Solving Regularized Total Least Squares Problems

Heinrich Voss

voss@tu-harburg.de

Hamburg University of Technology
Institute of Numerical Simulation

Joint work with Jörg Lampe

TUHH
Technische Universität Hamburg-Harburg

Total Least Squares Problem

The ordinary Least Squares (LS) method assumes that the system matrix A of a linear model is error free, and all errors are confined to the right hand side b.

Total Least Squares Problem

The ordinary Least Squares (LS) method assumes that the system matrix A of a linear model is error free, and all errors are confined to the right hand side b.

Given $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $m \geq n$

Find $x \in \mathbb{R}^n$ and $\Delta b \in \mathbb{R}^m$ such that

$$\|\Delta b\| = \text{min!} \quad \text{subject to } Ax = b + \Delta b.$$

Total Least Squares Problem

The ordinary Least Squares (LS) method assumes that the system matrix A of a linear model is error free, and all errors are confined to the right hand side b.

Given $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $m \geq n$

Find $x \in \mathbb{R}^n$ and $\Delta b \in \mathbb{R}^m$ such that

$$\|\Delta b\| = \min! \quad \text{subject to } Ax = b + \Delta b.$$

Obviously equivalent to: Find $x \in \mathbb{R}^n$ such that

$$\|Ax - b\| = \min!$$

Total Least Squares

In practical applications it may happen that all data are contaminated by noise, for instance because the matrix A is also obtained from measurements.

Total Least Squares

In practical applications it may happen that all data are contaminated by noise, for instance because the matrix A is also obtained from measurements.

If the true values of the observed variables satisfy linear relations, and if the errors in the observations are independent random variables with zero mean and equal variance, then the total least squares (TLS) approach often gives better estimates than LS.

Total Least Squares

In practical applications it may happen that all data are contaminated by noise, for instance because the matrix A is also obtained from measurements.

If the true values of the observed variables satisfy linear relations, and if the errors in the observations are independent random variables with zero mean and equal variance, then the total least squares (TLS) approach often gives better estimates than LS.

Given $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $m \geq n$

Find $\Delta A \in \mathbb{R}^{m \times n}$, $\Delta b \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$ such that

$$\|[\Delta A, \Delta b]\|_F^2 = \min! \quad \text{subject to } (A + \Delta A)x = b + \Delta b, \tag{1}$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

Total Least Squares

In practical applications it may happen that all data are contaminated by noise, for instance because the matrix A is also obtained from measurements.

If the true values of the observed variables satisfy linear relations, and if the errors in the observations are independent random variables with zero mean and equal variance, then the total least squares (TLS) approach often gives better estimates than LS.

Given $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $m \geq n$

Find $\Delta A \in \mathbb{R}^{m \times n}$, $\Delta b \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$ such that

$$\|[\Delta A, \Delta b]\|_F^2 = \min! \quad \text{subject to } (A + \Delta A)x = b + \Delta b, \qquad (1)$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

In statistics this approach is called errors-in-variables problem or orthogonal regression, in image deblurring blind deconvolution

Regularized Total Least Squares Problem

If A and $[A, b]$ are ill-conditioned, regularization is necessary.

Regularized Total Least Squares Problem

If A and $[A, b]$ are ill-conditioned, regularization is necessary.

Let $L \in \mathbb{R}^{k\times n}$, $k \leq n$ and $\delta > 0$. Then the quadratically constrained formulation of the Regularized Total Least Squares (RTLS) problem reads:

Find $\Delta A \in \mathbb{R}^{m\times n}$, $\Delta b \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$ such that

$$\|[\Delta A, \Delta b]\|_F^2 = \min! \quad \text{subject to } (A + \Delta A)x = b + \Delta b, \ \|Lx\|^2 \leq \delta^2.$$

Regularized Total Least Squares Problem

If A and $[A, b]$ are ill-conditioned, regularization is necessary.

Let $L \in \mathbb{R}^{k \times n}$, $k \leq n$ and $\delta > 0$. Then the quadratically constrained formulation of the Regularized Total Least Squares (RTLS) problem reads:

Find $\Delta A \in \mathbb{R}^{m \times n}$, $\Delta b \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$ such that

$$\|[\Delta A, \Delta b]\|_F^2 = \min! \quad \text{subject to } (A + \Delta A)x = b + \Delta b, \; \|Lx\|^2 = \delta^2.$$

Regularized Total Least Squares Problem

If A and $[A, b]$ are ill-conditioned, regularization is necessary.

Let $L \in \mathbb{R}^{k \times n}$, $k \leq n$ and $\delta > 0$. Then the quadratically constrained formulation of the Regularized Total Least Squares (RTLS) problem reads:

Find $\Delta A \in \mathbb{R}^{m \times n}$, $\Delta b \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$ such that

$$\|[\Delta A, \Delta b]\|_F^2 = \min! \quad \text{subject to } (A + \Delta A)x = b + \Delta b, \ \|Lx\|^2 = \delta^2.$$

Using the orthogonal distance this problems can be rewritten as (cf. Golub, Van Loan 1980)

Find $x \in \mathbb{R}^n$ such that

$$f(x) := \frac{\|Ax - b\|^2}{1 + \|x\|^2} = \min! \quad \text{subject to } \|Lx\|^2 = \delta^2.$$

Regularized Total Least Squares Problem ct.

Theorem 1: Let $\mathcal{N}(L)$ be the null space of L. If

$$f^* = \inf\{f(x) \,:\, \|Lx\|^2 = \delta^2\} < \min_{x\in\mathcal{N}(L),\, x\neq 0} \frac{\|Ax\|^2}{\|x\|^2} \qquad (1)$$

then

$$f(x) := \frac{\|Ax - b\|^2}{1 + \|x\|^2} = \min! \quad \text{subject to } \|Lx\|^2 = \delta^2. \qquad (2)$$

admits a global minimum.

Regularized Total Least Squares Problem ct.

Theorem 1: Let $\mathcal{N}(L)$ be the null space of L. If

$$f^* = \inf\{f(x) \,:\, \|Lx\|^2 = \delta^2\} < \min_{x \in \mathcal{N}(L),\, x \neq 0} \frac{\|Ax\|^2}{\|x\|^2} \qquad (1)$$

then

$$f(x) := \frac{\|Ax - b\|^2}{1 + \|x\|^2} = \min! \quad \text{subject to } \|Lx\|^2 = \delta^2. \qquad (2)$$

admits a global minimum.

Conversely, if problem (2) admits a global minimum, then

$$f^* \leq \min_{x \in \mathcal{N}(L),\, x \neq 0} \frac{\|Ax\|^2}{\|x\|^2}$$

Regularized Total Least Squares Problem ct.

Under the condition (1) problem (2) is equivalent to the quadratic optimization problem

$$\|Ax - b\|^2 - f^*(1 + \|x\|^2) = \min! \quad \text{subject to } \|Lx\|^2 = \delta^2, \qquad (3)$$

i.e. x^* is a global minimizer of problem (2) if and only if it is a global minimizer of (3).

Regularized Total Least Squares Problem ct.

Under the condition (1) problem (2) is equivalent to the quadratic optimization problem

$$\|Ax - b\|^2 - f^*(1 + \|x\|^2) = \min! \quad \text{subject to } \|Lx\|^2 = \delta^2, \qquad (3)$$

i.e. x^* is a global minimizer of problem (2) if and only if it is a global minimizer of (3).

For fixed $y \in \mathbb{R}^n$ find $x \in \mathbb{R}^n$ such that

$$g(x; y) := \|Ax - b\|^2 - f(y)(1 + \|x\|^2) = \min! \quad \text{subject to } \|Lx\|^2 = \delta^2. \qquad (P_y)$$

Regularized Total Least Squares Problem ct.

Under the condition (1) problem (2) is equivalent to the quadratic optimization problem

$$\|Ax - b\|^2 - f^*(1 + \|x\|^2) = \min! \quad \text{subject to } \|Lx\|^2 = \delta^2, \qquad (3)$$

i.e. x^* is a global minimizer of problem (2) if and only if it is a global minimizer of (3).

For fixed $y \in \mathbb{R}^n$ find $x \in \mathbb{R}^n$ such that

$$g(x; y) := \|Ax - b\|^2 - f(y)(1 + \|x\|^2) = \min! \quad \text{subject to } \|Lx\|^2 = \delta^2. \qquad (P_y)$$

Lemma 1 (Sima, Van Huffel, Golub 2004)
Problem (P_y) admits a global minimizer if and only if

$$f(y) \le \min_{x \in \mathcal{N}(L), x \neq 0} \frac{x^T A^T A x}{x^T x}.$$

RTLSQEP Method (Sima, Van Huffel, Golub 2004)

Lemma 2

Assume that y satisfies conditions of Lemma 1 and $\|Ly\| = \delta$, and let z be a global minimizer of problem (P_y). Then it holds that

$$f(z) \leq f(y).$$

RTLSQEP Method (Sima, Van Huffel, Golub 2004)

Lemma 2

Assume that y satisfies conditions of Lemma 1 and $\|Ly\| = \delta$, and let z be a global minimizer of problem (P_y). Then it holds that

$$f(z) \leq f(y).$$

Proof

$$(1 + \|z\|)^2(f(z) - f(y)) = g(z; y) \leq g(y; y) = (1 + \|y\|^2)(f(y) - f(y)) = 0.$$

RTLSQEP Method (Sima, Van Huffel, Golub 2004)

Lemma 2
Assume that y satisfies conditions of Lemma 1 and $\|Ly\| = \delta$, and let z be a global minimizer of problem (P_y). Then it holds that

$$f(z) \le f(y).$$

Proof

$$(1 + \|z\|)^2 (f(z) - f(y)) = g(z; y) \le g(y; y) = (1 + \|y\|^2)(f(y) - f(y)) = 0.$$

Require: x^0 satisfying conditions of Lemma 1 and $\|Lx^0\| = \delta$.
for $m = 0, 1, 2, \dots$ until convergence **do**
Determine global minimizer x^{m+1} of

$$g(x; x^m) = \min! \quad \text{subject to } \|Lx\|^2 = \delta^2.$$

end for

RTLS Method

Obviously,

$$0 \leq f(x^{m+1}) \leq f(x^m)$$

RTLS Method

Obviously,

$$0 \leq f(x^{m+1}) \leq f(x^m)$$

The (P_{x^m}) can be solved via the first order necessary optimality conditions

$$(A^T A - f(x^m)I)x + \lambda L^T L x = A^T b, \quad \|Lx\|^2 = \delta^2.$$

RTLS Method

Obviously,

$$0 \le f(x^{m+1}) \le f(x^m)$$

The (P_{x^m}) can be solved via the first order necessary optimality conditions

$$(A^T A - f(x^m)I)x + \lambda L^T Lx = A^T b, \quad \|Lx\|^2 = \delta^2.$$

Although $g(\cdot; x^m)$ in general is not convex these conditions are even sufficient if the Lagrange parameter is chosen maximal.

RTLS Method

Theorem 2
Assume that $(\hat{\lambda}, \hat{x})$ solves the first order conditions.

$$(A^T A - f(y)I)x + \lambda L^T Lx = A^T b, \quad \|Lx\|^2 = \delta^2. \qquad (*)$$

If $\|Ly\| = \delta$ and $\hat{\lambda}$ is the maximal Lagrange multiplier then $\hat{x}$ is a global minimizer of problem (P_y).

RTLS Method

Theorem 2
Assume that $(\hat{\lambda}, \hat{x})$ solves the first order conditions.

$$(A^T A - f(y)I)x + \lambda L^T Lx = A^T b, \quad \|Lx\|^2 = \delta^2. \qquad (*)$$

If $\|Ly\| = \delta$ and $\hat{\lambda}$ is the maximal Lagrange multiplier then $\hat{x}$ is a global minimizer of problem (P_y).

Proof
The statement follows immediately from the following equation which can be shown similarly as in W. Gander (1981):
If (λ_j, z^j), $j = 1, 2$, are solutions of $(*)$ then it holds that

$$g(z^2; y) - g(z^1; y) = \frac{1}{2}(\lambda_1 - \lambda_2)\|L(z^1 - z^2)\|^2.$$

A quadratic eigenproblem

$$(A^TA - f(y)I)x + \lambda L^TLx = A^Tb, \quad \|Lx\|^2 = \delta^2 \quad (*)$$

A quadratic eigenproblem

$$(A^TA - f(y)I)x + \lambda L^TLx = A^Tb, \quad \|Lx\|^2 = \delta^2 \quad (*)$$

If L is square and nonsingular: Let $z := Lx$. Then

$$Wz + \lambda z := L^{-T}(A^TA - f(y)I)L^{-1}z + \lambda z = L^{-T}A^Tb =: h, \; z^Tz = \delta^2.$$

A quadratic eigenproblem

$$(A^TA - f(y)I)x + \lambda L^TLx = A^Tb, \quad \|Lx\|^2 = \delta^2 \quad (*)$$

If L is square and nonsingular: Let $z := Lx$. Then

$$Wz + \lambda z := L^{-T}(A^TA - f(y)I)L^{-1}z + \lambda z = L^{-T}A^Tb =: h, \; z^Tz = \delta^2.$$

$$u := (W + \lambda I)^{-2}h \quad \Rightarrow \quad h^Tu = z^Tz = \delta^2 \quad \Rightarrow \quad h = \delta^{-2}hh^Tu$$

A quadratic eigenproblem

$$(A^TA - f(y)I)x + \lambda L^TLx = A^Tb, \quad \|Lx\|^2 = \delta^2 \quad (*)$$

If L is square and nonsingular: Let $z := Lx$. Then

$$Wz + \lambda z := L^{-T}(A^TA - f(y)I)L^{-1}z + \lambda z = L^{-T}A^Tb =: h, \; z^Tz = \delta^2.$$

$$u := (W + \lambda I)^{-2}h \quad \Rightarrow \quad h^Tu = z^Tz = \delta^2 \quad \Rightarrow \quad h = \delta^{-2}hh^Tu$$

$$(W + \lambda I)^2u - \delta^{-2}hh^Tu = 0.$$

A quadratic eigenproblem

$$(A^TA - f(y)I)x + \lambda L^TLx = A^Tb, \quad \|Lx\|^2 = \delta^2 \quad (*)$$

If L is square and nonsingular: Let $z := Lx$. Then

$$Wz + \lambda z := L^{-T}(A^TA - f(y)I)L^{-1}z + \lambda z = L^{-T}A^Tb =: h, \; z^Tz = \delta^2.$$

$$u := (W + \lambda I)^{-2}h \quad \Rightarrow \quad h^Tu = z^Tz = \delta^2 \quad \Rightarrow \quad h = \delta^{-2}hh^Tu$$

$$(W + \lambda I)^2 u - \delta^{-2}hh^Tu = 0.$$

If $\hat{\lambda}$ is the right-most real eigenvalue, and the corresponding eigenvector is scaled such that $h^Tu = \delta^2$ then the solution of problem $(*)$ is recovered as $x = L^{-1}(W + \hat{\lambda}I)u$.

A quadratic eigenproblem ct.

If $L \in \mathbb{R}^{k\times n}$ with linearly independent rows and $k < n$, the first order conditions can be reduced to a quadratic eigenproblem

$$(W + \lambda I)^2 u - \delta^{-2} hh^T u = 0.$$

where

$$\begin{aligned} W_m &= \left(C - f(x^m)D - S(T - f(x^m)I_{n-k})^{-1} S^T\right) \\ h_m &= g - D(T - f(x^m)I_{n-k})^{-1} c \end{aligned}$$

with $C, D \in \mathbb{R}^{k\times k}$, $S \in \mathbb{R}^{k\times n-k}$, $T \in \mathbb{R}^{n-k\times n-k}$, $g \in \mathbb{R}^k$, $c \in \mathbb{R}^{n-k}$, and C, D, T are symmetric matrices.

Nonlinear maxmin characterization

Let $T(\lambda) \in \mathbb{C}^{n \times n}$, $T(\lambda) = T(\lambda)^H$, $\lambda \in J \subset \mathbb{R}$ an open interval (maybe unbounded).

Nonlinear maxmin characterization

Let $T(\lambda) \in \mathbb{C}^{n \times n}$, $T(\lambda) = T(\lambda)^H$, $\lambda \in J \subset \mathbb{R}$ an open interval (maybe unbounded).

For every fixed $x \in \mathbb{C}^n$, $x \neq 0$ assume that the real function

$$f(\cdot; x) \, : \, J \to \mathbb{R}, \; f(\lambda; x) := x^H T(\lambda) x$$

is continuous, and that the real equation

$$f(\lambda, x) = 0$$

has at most one solution $\lambda =: p(x)$ in J.

Nonlinear maxmin characterization

Let $T(\lambda) \in \mathbb{C}^{n\times n}$, $T(\lambda) = T(\lambda)^H$, $\lambda \in J \subset \mathbb{R}$ an open interval (maybe unbounded).

For every fixed $x \in \mathbb{C}^n$, $x \neq 0$ assume that the real function

$$f(\cdot; x) \; : \; J \to \mathbb{R}, \;\; f(\lambda; x) := x^H T(\lambda) x$$

is continuous, and that the real equation

$$f(\lambda, x) = 0$$

has at most one solution $\lambda =: p(x)$ in J.

Then equation $f(\lambda, x) = 0$ implicitly defines a functional p on some subset D of $\mathbb{C}^n$ which we call the Rayleigh functional.

Nonlinear maxmin characterization

Let $T(\lambda) \in \mathbb{C}^{n \times n}$, $T(\lambda) = T(\lambda)^H$, $\lambda \in J \subset \mathbb{R}$ an open interval (maybe unbounded).

For every fixed $x \in \mathbb{C}^n$, $x \neq 0$ assume that the real function

$$f(\cdot; x) : J \to \mathbb{R}, \; f(\lambda; x) := x^H T(\lambda) x$$

is continuous, and that the real equation

$$f(\lambda, x) = 0$$

has at most one solution $\lambda =: p(x)$ in J.

Then equation $f(\lambda, x) = 0$ implicitly defines a functional p on some subset D of $\mathbb{C}^n$ which we call the Rayleigh functional.

Assume that

$$(\lambda - p(x)) f(\lambda, x) > 0 \quad \text{for every } x \in D, \; \lambda \neq p(x).$$

maxmin characterization (V., Werner 1982)

Let $\sup_{v \in D} p(v) \in J$ and assume that there exists a subspace $V \subset \mathbb{C}^n$ of dimension ℓ such that

$$V \cap D \neq \emptyset \quad \text{and} \quad \inf_{v \in V \cap D} p(v) \in J.$$

maxmin characterization (V., Werner 1982)

Let $\sup_{v\in D} p(v) \in J$ and assume that there exists a subspace $V \subset \mathbb{C}^n$ of dimension ℓ such that

$$V \cap D \neq \emptyset \quad \text{and} \quad \inf_{v\in V\cap D} p(v) \in J.$$

- Then $T(\lambda)x = 0$ has at least ℓ eigenvalues in J, and for $j = 1, \dots, \ell$ the j-largest eigenvalue λ_j can be characterized by

$$\lambda_j = \max_{\substack{\dim V=j,\\ V\cap D\neq\emptyset}} \inf_{v\in V\cap D} p(v). \qquad (1)$$

maxmin characterization (V., Werner 1982)

Let $\sup_{v \in D} p(v) \in J$ and assume that there exists a subspace $V \subset \mathbb{C}^n$ of dimension ℓ such that

$$V \cap D \neq \emptyset \quad \text{and} \quad \inf_{v \in V \cap D} p(v) \in J.$$

- Then $T(\lambda)x = 0$ has at least ℓ eigenvalues in J, and for $j = 1, \dots, \ell$ the j-largest eigenvalue λ_j can be characterized by

$$\lambda_j = \max_{\substack{\dim V = j, \\ V \cap D \neq \emptyset}} \inf_{v \in V \cap D} p(v). \qquad (1)$$

- For $j = 1, \dots, \ell$ every j dimensional subspace $\tilde{V} \subset \mathbb{C}^n$ with

$$\tilde{V} \cap D \neq \emptyset \quad \text{and} \quad \lambda_j = \inf_{v \in \tilde{V} \cap D} p(v)$$

is contained in $D \cup \{0\}$, and the maxmin characterization of λ_j can be replaced by

$$\lambda_j = \max_{\substack{\dim V = j, \\ V \setminus \{0\} \subset D}} \min_{v \in V \setminus \{0\}} p(v).$$

Back to

$$T(\lambda)x := \left((W + \lambda I)^2 - \delta^{-2} hh^T\right)x = 0 \qquad (QEP)$$

Back to

$$T(\lambda)x := \left((W + \lambda I)^2 - \delta^{-2} hh^T\right)x = 0 \qquad (QEP)$$

$$f(\lambda, x) = x^H T(\lambda)x = \lambda^2\|x\|^2 + 2\lambda x^H Wx + \|Wx\|^2 - |x^H h|^2/\delta^2, \;\; x \neq 0$$

is a parabola which attains its minimum at

$$\lambda = -\frac{x^H Wx}{x^H x}.$$

Back to

$$T(\lambda)x := \left((W+\lambda I)^2 - \delta^{-2}hh^T\right)x = 0 \qquad (QEP)$$

$$f(\lambda, x) = x^H T(\lambda)x = \lambda^2\|x\|^2 + 2\lambda x^H Wx + \|Wx\|^2 - |x^H h|^2/\delta^2,\ \ x \neq 0$$

is a parabola which attains its minimum at

$$\lambda = -\frac{x^H Wx}{x^H x}.$$

Let $J = (-\lambda_{\min}, \infty)$ where $\lambda_{\min}$ is the minimum eigenvalue of W. Then $f(\lambda, x) = 0$ has at most one solution $p(x) \in J$ for every $x \neq 0$. Hence, the Rayleigh functional p of (QEP) corresponding to J is defined, and the general conditions are satisfied.

Characterization of maximal real eigenvalue

Let $V_{\min}$ be the eigenspace of W corresponding to $\lambda_{\min}$. Then for every $x_{\min} \in V_{\min}$

$$f(-\lambda_{\min}, x_{min}) = x_{\min}^H (W - \lambda_{\min})^2 x_{\min} - |x_{\min}^H h|^2/\delta^2 = -|x_{\min}^H h|^2/\delta^2 \leq 0$$

Characterization of maximal real eigenvalue

Let $V_{\min}$ be the eigenspace of W corresponding to $\lambda_{\min}$. Then for every $x_{\min} \in V_{\min}$

$$f(-\lambda_{\min}, x_{min}) = x_{\min}^H (W - \lambda_{\min})^2 x_{\min} - |x_{\min}^H h|^2/\delta^2 = -|x_{\min}^H h|^2/\delta^2 \leq 0$$

Hence, if $x_{min}^H h \neq 0$ for some $x_{\min} \in V_{\min}$, then $x_{\min} \in D$.

Characterization of maximal real eigenvalue

Let $V_{\min}$ be the eigenspace of W corresponding to $\lambda_{\min}$. Then for every $x_{\min} \in V_{\min}$

$$f(-\lambda_{\min}, x_{min}) = x_{\min}^H (W - \lambda_{\min})^2 x_{\min} - |x_{\min}^H h|^2/\delta^2 = -|x_{\min}^H h|^2/\delta^2 \leq 0$$

Hence, if $x_{min}^H h \neq 0$ for some $x_{\min} \in V_{\min}$, then $x_{\min} \in D$.

If $h \perp V_{\min}$, and if the minimum eigenvalue $\mu_{\min}$ of $T(-\lambda_{\min})$ is negative, then for the corresponding eigenvector y_{min} it holds

$$f(-\lambda_{\min}, y_{\min}) = y_{\min}^H T(-\lambda_{\min}) y_{\min} = \mu_{min} \|y_{\min}\|^2 < 0,$$

and $y_{\min} \in D$.

Characterization of maximal real eigenvalue

Let $V_{\min}$ be the eigenspace of W corresponding to $\lambda_{\min}$. Then for every $x_{\min} \in V_{\min}$

$$f(-\lambda_{\min}, x_{min}) = x_{\min}^H (W - \lambda_{\min})^2 x_{\min} - |x_{\min}^H h|^2/\delta^2 = -|x_{\min}^H h|^2/\delta^2 \le 0$$

Hence, if $x_{min}^H h \neq 0$ for some $x_{\min} \in V_{\min}$, then $x_{\min} \in D$.

If $h \perp V_{\min}$, and if the minimum eigenvalue $\mu_{\min}$ of $T(-\lambda_{\min})$ is negative, then for the corresponding eigenvector y_{min} it holds

$$f(-\lambda_{\min}, y_{\min}) = y_{\min}^H T(-\lambda_{\min}) y_{\min} = \mu_{min} \|y_{\min}\|^2 < 0,$$

and $y_{\min} \in D$.

If $h \perp V_{\min}$, and $T(-\lambda_{\min})$ is positive semi-definite, then

$$f(-\lambda_{\min}, x) = x^H T(-\lambda_{\min}) x \ge 0 \quad \text{for every } x \neq 0,$$

and $D = \emptyset$.

Characterization of maximal real eigenvalue ct.

Assume that $D \neq \emptyset$. For $x^H h = 0$ it holds that

$$f(\lambda, x) = \|(W + \lambda I)x\|^2 > 0 \quad \text{for every } \lambda \in J,$$

i.e. $x \notin D$.

Characterization of maximal real eigenvalue ct.

Assume that $D \neq \emptyset$. For $x^H h = 0$ it holds that

$$f(\lambda, x) = \|(W + \lambda I)x\|^2 > 0 \quad \text{for every } \lambda \in J,$$

i.e. $x \notin D$.

Hence, D does not contain a two-dimensional subspace of $\mathbb{R}^n$, and therefore J contains at most one eigenvalue of (QEP).

Characterization of maximal real eigenvalue ct.

Assume that $D \neq \emptyset$. For $x^H h = 0$ it holds that

$$f(\lambda, x) = \|(W + \lambda I)x\|^2 > 0 \quad \text{for every } \lambda \in J,$$

i.e. $x \notin D$.

Hence, D does not contain a two-dimensional subspace of $\mathbb{R}^n$, and therefore J contains at most one eigenvalue of (QEP).

If $\lambda \in \mathbb{C}$ is a non-real eigenvalue of (QEP) and x a corresponding eigenvector, then

$$x^H T(\lambda)x = \lambda^2 \|x\|^2 + 2\lambda x^H Wx + \|Wx\|^2 - |x^H h|^2/\delta^2 = 0.$$

Hence, the real part of λ satisfies

$$\text{real}(\lambda) = -\frac{x^H Wx}{x^H x} \leq -\lambda_{\min}.$$

Theorem 3

Let $\lambda_{\min}$ be the minimal eigenvalue of W, and $V_{\min}$ be the corresponding eigenspace.

Theorem 3

Let $\lambda_{\min}$ be the minimal eigenvalue of W, and $V_{\min}$ be the corresponding eigenspace.

- If $h \perp V_{\min}$ and $T(-\lambda_{min})$ is positive semi-definite, then $\hat{\lambda} := -\lambda_{\min}$ is the maximal real eigenvalue of (QEP).

Theorem 3

Let $\lambda_{\min}$ be the minimal eigenvalue of W, and $V_{\min}$ be the corresponding eigenspace.

- If $h \perp V_{\min}$ and $T(-\lambda_{min})$ is positive semi-definite, then $\hat{\lambda} := -\lambda_{\min}$ is the maximal real eigenvalue of (QEP).
- Otherwise, the maximal real eigenvalue is the unique eigenvalue $\hat{\lambda}$ of (QEP) in $J = (-\lambda_{\min}, \infty)$, and it holds

$$\hat{\lambda} = \max_{x \in D} p(x).$$

Theorem 3

Let $\lambda_{\min}$ be the minimal eigenvalue of W, and $V_{\min}$ be the corresponding eigenspace.

- If $h \perp V_{\min}$ and $T(-\lambda_{min})$ is positive semi-definite, then $\hat{\lambda} := -\lambda_{\min}$ is the maximal real eigenvalue of (QEP).

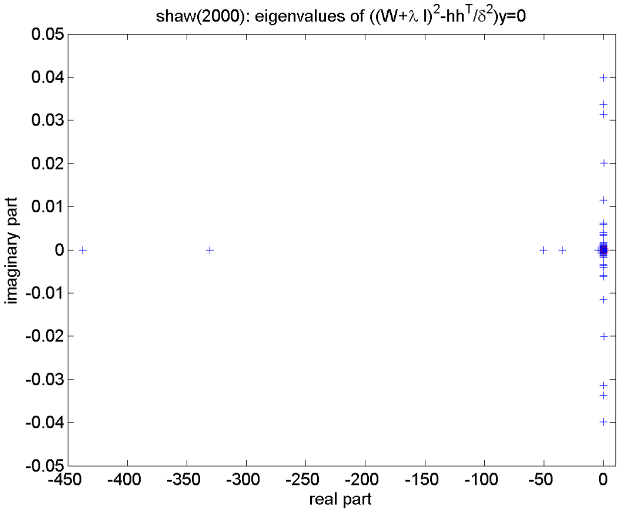
- Otherwise, the maximal real eigenvalue is the unique eigenvalue $\hat{\lambda}$ of (QEP) in $J = (-\lambda_{\min}, \infty)$, and it holds

$$\hat{\lambda} = \max_{x \in D} p(x).$$

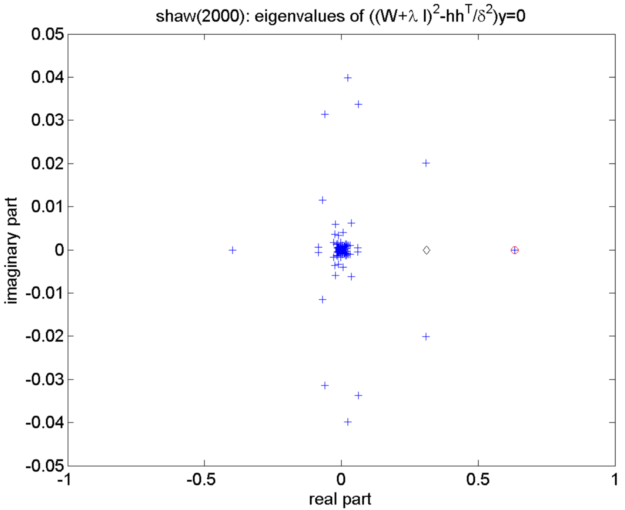
- $\hat{\lambda}$ is the right most eigenvalue of (QEP), i.e.

$$\text{real}(\lambda) \le -\lambda_{\min} \le \hat{\lambda} \quad \text{for every eigenvalue } \lambda \ne \hat{\lambda} \text{ of (QEP).}$$

Example


shaw(2000): eigenvalues of ((W+λ I)^2-hh^T/δ^2)y=0
0.05
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
-0.05
imaginary part
-450
-400
-350
-300
-250
-200
-150
-100
-50
0
real part

Example: close up



Positivity of $\hat{\lambda}$

Sima et al. claimed that the right-most eigenvalue problem is always positive.

Positivity of $\hat{\lambda}$

Sima et al. claimed that the right-most eigenvalue problem is always positive.

Simplest counter–example: If W is positive definite with eigenvalue $\lambda_j > 0$, then $-\lambda_j$ are the only eigenvalues of the quadratic eigenproblem $(W + \lambda I)^2 x = 0$, and if the term $\delta^{-2} hh^T$ is small enough, then the quadratic problem will have no positive eigenvalue, but the right–most eigenvalue will be negative.

Positivity of $\hat{\lambda}$

Sima et al. claimed that the right-most eigenvalue problem is always positive.

Simplest counter–example: If W is positive definite with eigenvalue $\lambda_j > 0$, then $-\lambda_j$ are the only eigenvalues of the quadratic eigenproblem $(W + \lambda I)^2 x = 0$, and if the term $\delta^{-2} hh^T$ is small enough, then the quadratic problem will have no positive eigenvalue, but the right–most eigenvalue will be negative.

However, in quadratic eigenproblems occurring in regularized total least squares problems δ and h are not arbitrary, but regularization only makes sense if $\delta \leq \|Lx_{\text{TLS}}\|$ where x_{TLS} denotes the solution of the total least squares problem without regularization.

The following theorem characterizes the case that the right–most eigenvalue is negative.

Positivity of $\hat\lambda$ ct.

Theorem 4
The maximal real eigenvalue $\hat\lambda$ of the quadratic problem

$$(W + \lambda I)^2 x - \delta^{-2} h h^T x = 0$$

is negative if and only if W is positive definite and

$$\|W^{-1} h\| < \delta.$$

Positivity of $\hat\lambda$ ct.

Theorem 4
The maximal real eigenvalue $\hat\lambda$ of the quadratic problem

$$(W+\lambda I)^2 x - \delta^{-2} h h^T x = 0$$

is negative if and only if W is positive definite and

$$\|W^{-1}h\| < \delta.$$

For the quadratic eigenproblem occuring in regularized total least squares it holds that

$$\|W^{-1}h\| = \|L(A^TA - f(x)I)^{-1}A^Tb\|.$$

Positivity of $\hat\lambda$ ct.

Theorem 4
The maximal real eigenvalue $\hat\lambda$ of the quadratic problem

$$(W+\lambda I)^2x-\delta^{-2}hh^Tx=0$$

is negative if and only if W is positive definite and

$$\|W^{-1}h\|<\delta.$$

For the quadratic eigenproblem occuring in regularized total least squares it holds that

$$\|W^{-1}h\|=\|L(A^TA-f(x)I)^{-1}A^Tb\|.$$

For the standard case $L=I$ the right-most eigenvalue $\hat\lambda$ is always nonnegative if $\delta<\|x_{TLS}\|$.

Convergence

Theorem 5
Any limit point x^* of the sequence $\{x^m\}$ constructed by RTLSQEP is a global minimizer of

$$f(x) = \frac{\|Ax - b\|^2}{1 + \|x\|^2} \quad \text{subject to } \|Lx\|^2 = \delta^2.$$

Convergence

Theorem 5
Any limit point x^* of the sequence $\{x^m\}$ constructed by RTLSQEP is a global minimizer of

$$f(x) = \frac{\|Ax - b\|^2}{1 + \|x\|^2} \quad \text{subject to } \|Lx\|^2 = \delta^2.$$

Proof: Let x^* be a limit point of $\{x^m\}$, and let $\{x^{m_j}\}$ be a subsequence converging to x^*. Then x^{m_j} solves the first order conditions

$$(A^T A - f(x^{m_j - 1})I)x^{m_j} + \lambda_{m_j} L^T L x^{m_j} = A^T b.$$

Convergence

Theorem 5
Any limit point x^* of the sequence $\{x^m\}$ constructed by RTLSQEP is a global minimizer of

$$f(x) = \frac{\|Ax - b\|^2}{1 + \|x\|^2} \quad \text{subject to } \|Lx\|^2 = \delta^2.$$

Proof: Let x^* be a limit point of $\{x^m\}$, and let $\{x^{m_j}\}$ be a subsequence converging to x^*. Then x^{m_j} solves the first order conditions

$$(A^T A - f(x^{m_j - 1})I)x^{m_j} + \lambda_{m_j} L^T L x^{m_j} = A^T b.$$

From the monotonicity of $f(x^m)$ it follows that $\lim_{j\to\infty} f(x^{m_j-1}) = f(x^*)$

Proof ct.

Since $W(y)$ and $h(y)$ depend continuously on y the sequence of right-most eigenvalues $\{\lambda_{m_{j'}}\}$ converges to some λ^*, and x^* satisfies

$$(A^TA - f(x^*)I)x^* - \lambda^* L^T L x^* = A^T b, \quad \|Lx^*\|^2 = \delta^2,$$

where λ^* is the maximal Lagrange multiplier.

Proof ct.

Since $W(y)$ and $h(y)$ depend continuously on y the sequence of right-most eigenvalues $\{\lambda_{m_{j'}}\}$ converges to some λ^*, and x^* satisfies

$$(A^TA - f(x^*)I)x^* - \lambda^* L^T Lx^* = A^T b, \quad \|Lx^*\|^2 = \delta^2,$$

where λ^* is the maximal Lagrange multiplier.

Hence x^* is a global minimizer of

$$g(x; x^*) = \text{min!} \quad \text{subject to } \|Lx\|^2 = \delta^2,$$

Proof ct.

Since $W(y)$ and $h(y)$ depend continuously on y the sequence of right-most eigenvalues $\{\lambda_{m_{j'}}\}$ converges to some λ^*, and x^* satisfies

$$(A^TA - f(x^*)I)x^* - \lambda^* L^T L x^* = A^T b, \quad \|Lx^*\|^2 = \delta^2,$$

where λ^* is the maximal Lagrange multiplier.

Hence x^* is a global minimizer of

$$g(x; x^*) = \min! \quad \text{subject to } \|Lx\|^2 = \delta^2,$$

and for $y \in \mathbb{R}^n$ with $\|Ly\|^2 = \delta^2$ it follows that

$$\begin{aligned} 0 &= g(x^*; x^*) \le g(y; x^*) \\ &= \|Ay - b\|^2 - f(x^*)(1 + \|y\|^2) \\ &= (f(y) - f(x^*))(1 + \|y\|^2), \quad \text{i.e.} \quad f(y) \ge f(x^*). \end{aligned}$$

Quadratic eigenproblem

The quadratic eigenproblems

$$T_m(\lambda)z = (W_m + \lambda I)^2 z - \frac{1}{\delta^2} h_m h_m^T z = 0$$

can be solved by

- linearization
- Krylov subspace method for QEP (Li & Ye 2003)
- SOAR (Bai & Su 2005)
- nonlinear Arnoldi method (Meerbergen 2001, V. 2004)

Krylov subspace method: Li & Ye 2003

For $A, B \in \mathbb{R}^{n\times n}$ such that some linear combination of A and B is a matrix of rank q with an Arnoldi-type process a matrix $Q \in \mathbb{R}^{n\times m+q+1}$ with orthonormal columns and two matrices $H_a \in \mathbb{R}^{m+q+1\times m}$ and $H_b \in \mathbb{R}^{m+q+1\times m}$ with lower bandwidth $q+1$ are determined such that

$$AQ(:,1:m) = Q(:,1:m+q+1)H_a \text{ and } BQ(:,1:m) = Q(:,1:m+q+1)H_b.$$

Krylov subspace method: Li & Ye 2003

For $A, B \in \mathbb{R}^{n\times n}$ such that some linear combination of A and B is a matrix of rank q with an Arnoldi-type process a matrix $Q \in \mathbb{R}^{n\times m+q+1}$ with orthonormal columns and two matrices $H_a \in \mathbb{R}^{m+q+1\times m}$ and $H_b \in \mathbb{R}^{m+q+1\times m}$ with lower bandwidth $q+1$ are determined such that

$$AQ(:,1:m) = Q(:,1:m+q+1)H_a \text{ and } BQ(:,1:m) = Q(:,1:m+q+1)H_b.$$

Then approximations to eigenpairs of the quadratic eigenproblem

$$(\lambda^2 I - \lambda A - B)y = 0$$

are obtained from its projection onto span$Q(:,1:m)$ which reads

$$(\lambda^2 I_m - \lambda H_a(1:m,1:m) - H_b(1:m,1:m))z = 0.$$

Krylov subspace method: Li & Ye 2003

For $A, B \in \mathbb{R}^{n\times n}$ such that some linear combination of A and B is a matrix of rank q with an Arnoldi-type process a matrix $Q \in \mathbb{R}^{n\times m+q+1}$ with orthonormal columns and two matrices $H_a \in \mathbb{R}^{m+q+1\times m}$ and $H_b \in \mathbb{R}^{m+q+1\times m}$ with lower bandwidth $q+1$ are determined such that

$$AQ(:,1:m) = Q(:,1:m+q+1)H_a \text{ and } BQ(:,1:m) = Q(:,1:m+q+1)H_b.$$

Then approximations to eigenpairs of the quadratic eigenproblem

$$(\lambda^2 I - \lambda A - B)y = 0$$

are obtained from its projection onto span$Q(:,1:m)$ which reads

$$(\lambda^2 I_m - \lambda H_a(1:m,1:m) - H_b(1:m,1:m))z = 0.$$

For the quadratic eigenproblem $((W+\lambda I)^2 - \delta^{-2}hh^T)x = 0$ the algorithm of Li & Ye is applied with $A = -W$ and $B = hh^T$ from which the projected problem

$$(\theta^2 I - 2\theta H_a(1:m,1:m) - H_a(1:m+2,1:m)^T H_a(1:m+2,m) - \delta^{-2}H_b(1:m,1:m))z = 0$$

is easily obtained.

SOAR: Bai & Su 2005

The Second Order Arnoldi Reduction method is based on the observation that the Krylov space of the linearization

$$\begin{pmatrix} A & B \\ I & O \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} = \lambda \begin{pmatrix} I & O \\ O & I \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix}$$

of $(\lambda^2 I - \lambda A - B)y = 0$ with initial vector $\begin{pmatrix} r_0 \\ 0 \end{pmatrix}$ has the form

$$\mathcal{K}_k = \left\{ \begin{pmatrix} r_0 \\ 0 \end{pmatrix}, \begin{pmatrix} r_1 \\ r_0 \end{pmatrix}, \begin{pmatrix} r_2 \\ r_1 \end{pmatrix}, \dots, \begin{pmatrix} r_{k-1} \\ r_{k-2} \end{pmatrix} \right\}$$

where

$$\begin{aligned} r_1 &= Ar_0 \\ r_j &= Ar_{j-1} + Br_{j-2}, \text{ for } j \geq 2. \end{aligned}$$

SOAR: Bai & Su 2005

The Second Order Arnoldi Reduction method is based on the observation that the Krylov space of the linearization

$$\begin{pmatrix} A & B \\ I & O \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} = \lambda \begin{pmatrix} I & O \\ O & I \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix}$$

of $(\lambda^2 I - \lambda A - B)y = 0$ with initial vector $\begin{pmatrix} r_0 \\ 0 \end{pmatrix}$ has the form

$$\mathcal{K}_k = \left\{ \begin{pmatrix} r_0 \\ 0 \end{pmatrix}, \begin{pmatrix} r_1 \\ r_0 \end{pmatrix}, \begin{pmatrix} r_2 \\ r_1 \end{pmatrix}, \dots, \begin{pmatrix} r_{k-1} \\ r_{k-2} \end{pmatrix} \right\}$$

where

$$\begin{aligned} r_1 &= Ar_0 \\ r_j &= Ar_{j-1} + Br_{j-2}, \text{ for } j \geq 2. \end{aligned}$$

The entire information on $\mathcal{K}_k$ is therefore contained in the Second Order Krylov Space

$$\mathcal{G}_k(A, B) = \text{span}\{r_0, r_1, \dots, r_{k-1}\}.$$

SOAR determines an orthonormal basis of $\mathcal{G}_k(A, B)$.

Nonlinear Arnoldi Method

start with initial basis V, $V^T V = I$
determine preconditioner $M \approx T(\sigma)^{-1}$, σ close to wanted eigenvalue
find largest eigenvalue μ of $V^T T(\mu) Vy = 0$ and corresponding eigenvector y
set $u = Vy$, $r = T(\mu)u$
while $\|r\|/\|u\| > \epsilon$ **do**
    $v = Mr$
    $v = v - VV^T v$
    $\tilde{v} = v/\|v\|$, $V = [V, \tilde{v}]$
    find largest eigenvalue μ of $V^T T(\mu) Vy = 0$ and corresponding eigenvector y
    set $u = Vy$, $r = T(\mu)u$
end while

Reuse of information

The convergence of W_m and h_m suggests to reuse information from the previous iterations when solving $T_m(\lambda)z = 0$.

Reuse of information

The convergence of W_m and h_m suggests to reuse information from the previous iterations when solving $T_m(\lambda)z = 0$.

Krylov subspace methods for $T_m(\lambda)z = 0$ can be started with the solution z^{m-1} of $T_{m-1}(\lambda)z = 0$.

Reuse of information

The convergence of W_m and h_m suggests to reuse information from the previous iterations when solving $T_m(\lambda)z = 0$.

Krylov subspace methods for $T_m(\lambda)z = 0$ can be started with the solution z^{m-1} of $T_{m-1}(\lambda)z = 0$.

The nonlinear Arnoldi method can use thick starts, i.e. the projection method for

$$T_m(\lambda)z = 0$$

can be initialized by V_{m-1} where $z^{m-1} = V_{m-1}u^{j-1}$, and u^{j-1} is an eigenvector of $V_{m-1}^T T_{m-1}(\lambda)V_{m-1}u = 0$.

Thick and early updates

$$\begin{aligned} W_j &= C - f(x^j)D - S(T - f(x^j)I_{n-k})^{-1}S^T \\ h_j &= g - D(T - f(x^j)I_{n-k})^{-1}c \end{aligned}$$

with $C, D \in \mathbb{R}^{k\times k}$, $S \in \mathbb{R}^{k\times n-k}$, $T \in \mathbb{R}^{n-k\times n-k}$, $g \in \mathbb{R}^k$, $c \in \mathbb{R}^{n-k}$.

Thick and early updates

$$
\begin{aligned}
W_j &= C - f(x^j)D - S(T - f(x^j)I_{n-k})^{-1}S^T \\
h_j &= g - D(T - f(x^j)I_{n-k})^{-1}c
\end{aligned}
$$

with $C, D \in \mathbb{R}^{k\times k}$, $S \in \mathbb{R}^{k\times n-k}$, $T \in \mathbb{R}^{n-k\times n-k}$, $g \in \mathbb{R}^k$, $c \in \mathbb{R}^{n-k}$.

Hence, in order to update the projected problem

$$
V^T(W_{m+1} + \lambda I)^2 Vu - \frac{1}{\delta^2} V^T h_m h_m^T Vu = 0
$$

one has to keep only CV, DV, $S^T V$, and $g^T V$.

Thick and early updates

$$W_j = C - f(x^j)D - S(T - f(x^j)I_{n-k})^{-1}S^T$$
$$h_j = g - D(T - f(x^j)I_{n-k})^{-1}c$$

with $C, D \in \mathbb{R}^{k\times k}$, $S \in \mathbb{R}^{k\times n-k}$, $T \in \mathbb{R}^{n-k\times n-k}$, $g \in \mathbb{R}^k$, $c \in \mathbb{R}^{n-k}$.

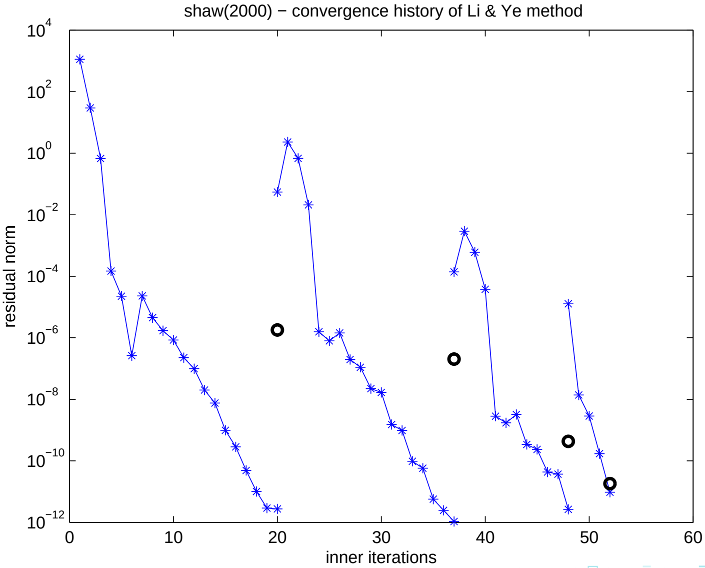
Hence, in order to update the projected problem

$$V^T(W_{m+1} + \lambda I)^2 Vu - \frac{1}{\delta^2} V^T h_m h_m^T Vu = 0$$

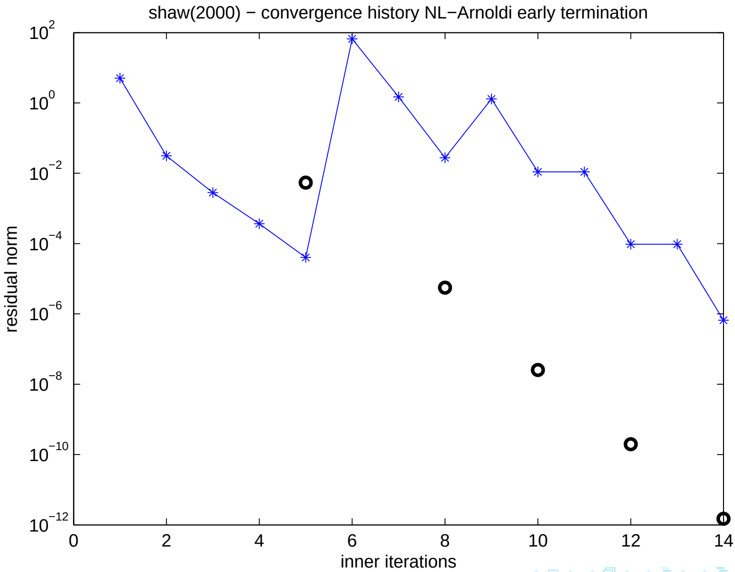
one has to keep only CV, DV, $S^T V$, and $g^T V$.

Since it is inexpensive to obtain updates of W_m and h_m we decided to terminate the inner iteration long before convergence, namely if the residual of the quadratic eigenvalue was reduced by at least 10^{-2}. This reduced the computing time further.

Example: shaw(2000); Li & Ye



shaw(2000); Early updates



Necessary Condition: Golub, Hansen, O'Leary 1999

The RTLS solution x satisfies

$$(A^TA + \lambda_I I + \lambda_L L^T L)x = A^T b, \quad \|Lx\|^2 = \delta^2$$

where

$$\begin{aligned} \lambda_I &= -f(x) \\ \lambda_L &= -\frac{1}{\delta^2}(b^T(Ax - b) - \lambda_I) \end{aligned}$$

Necessary Condition: Golub, Hansen, O'Leary 1999

The RTLS solution x satisfies

$$(A^TA + \lambda_I I + \lambda_L L^T L)x = A^T b, \quad \|Lx\|^2 = \delta^2$$

where

$$\begin{aligned} \lambda_I &= -f(x) \\ \lambda_L &= -\frac{1}{\delta^2}(b^T(Ax - b) - \lambda_I) \end{aligned}$$

Eliminating λ_I:

$$(A^TA - f(x)I + \lambda_L L^T L)x = A^T b, \quad \|Lx\|^2 = \delta^2,$$

iterating on $f(x^m)$, and solving via quadratic eigenproblems yields the method of Sima, Van Huffel and Golub.

Necessary Condition: Renaut, Guo 2002,2005

The RTLS solution x satisfies the eigenproblem

$$B(x)\begin{pmatrix} x \\ -1 \end{pmatrix} = -\lambda_I \begin{pmatrix} x \\ -1 \end{pmatrix}$$

where

$$\begin{aligned} B(x) &= [A,b]^T[A,b] + \lambda_L(x)\begin{pmatrix} L^TL & 0 \\ 0 & -\delta^2 \end{pmatrix} \\ \lambda_I &= -f(x) \\ \lambda_L &= -\frac{1}{\delta^2}(b^T(Ax-b)-\lambda_I) \end{aligned}$$

Necessary Condition: Renaut, Guo 2002,2005

The RTLS solution x satisfies the eigenproblem

$$B(x)\begin{pmatrix} x \\ -1 \end{pmatrix} = -\lambda_I \begin{pmatrix} x \\ -1 \end{pmatrix}$$

where

$$\begin{aligned} B(x) &= [A,b]^T[A,b] + \lambda_L(x)\begin{pmatrix} L^TL & 0 \\ 0 & -\delta^2 \end{pmatrix} \\ \lambda_I &= -f(x) \\ \lambda_L &= -\frac{1}{\delta^2}(b^T(Ax-b)-\lambda_I) \end{aligned}$$

Conversely, if $\left(-\hat{\lambda}, \begin{pmatrix} \hat{x} \\ -1 \end{pmatrix}\right)$ is an eigenpair of $B(\hat{x})$, and $\lambda_L(\hat{x}) = -\frac{1}{\delta^2}(b^T(A\hat{x}-b)+f(\hat{x}))$, then $\hat{x}$ satisfies the necessary conditions of the last slide, and the eigenvalue is given by $\hat{\lambda} = -f(\hat{x})$.

Method of Renaut, Guo 2005

For $\theta \in \mathbb{R}_+$ let $(x_\theta^T, -1)^T$ be the eigenvector corresponding to the smallest eigenvalue of

$$B(\theta) := [A, b]^T [A, b] + \theta \begin{pmatrix} L^T L & 0 \\ 0 & -\delta^2 \end{pmatrix} \qquad (*)$$

Method of Renaut, Guo 2005

For $\theta \in \mathbb{R}_+$ let $(x_\theta^T, -1)^T$ be the eigenvector corresponding to the smallest eigenvalue of

$$B(\theta) := [A, b]^T [A, b] + \theta \begin{pmatrix} L^T L & 0 \\ 0 & -\delta^2 \end{pmatrix} \qquad (*)$$

Problem Determine θ such that $g(\theta) = 0$ where

$$g(\theta) := \frac{\|Lx_\theta\|^2 - \delta^2}{1 + \|x_\theta\|^2}.$$

Method of Renaut, Guo 2005

For $\theta \in \mathbb{R}_+$ let $(x_\theta^T, -1)^T$ be the eigenvector corresponding to the smallest eigenvalue of

$$B(\theta) := [A, b]^T [A, b] + \theta \begin{pmatrix} L^T L & 0 \\ 0 & -\delta^2 \end{pmatrix} \qquad (*)$$

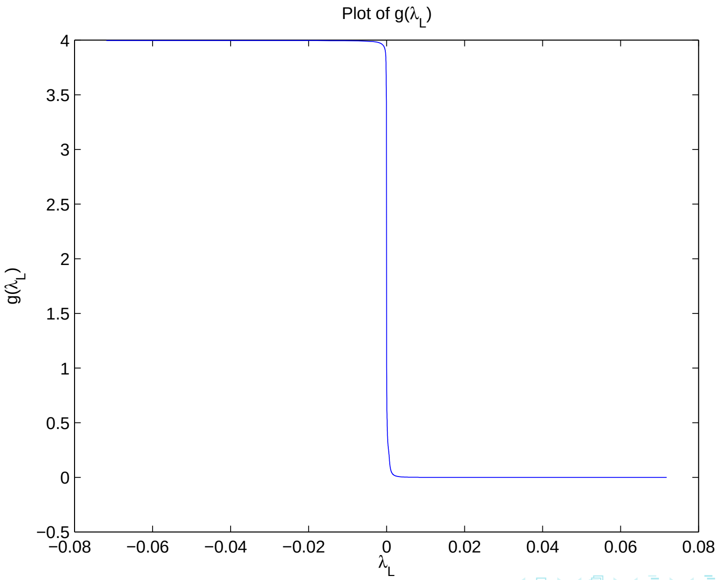
Problem Determine θ such that $g(\theta) = 0$ where

$$g(\theta) := \frac{\|Lx_\theta\|^2 - \delta^2}{1 + \|x_\theta\|^2}.$$

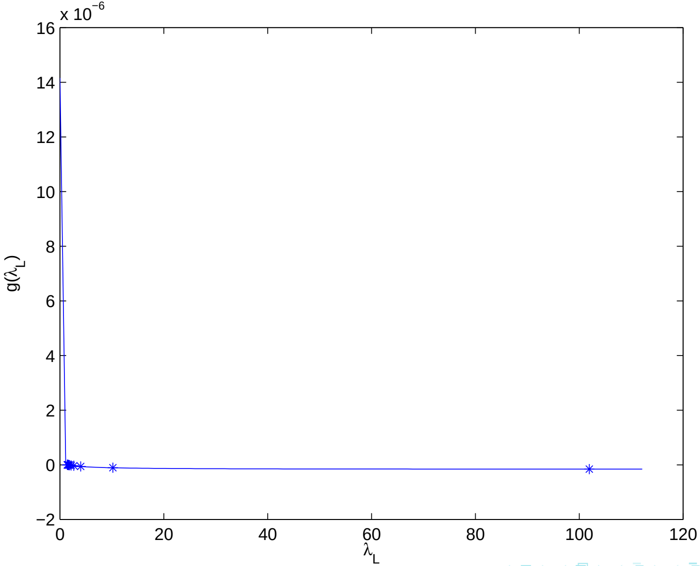
Given θ_k, the eigenvalue problem (*) is solved by the Rayleigh quotient iteration, and θ_k is updated according to

$$\theta_{k+1} = \theta_k + \frac{\theta_k}{\delta^2} g(\theta_k).$$

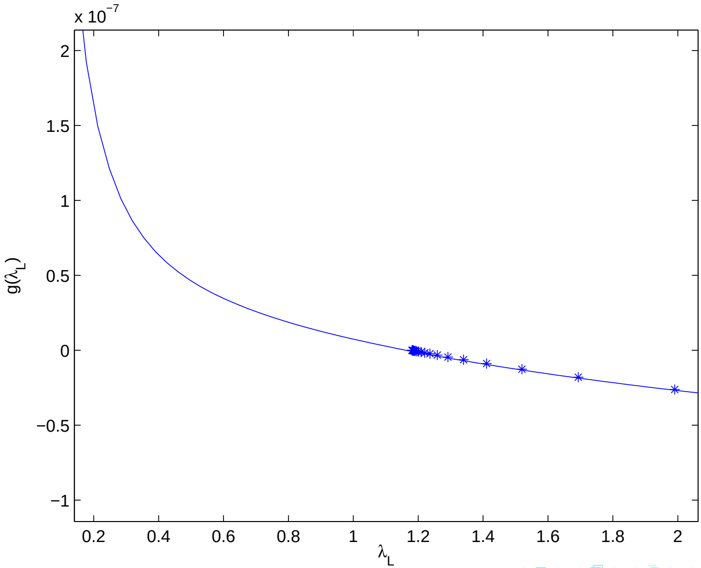
Typical g


Plot of g(λ_L)
4
3.5
3
2.5
2
1.5
1
0.5
0
−0.5
−0.08
−0.06
−0.04
−0.02
0
0.02
0.04
0.06
0.08
λ_L
g(λ_L)

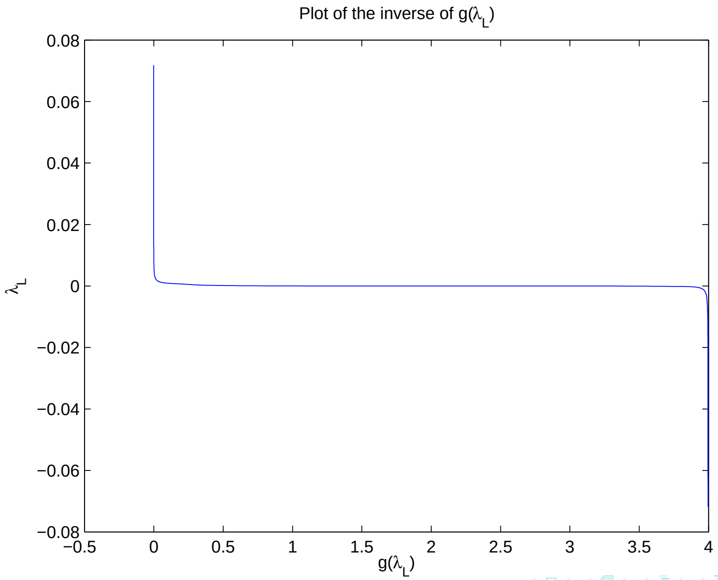
Back tracking



Close up



Typical g^{-1}



Plot of the inverse of g(λL)
0.08
0.06
0.04
0.02
0
−0.02
−0.04
−0.06
−0.08
λL
−0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
g(λL)

Rational interpolation

Assumption A: Let $\theta_1 < \theta_2 < \theta_3$ such that $g(\theta_3) < 0 < g(\theta_1)$.

Rational interpolation

Assumption A: Let $\theta_1 < \theta_2 < \theta_3$ such that $g(\theta_3) < 0 < g(\theta_1)$.

Let $p_1 < p_2$ be the poles of g^{-1}, let

$$h := \frac{\alpha + \beta x + \gamma x^2}{(p_2 - x)(x - p_1)}$$

be the interpolation of g^{-1} at $(g(\theta_j), \theta_j)$, $j = 1, 2, 3$, and set $\theta_4 = h(0)$.

Rational interpolation

Assumption A: Let $\theta_1 < \theta_2 < \theta_3$ such that $g(\theta_3) < 0 < g(\theta_1)$.

Let $p_1 < p_2$ be the poles of g^{-1}, let

$$h := \frac{\alpha + \beta x + \gamma x^2}{(p_2 - x)(x - p_1)}$$

be the interpolation of g^{-1} at $(g(\theta_j), \theta_j)$, $j = 1, 2, 3$, and set $\theta_4 = h(0)$.

Drop θ_1 or θ_3 such that the remaining θ values satisfy assumption A, and repeat until convergence.

Rational interpolation

Assumption A: Let $\theta_1 < \theta_2 < \theta_3$ such that $g(\theta_3) < 0 < g(\theta_1)$.

Let $p_1 < p_2$ be the poles of g^{-1}, let

$$h := \frac{\alpha + \beta x + \gamma x^2}{(p_2 - x)(x - p_1)}$$

be the interpolation of g^{-1} at $(g(\theta_j), \theta_j)$, $j = 1, 2, 3$, and set $\theta_4 = h(0)$.

Drop θ_1 or θ_3 such that the remaining θ values satisfy assumption A, and repeat until convergence.

To evaluate $g(\theta)$ one has to solve the eigenvalue problem $B(\theta)x = \mu x$ which can be done efficiently by the nonlinear Arnoldi method starting with the entire search space of the previous step.

Numerical example

We added white noise to the data of phillips(n) and deriv2(n) with noise level 1%, and chose L to be an approximate first derivative.

The following tabel contains the average CPU time for 100 test problems of dimensions $n = 1000$, $n = 2000$, and $n = 4000$ (CPU: Pentium D, 3.4 GHz).

Numerical example

We added white noise to the data of phillips(n) and deriv2(n) with noise level 1%, and chose L to be an approximate first derivative.

The following tabel contains the average CPU time for 100 test problems of dimensions $n = 1000$, $n = 2000$, and $n = 4000$ (CPU: Pentium D, 3.4 GHz).

problem	n	SOAR	Li & Ye	NL Arn.	R & G
phillips	1000	0.11	0.13	0.05	0.07
	2000	0.38	0.39	0.15	0.19
	4000	1.30	1.30	0.66	0.70
deriv2	1000	0.12	0.09	0.06	0.05
	2000	0.31	0.28	0.18	0.17
	4000	1.25	1.06	0.69	0.38

Numerical example ct.

We added white noise to the data of phillips(n) and deriv2(n) with noise level 10%, and chose L to be an approximate first derivative.

The following tabel contains the average CPU time for 100 test problems of dimensions $n = 1000$, $n = 2000$, and $n = 4000$ (CPU: Pentium D, 3.4 GHz).

problem	n	SOAR	Li & Ye	NL Arn.	R & G
phillips	1000	0.11	0.14	0.05	0.07
	2000	0.45	0.44	0.16	0.23
	4000	1.16	1.39	0.64	0.74
deriv2	1000	0.10	0.08	0.05	0.05
	2000	0.27	0.24	0.16	0.18
	4000	0.99	0.93	0.64	0.43

Conclusions

- Using minmax theory for nonlinear eigenproblems we proved a localization and characterization result for the maximum real eigenvalue of a quadratic eigenproblem occuring in TLS problems.

Conclusions

- Using minmax theory for nonlinear eigenproblems we proved a localization and characterization result for the maximum real eigenvalue of a quadratic eigenproblem occuring in TLS problems.
- The right most eigenvalue is not necesarily nonnegative

Conclusions

- Using minmax theory for nonlinear eigenproblems we proved a localization and characterization result for the maximum real eigenvalue of a quadratic eigenproblem occuring in TLS problems.
- The right most eigenvalue is not necesarily nonnegative
- The maximum eigenvalue can be determined efficiently by the nonlinear Arnoldi method, and taking advantage of thick starts and early updates it can be accelerated substantially.

Conclusions

- Using minmax theory for nonlinear eigenproblems we proved a localization and characterization result for the maximum real eigenvalue of a quadratic eigenproblem occuring in TLS problems.
- The right most eigenvalue is not necesarily nonnegative
- The maximum eigenvalue can be determined efficiently by the nonlinear Arnoldi method, and taking advantage of thick starts and early updates it can be accelerated substantially.
- Likewise, The approach of Renaut & Guo can be significantly improved by rational interpolation and the nonlinear Arnoldi method with thick starts.

Conclusions

- Using minmax theory for nonlinear eigenproblems we proved a localization and characterization result for the maximum real eigenvalue of a quadratic eigenproblem occuring in TLS problems.
- The right most eigenvalue is not necesarily nonnegative
- The maximum eigenvalue can be determined efficiently by the nonlinear Arnoldi method, and taking advantage of thick starts and early updates it can be accelerated substantially.
- Likewise, The approach of Renaut & Guo can be significantly improved by rational interpolation and the nonlinear Arnoldi method with thick starts.

THANK YOU FOR YOUR ATTENTION