ such that $m \geq n$ and $\mathrm{rank}(B) = n$, $\mathcal{S} = \{\lambda_1, \dots, \lambda_k\}$ be a set consisting of distinct complex scalars and $r \in \mathbb{Z}^+$. Then the following two statements are equivalent.

1. $\sum_{j=1}^{k} m_j(A,B) \geq r$
2. *There exists a $\mu \in \mathcal{S}^r$ such that*

$$\dim\{X \in \mathbb{C}^{n\times r} : AX - BXC(\mu,\Gamma) = 0\} \geq r$$

for all $\Gamma \in \mathcal{G}(\mu) := \{\Gamma : C(\mu,\Gamma)$ has Jordan blocks of maximal size.$\}$.

Rank Characterization

$$C(\mu,\Gamma) := \begin{bmatrix} \mu_1 & -\gamma_{21} & \dots & -\gamma_{r1} \\ 0 & \mu_2 & \dots & -\gamma_{r2} \\ & & \ddots & \\ 0 & & & \mu_r \end{bmatrix}$$

Theorem (Sylvester Characterization)

Let $A - \lambda B$ be a pencil with $A, B \in \mathbb{C}^{m\times n}$ such that $m \geq n$ and $\text{rank}(B) = n$, $\mathcal{S} = \{\lambda_1, \dots, \lambda_k\}$ be a set consisting of distinct complex scalars and $r \in \mathbb{Z}^+$. Then the following two statements are equivalent.

1. $\sum_{j=1}^{k} m_j(A, B) \geq r$
2. *There exists a $\mu \in \mathcal{S}^r$ such that*

$$\dim\{X \in \mathbb{C}^{n\times r} : AX - BXC(\mu,\Gamma) = 0\} \geq r$$

for all $\Gamma \in \mathcal{G}(\mu) := \{\Gamma : C(\mu,\Gamma)$ has Jordan blocks of maximal size.$\}$.

Rank Characterization

$$C(\mu,\Gamma) := \begin{bmatrix} \mu_1 & -\gamma_{21} & \dots & -\gamma_{r1} \\ 0 & \mu_2 & \dots & -\gamma_{r2} \\ & & \ddots & \\ 0 & & & \mu_r \end{bmatrix}$$

Theorem (Sylvester Characterization)

Let $A - \lambda B$ be a pencil with $A, B \in \mathbb{C}^{m\times n}$ such that $m \geq n$ and $\operatorname{rank}(B) = n$, $\mathcal{S} = \{\lambda_1, \dots, \lambda_k\}$ be a set consisting of distinct complex scalars and $r \in \mathbb{Z}^+$. Then the following two statements are equivalent.

1. $\sum_{j=1}^{k} m_j(A, B) \geq r$
2. *There exists a $\mu \in \mathcal{S}^r$ such that*

$$\dim\{X \in \mathbb{C}^{n\times r} : AX - BXC(\mu,\Gamma) = 0\} \geq r$$

for all $\Gamma \in \mathcal{G}(\mu) := \{\Gamma : C(\mu,\Gamma)$ has Jordan blocks of maximal size.$\}$.

Rank Characterization

Kroneckerization of the Sylvester Equation

Recall the Identity

$$\mathrm{vec}(FXG) = (G^T \otimes F)\mathrm{vec}(X)$$

where $X = \begin{bmatrix} x_1 & \dots & x_r \end{bmatrix} \in \mathbb{C}^{n \times r}$ and $\mathrm{vec}(X) = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{bmatrix} \in \mathbb{C}^{nr}$.

Rank Characterization

Kroneckerization of the Sylvester Equation

- In particular

  $AX - BXC(\mu, \Gamma) = 0 \Leftrightarrow ((I \otimes A) - (C^T(\mu, \Gamma) \otimes B))\text{vec}(X) = 0.$

- Consequently

$$\dim\{X \in \mathbb{C}^{n \times r} : AX - BXC(\mu, \Gamma) = 0\} \geq r$$
$$\Longleftrightarrow$$
$$\text{rank}\left(((I \otimes A) - (C^T(\mu, \Gamma) \otimes B))\right) \leq nr - r$$

Rank Characterization

Kroneckerization of the Sylvester Equation

- In particular
  $AX - BXC(\mu,\Gamma) = 0 \Leftrightarrow ((I \otimes A) - (C^T(\mu,\Gamma) \otimes B))\mathrm{vec}(X) = 0.$

- Consequently

$$\dim\{X \in \mathbb{C}^{n \times r} : AX - BXC(\mu,\Gamma) = 0\} \geq r$$
$$\Longleftrightarrow$$
$$\mathrm{rank}\left(((I \otimes A) - (C^T(\mu,\Gamma) \otimes B))\right) \leq nr - r$$

Rank Characterization

Kroneckerization of the Sylvester Equation

$$\mathcal{L}(\mu, \Gamma, A, B) := \left(((I \otimes A) - (C^T(\mu, \Gamma) \otimes B)\right)$$

$$= \begin{bmatrix} A - \mu_1 B & 0 & & & 0 \\ \gamma_{21} B & A - \mu_2 B & & & 0 \\ & & \ddots & & \\ & & & A - \mu_{r-1} B & 0 \\ \gamma_{r1} B & \gamma_{r2} B & & \gamma_{r(r-1)} B & A - \mu_r B \end{bmatrix}.$$

Theorem (Rank Characterization)

Given a pencil $A - \lambda B$ with $A, B \in \mathbb{C}^{m \times n}$ such that $m \geq n$ and $\mathrm{rank}(B) = n$, a set $\mathcal{S} = \{\lambda_1, \ldots, \lambda_k\}$ consisting of distinct complex scalars and $r \in \mathbb{Z}^+$. Then the following two statements are equivalent.

1. $\sum_{j=1}^{k} m_j(A, B) \geq r$
2. *There exists a $\mu \in \mathcal{S}^r$ such that*

$$\mathrm{rank}\left(\mathcal{L}(\mu, \Gamma, A, B)\right) \leq nr - r$$

for all $\Gamma \in \mathcal{G}(\mu)$.

Rank Characterization

Kroneckerization of the Sylvester Equation

$$\mathcal{L}(\mu, \Gamma, A, B) \quad := \quad \Big(((I \otimes A) - (C^T(\mu, \Gamma) \otimes B)\Big)$$

$$= \begin{bmatrix} A - \mu_1 B & 0 & & & 0 \\ \gamma_{21} B & A - \mu_2 B & & & 0 \\ & & \ddots & & \\ & & & A - \mu_{r-1} B & 0 \\ \gamma_{r1} B & \gamma_{r2} B & & \gamma_{r(r-1)} B & A - \mu_r B \end{bmatrix}.$$

Theorem (Rank Characterization)

Given a pencil $A - \lambda B$ with $A, B \in \mathbb{C}^{m \times n}$ such that $m \geq n$ and $\operatorname{rank}(B) = n$, a set $\mathcal{S} = \{\lambda_1, \ldots, \lambda_k\}$ consisting of distinct complex scalars and $r \in \mathbb{Z}^+$. Then the following two statements are equivalent.

1. $\sum_{j=1}^{k} m_j(A, B) \geq r$
2. *There exists a $\mu \in \mathcal{S}^r$ such that*

$$\operatorname{rank}\left(\mathcal{L}(\mu, \Gamma, A, B)\right) \leq nr - r$$

for all $\Gamma \in \mathcal{G}(\mu)$.

Rank Characterization

Kroneckerization of the Sylvester Equation

$$\mathcal{L}(\mu, \Gamma, A, B) \quad := \quad \Big(((I \otimes A) - (C^T(\mu, \Gamma) \otimes B)\Big)$$

$$= \begin{bmatrix} A - \mu_1 B & 0 & & & 0 \\ \gamma_{21} B & A - \mu_2 B & & & 0 \\ & & \ddots & & \\ & & & A - \mu_{r-1} B & 0 \\ \gamma_{r1} B & \gamma_{r2} B & & \gamma_{r(r-1)} B & A - \mu_r B \end{bmatrix}.$$

Theorem (Rank Characterization)

Given a pencil $A - \lambda B$ with $A, B \in \mathbb{C}^{m \times n}$ such that $m \geq n$ and $\mathrm{rank}(B) = n$, a set $\mathcal{S} = \{\lambda_1, \ldots, \lambda_k\}$ consisting of distinct complex scalars and $r \in \mathbb{Z}^+$. Then the following two statements are equivalent.

1. $\sum_{j=1}^{k} m_j(A, B) \geq r$
2. *There exists a $\mu \in \mathcal{S}^r$ such that*

$$\mathrm{rank}\left(\mathcal{L}(\mu, \Gamma, A, B)\right) \leq nr - r$$

for all $\Gamma \in \mathcal{G}(\mu)$.

Rank Characterization

Kroneckerization of the Sylvester Equation

$$\mathcal{L}(\mu, \Gamma, A, B) := \Big(((I \otimes A) - (C^T(\mu,\Gamma) \otimes B) \Big)$$
$$= \begin{bmatrix} A-\mu_1 B & 0 & & & 0 \\ \gamma_{21}B & A-\mu_2 B & & & 0 \\ & & \ddots & & \\ & & & A-\mu_{r-1}B & 0 \\ \gamma_{r1}B & \gamma_{r2}B & & \gamma_{r(r-1)}B & A-\mu_r B \end{bmatrix}.$$

Theorem (Rank Characterization)

Given a pencil $A - \lambda B$ with $A, B \in \mathbb{C}^{m\times n}$ such that $m \geq n$ and $\operatorname{rank}(B) = n$, a set $\mathcal{S} = \{\lambda_1, \ldots, \lambda_k\}$ consisting of distinct complex scalars and $r \in \mathbb{Z}^+$. Then the following two statements are equivalent.

1. $\sum_{j=1}^{k} m_j(A, B) \geq r$
2. *There exists a $\mu \in \mathcal{S}^r$ such that*
$$\operatorname{rank}(\mathcal{L}(\mu, \Gamma, A, B)) \leq nr - r$$
for all $\Gamma \in \mathcal{G}(\mu)$.

Construction of an Optimal Perturbation

Theorem (Nearest Pencils with Specified Eigenvalues)

Let $A - \lambda B$ be an $m \times n$ pencil with $m \geq n$, $r \in \mathbb{Z}^+$ and $\mathcal{S} = \{\lambda_1, \dots, \lambda_k\}$ be a set of distinct complex scalars. Then the equality

$$\tau_r(A, B, \mathcal{S}) = \inf_{\mu \in \mathcal{S}^r} \sup_{\Gamma} \sigma_{nr-r+1} \left(\mathcal{L}\left(\mu, \Gamma, A, B\right)\right)$$

holds provided that the optimization problem on the right is attained at a (μ_*, Γ_*) where the multiplicity and linear independence qualifications hold.

Construction of an Optimal Perturbation

Theorem (Nearest Pencils with Specified Eigenvalues)

Let $A - \lambda B$ be an $m \times n$ pencil with $m \geq n$, $r \in \mathbb{Z}^+$ and $\mathcal{S} = \{\lambda_1, \ldots, \lambda_k\}$ be a set of distinct complex scalars. Then the equality

$$\tau_r(A, B, \mathcal{S}) = \inf_{\mu \in \mathcal{S}^r} \sup_{\Gamma} \sigma_{nr-r+1}\left(\mathcal{L}\left(\mu, \Gamma, A, B\right)\right)$$

holds provided that the optimization problem on the right is attained at a (μ_*, Γ_*) where the multiplicity and linear independence qualifications hold.

Summary

- A generic algorithm is introduced for the optimization of symmetric eigenvalues based on their analyticity.
- The algorithm is globally convergent and the rate of convergence is linear in practice.
- Software eigopt at http://home.ku.edu.tr/~emengi/software.html
- A singular value characterization for $\tau_r(A, B, \mathcal{S})$
- Future
  - Improvements on the algorithm for the optimization of eigenvalues in the multivariate-case
  - A singular value charac for $\tau_r(A, B, \mathcal{S})$ when both A and B are perturbed.

Summary

- A generic algorithm is introduced for the optimization of symmetric eigenvalues based on their analyticity.
- The algorithm is globally convergent and the rate of convergence is linear in practice.
- Software eigopt at http://home.ku.edu.tr/∼emengi/software.html
- A singular value characterization for $\tau_r(A, B, \mathcal{S})$
- Future
  - Improvements on the algorithm for the optimization of eigenvalues in the multivariate-case
  - A singular value charac for $\tau_r(A, B, \mathcal{S})$ when both A and B are perturbed.

Summary

- A generic algorithm is introduced for the optimization of symmetric eigenvalues based on their analyticity.
- The algorithm is globally convergent and the rate of convergence is linear in practice.
- Software eigopt at http://home.ku.edu.tr/∼emengi/software.html
- A singular value characterization for $\tau_r(A, B, \mathcal{S})$
- Future
  - Improvements on the algorithm for the optimization of eigenvalues in the multivariate-case
  - A singular value charac for $\tau_r(A, B, \mathcal{S})$ when both A and B are perturbed.

Summary

- A generic algorithm is introduced for the optimization of symmetric eigenvalues based on their analyticity.
- The algorithm is globally convergent and the rate of convergence is linear in practice.
- Software eigopt at http://home.ku.edu.tr/~emengi/software.html
- A singular value characterization for $\tau_r(A, B, \mathcal{S})$

- Future
  - Improvements on the algorithm for the optimization of eigenvalues in the multivariate-case
  - A singular value charac for $\tau_r(A, B, \mathcal{S})$ when both A and B are perturbed.

Summary

- A generic algorithm is introduced for the optimization of symmetric eigenvalues based on their analyticity.
- The algorithm is globally convergent and the rate of convergence is linear in practice.
- Software eigopt at http://home.ku.edu.tr/∼emengi/software.html
- A singular value characterization for $\tau_r(A, B, \mathcal{S})$

- Future
  - Improvements on the algorithm for the optimization of eigenvalues in the multivariate-case
  - A singular value charac for $\tau_r(A, B, \mathcal{S})$ when both A and B are perturbed.

Summary

- A generic algorithm is introduced for the optimization of symmetric eigenvalues based on their analyticity.
- The algorithm is globally convergent and the rate of convergence is linear in practice.
- Software eigopt at http://home.ku.edu.tr/∼emengi/software.html
- A singular value characterization for $\tau_r(A, B, \mathcal{S})$

- Future
  - Improvements on the algorithm for the optimization of eigenvalues in the multivariate-case
  - A singular value charac for $\tau_r(A, B, \mathcal{S})$ when both A and B are perturbed.

Summary

- A generic algorithm is introduced for the optimization of symmetric eigenvalues based on their analyticity.
- The algorithm is globally convergent and the rate of convergence is linear in practice.
- Software eigopt at http://home.ku.edu.tr/∼emengi/software.html
- A singular value characterization for $\tau_r(A, B, \mathcal{S})$

- Future
  - Improvements on the algorithm for the optimization of eigenvalues in the multivariate-case
  - A singular value charac for $\tau_r(A, B, \mathcal{S})$ when both A and B are perturbed.

References

- J.H. Wilkinson, Note on matrices with a very ill-conditioned eigenproblem, *Numer Math*, 19:176-178, 1972
- A.N. Malyshev, A formula for the 2-norm distance from a matrix to the set of matrices with multiple eigenvalues, *Numer Math*, 83:443-454, 1999
- R. Alam and S. Bora, On sensitivity of eigenvalues and eigendecompositions of matrices, *Linear Algebra Appl*, 396:273-301, 2005
- E. Mengi, Locating a nearest matrix with an eigenvalue of prespecified algebraic multiplicity, *Numer Math*, 118:109-135, 2011
- D. Kressner, E. Mengi, I. Nakic and N. Truhar, Generalized eigenvalue problems with specified eigenvalues, *IMA J Num Anal*, accepted subject to minor revision
- M. Kilic, E. Mengi and E.A. Yildirim, Numerical Optimization of Eigenvalues of Hermitian Matrix Functions, *SIAM J Matrix Anal Appl*, submitted
- F. Rellich, Perturbation Theory of Eigenvalue Problems, *Gordon and Breach*, 1969
- P. Lancaster, On eigenvalues of matrices dependent on a parameter, *Numer Math*, 6:377-387, 1964