Secular Equations - Applications and Methods

Nick Gould
STFC Rutherford Appleton Laboratory

with contributions from

Coralia Cartis (Oxford), **Chris Farmer** (Oxford)
Jari Fowkes (Edinburgh), **Stefano Lucidi** (Rome)
Daniel Robinson (Johns Hopkins), **Massimo Roma** (Rome)
Sue Thorne (RAL), **Philippe Toint** (Namur)
. . . and **Gene Golub**

$$(A + \lambda B)x(\lambda) = b, \quad \theta(x(\lambda)) = \tau(\lambda)$$

SIAM UKIE Annual Meeting 2014
University College London, 9th January 2014

Summary of the talk

- motivation and examples
- solving the explicit problem
- solving the implicit problem by factorization
- solving the implicit problem by iteration
- comments and conclusions

History

- **Secular** (adj.) 3. astronomy, of or denoting slow changes in the motion of the sun or planets (c.f. 1. not connected with religious or spiritual matters) (O.E.D.)
- early use of the term **secular equation** in papers by Cauchy and Sylvester refer to equations from which the eigenvalues of a real symmetric matrix may be obtained
- Hilbert (1924) and E.T. Browne (1930) explicitly refer to the characteristic equation as a secular equation
- Golub (SIAM Review, 1973) gives a number of examples of more general secular equations (to follow). See also Golub & Meurant (Matrices, Moments and Quadrature with Applications, 2010)

Secular equations

- given real symmetric $\boldsymbol{A}$ and $\boldsymbol{B}$ and real vector $\boldsymbol{b}$
- let $\boldsymbol{x}(\boldsymbol{\lambda})$ be a solution of the linear system

$$(\boldsymbol{A} + \boldsymbol{\lambda} \boldsymbol{B})\boldsymbol{x}(\boldsymbol{\lambda}) = \boldsymbol{b}$$

  for given real scalar $\boldsymbol{\lambda}$

Secular equations

- given real symmetric $\boldsymbol{A}$ and $\boldsymbol{B}$ and real vector $\boldsymbol{b}$
- let $\boldsymbol{x}(\boldsymbol{\lambda})$ be a solution of the linear system

$$(\boldsymbol{A} + \boldsymbol{\lambda} \boldsymbol{B})\boldsymbol{x}(\boldsymbol{\lambda}) = \boldsymbol{b}$$

  for given real scalar $\boldsymbol{\lambda}$
- find values of $\boldsymbol{\lambda}$ that solve the **secular equation**

$$\boldsymbol{\theta}(\boldsymbol{x}(\boldsymbol{\lambda})) = \boldsymbol{\tau}(\boldsymbol{\lambda})$$

  for given real **secular functions** $\boldsymbol{\theta} : \mathbb{R}^n \rightarrow \mathbb{R}$ and $\boldsymbol{\tau} : \mathbb{R} \rightarrow \mathbb{R}$

Secular equations

- given real symmetric A and B and real vector b
- let $x(\lambda)$ be a solution of the linear system

$$(A + \lambda B)x(\lambda) = b$$

for given real scalar λ

- find values of λ that solve the **secular equation**

$$\theta(x(\lambda)) = \tau(\lambda)$$

for given real **secular functions** $\theta : \mathbb{R}^n \to \mathbb{R}$ and $\tau : \mathbb{R} \to \mathbb{R}$

- often impose extra conditions that identify particular λ

Examples

- eigenvalues of a symmetric bordered matrix
- eigenvalues following a symmetric rank-one change
- the trust-region subproblem
- regularization of quadratic minimization
- global optimization
- total least squares

Eigenvalues of a symmetric bordered matrix

- given real symmetric $\boldsymbol{A}$, vector $\boldsymbol{b}$ and scalar $\boldsymbol{\beta}$, find the eigenvalues of

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{b} \\ \boldsymbol{b}^T & \boldsymbol{\beta} \end{pmatrix}$$

- require eigenvalues $\boldsymbol{\lambda}$ for which

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{b} \\ \boldsymbol{b}^T & \boldsymbol{\beta} \end{pmatrix} \begin{pmatrix} \boldsymbol{v} \\ \boldsymbol{\xi} \end{pmatrix} = \boldsymbol{\lambda} \begin{pmatrix} \boldsymbol{v} \\ \boldsymbol{\xi} \end{pmatrix}$$

Eigenvalues of a symmetric bordered matrix

- given real symmetric $\boldsymbol{A}$, vector $\boldsymbol{b}$ and scalar $\boldsymbol{\beta}$, find the eigenvalues of

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{b} \\ \boldsymbol{b}^T & \boldsymbol{\beta} \end{pmatrix}$$

- require eigenvalues $\boldsymbol{\lambda}$ for which

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{b} \\ \boldsymbol{b}^T & \boldsymbol{\beta} \end{pmatrix} \begin{pmatrix} \boldsymbol{v} \\ \boldsymbol{\xi} \end{pmatrix} = \boldsymbol{\lambda} \begin{pmatrix} \boldsymbol{v} \\ \boldsymbol{\xi} \end{pmatrix}$$

- if $\boldsymbol{\xi} = \boldsymbol{0} \Longrightarrow \boldsymbol{A}\boldsymbol{v} = \boldsymbol{\lambda}\boldsymbol{v}$
- otherwise $\boldsymbol{x} = -\boldsymbol{v}/\boldsymbol{\xi} \Longrightarrow$

$$(\boldsymbol{A} - \boldsymbol{\lambda}\boldsymbol{I})\boldsymbol{x}(\boldsymbol{\lambda}) = \boldsymbol{b} \quad \text{and} \quad \boldsymbol{b}^T\boldsymbol{x}(\boldsymbol{\lambda}) = \boldsymbol{\beta} - \boldsymbol{\lambda}$$

Eigenvalues of a symmetric bordered matrix

- given real symmetric A, vector b and scalar β, find the eigenvalues of

$$\begin{pmatrix} A & b \\ b^T & \beta \end{pmatrix}$$

- require eigenvalues λ for which

$$\begin{pmatrix} A & b \\ b^T & \beta \end{pmatrix} \begin{pmatrix} v \\ \xi \end{pmatrix} = \lambda \begin{pmatrix} v \\ \xi \end{pmatrix}$$

- if $\xi = 0 \Longrightarrow \; Av = \lambda v$
- otherwise $x = -v/\xi \Longrightarrow$

$$(A - \lambda I)x(\lambda) = b \;\; \text{and} \;\; b^T x(\lambda) = \beta - \lambda$$

- secular functions

$$\theta(x) = b^T x \;\; \text{and} \;\; \tau(\lambda) = \beta - \lambda$$

Eigenvalues following a symmetric rank-one change

- given real symmetric $\boldsymbol{A}$, vector $\boldsymbol{b}$ and scalar $\boldsymbol{\beta}$, find the eigenvalues of

$$\boldsymbol{A} + \boldsymbol{\beta} \boldsymbol{b} \boldsymbol{b}^T$$

- prime example: Divide & Conquer method for tridiagonal eigenproblem

  (Cuppen, 1981)

- require eigenvalues $\boldsymbol{\lambda}$ for which

$$(\boldsymbol{A} + \boldsymbol{\beta} \boldsymbol{b} \boldsymbol{b}^T)\boldsymbol{v} = \boldsymbol{\lambda} \boldsymbol{v}$$

Eigenvalues following a symmetric rank-one change

- given real symmetric $\boldsymbol{A}$, vector $\boldsymbol{b}$ and scalar $\boldsymbol{\beta}$, find the eigenvalues of
$$\boldsymbol{A} + \boldsymbol{\beta} \boldsymbol{b} \boldsymbol{b}^T$$
- prime example: Divide & Conquer method for tridiagonal eigenproblem

(Cuppen, 1981)

- require eigenvalues $\boldsymbol{\lambda}$ for which
$$(\boldsymbol{A} + \boldsymbol{\beta} \boldsymbol{b} \boldsymbol{b}^T)\boldsymbol{v} = \boldsymbol{\lambda} \boldsymbol{v}$$
- if $\boldsymbol{b}^T \boldsymbol{v} = 0 \Longrightarrow \boldsymbol{A}\boldsymbol{v} = \boldsymbol{\lambda}\boldsymbol{v}$,
- otherwise with $\boldsymbol{x} = -\boldsymbol{v}/\boldsymbol{\beta} \boldsymbol{b}^T \boldsymbol{v} \Longrightarrow$

$$(\boldsymbol{A} - \boldsymbol{\lambda} \boldsymbol{I})\boldsymbol{x}(\boldsymbol{\lambda}) = \boldsymbol{b} \quad \text{and} \quad \boldsymbol{b}^T \boldsymbol{x}(\boldsymbol{\lambda}) = -1/\boldsymbol{\beta}$$

Eigenvalues following a symmetric rank-one change

- given real symmetric $\boldsymbol{A}$, vector $\boldsymbol{b}$ and scalar β, find the eigenvalues of

$$\boldsymbol{A} + \beta \boldsymbol{b}\boldsymbol{b}^T$$

- prime example: Divide & Conquer method for tridiagonal eigenproblem

(Cuppen, 1981)

- require eigenvalues $\boldsymbol{\lambda}$ for which

$$(\boldsymbol{A} + \beta \boldsymbol{b}\boldsymbol{b}^T)\boldsymbol{v} = \boldsymbol{\lambda}\boldsymbol{v}$$

- if $\boldsymbol{b}^T\boldsymbol{v} = 0 \Longrightarrow \boldsymbol{A}\boldsymbol{v} = \boldsymbol{\lambda}\boldsymbol{v}$,
- otherwise with $\boldsymbol{x} = -\boldsymbol{v}/\beta\boldsymbol{b}^T\boldsymbol{v} \Longrightarrow$

$$(\boldsymbol{A} - \boldsymbol{\lambda}\boldsymbol{I})\boldsymbol{x}(\boldsymbol{\lambda}) = \boldsymbol{b} \quad \text{and} \quad \boldsymbol{b}^T\boldsymbol{x}(\boldsymbol{\lambda}) = -1/\beta$$

- secular functions

$$\theta(\boldsymbol{x}) = \boldsymbol{b}^T\boldsymbol{x} \quad \text{and} \quad \tau(\boldsymbol{\lambda}) = -1/\beta$$

The trust-region subproblem

- given real symmetric $\boldsymbol{H}$, vector $\boldsymbol{g}$ & radius $\boldsymbol{\Delta > 0}$

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad q(x) = \tfrac{1}{2} x^T H x + g^T x \quad \text{subject to} \quad \|x\|_2 \leq \Delta$$

- fundamental subproblem in trust-region-based optimization

The trust-region subproblem

- given real symmetric $\boldsymbol{H}$, vector $\boldsymbol{g}$ & radius $\boldsymbol{\Delta > 0}$

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad q(x) = \tfrac{1}{2}x^T H x + g^T x \quad \text{subject to} \quad \|x\|_2 \leq \Delta$$

- fundamental subproblem in trust-region-based optimization
- critical point of $q(x)$ subject to $\frac{1}{2}\|x\|_2^2 \leq \frac{1}{2}\Delta^2 \Longrightarrow$

$$(H + \lambda I)x(\lambda) = -g \qquad\qquad (*)$$

The trust-region subproblem

- given real symmetric H, vector g & radius $\Delta > 0$

$$\operatorname*{minimize}_{x \in \mathbb{R}^n} \; q(x) = \tfrac{1}{2} x^T H x + g^T x \;\; \text{subject to} \;\; \|x\|_2 \le \Delta$$

- fundamental subproblem in trust-region-based optimization
- critical point of $q(x)$ subject to $\frac{1}{2}\|x\|_2^2 \le \frac{1}{2}\Delta^2 \Longrightarrow$

$$(H + \lambda I) x(\lambda) = -g \qquad (*)$$

Theorem: any **global** minimizer $x_* = x(\lambda_*)$ of $q(x)$ subject to $\|x\|_2 \le \Delta$ satisfies $(*)$ where $H + \lambda_* I \succeq 0$, $\lambda_* \ge 0$ and $\lambda_*(\|x_*\|_2 - \Delta) = 0$. If $H + \lambda_* I \succ 0$, x_* is unique.

(Gay, Moré, Sorensen, 1981-3)

The trust-region subproblem

- given real symmetric H, vector g & radius $\Delta > 0$

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad q(x) = \tfrac{1}{2} x^T H x + g^T x \ \text{ subject to } \ \|x\|_2 \leq \Delta$$

- fundamental subproblem in trust-region-based optimization
- critical point of $q(x)$ subject to $\frac{1}{2}\|x\|_2^2 \leq \frac{1}{2}\Delta^2 \Longrightarrow$

$$(H + \lambda I) x(\lambda) = -g \qquad\qquad (*)$$

Theorem: any **global** minimizer $x_* = x(\lambda_*)$ of $q(x)$ subject to $\|x\|_2 \leq \Delta$ satisfies $(*)$ where $H + \lambda_* I \succeq 0$, $\lambda_* \geq 0$ and $\lambda_*(\|x_*\|_2 - \Delta) = 0$. If $H + \lambda_* I \succ 0$, x_* is unique.

(Gay, Moré, Sorensen, 1981-3)

- if $\|x_*\|_2 = \Delta$, need to find largest root of $(*)$ with secular functions

$$\theta(x) = \|x\|_2 \ \text{ and } \ \tau(\lambda) = \Delta$$

Regularization of quadratic minimization

- given real symmetric H, vector g weight $\sigma > 0$ & index $p \geq 2$

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad r(x) = \tfrac{1}{2} x^T H x + g^T x + \tfrac{1}{p} \sigma \|x\|_2^p$$

- fundamental subproblem in optimal-complexity optimization

Science & Technology Facilities Council

Regularization of quadratic minimization

- given real symmetric H, vector g weight $\sigma > 0$ & index $p \geq 2$

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad r(x) = \tfrac{1}{2} x^T H x + g^T x + \tfrac{1}{p} \sigma \|x\|_2^p$$

- fundamental subproblem in optimal-complexity optimization
- critical point of $r(x) \Longrightarrow$

$$H x + g + \sigma \|x\|_2^{p-2} x = 0 \quad \Longrightarrow$$
$$(H + \lambda I) x(\lambda) = -g \quad \text{and} \quad \lambda = \sigma \|x(\lambda)\|_2^{p-2} \qquad (*)$$

Regularization of quadratic minimization

- given real symmetric H, vector g weight $\sigma > 0$ & index $p \geq 2$

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad r(x) = \tfrac{1}{2} x^T H x + g^T x + \tfrac{1}{p} \sigma \|x\|_2^p$$

- fundamental subproblem in optimal-complexity optimization
- critical point of $r(x) \Longrightarrow$

$$Hx + g + \sigma \|x\|_2^{p-2} x = 0 \quad \Longrightarrow$$

$$(H + \lambda I) x(\lambda) = -g \quad \text{and} \quad \lambda = \sigma \|x(\lambda)\|_2^{p-2} \qquad (*)$$

Theorem: Any **global** minimizer $x_* = x(\lambda_*)$ of $r(x)$ satisfies $(*)$ where $H + \lambda_* I \succeq 0$. If $H + \lambda_* I \succ 0$, x_* is unique.

(Cartis, G., Toint, 2011)

Regularization of quadratic minimization

- given real symmetric H, vector g weight $\sigma > 0$ & index $p \geq 2$

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad r(x) = \tfrac{1}{2} x^T H x + g^T x + \tfrac{1}{p} \sigma \|x\|_2^p$$

- fundamental subproblem in optimal-complexity optimization
- critical point of $r(x) \Longrightarrow$

$$H x + g + \sigma \|x\|_2^{p-2} x = 0 \quad \Longrightarrow$$
$$(H + \lambda I) x(\lambda) = -g \quad \text{and} \quad \lambda = \sigma \|x(\lambda)\|_2^{p-2} \qquad (*)$$

Theorem: Any **global** minimizer $x_* = x(\lambda_*)$ of $r(x)$ satisfies $(*)$ where $H + \lambda_* I \succeq 0$. If $H + \lambda_* I \succ 0$, x_* is unique.

(Cartis, G., Toint, 2011)

- need to find largest root of $(*)$ with secular functions

$$\theta(x) = \|x(\lambda)\|_2 \quad \text{and} \quad \tau(\lambda) = (\lambda/\sigma)^{1/(p-2)}$$

Science & Technology Facilities Council

Global Lipschitz optimization

- given a C^2 function $f(x)$ with Lipschitz Hessian, Taylor $\Longrightarrow$

$$c^{\mathrm{L}}(s) \equiv f(x) + s^T g(x) + \tfrac{1}{2} s^T H(x) s - \tfrac{1}{3}\sigma\|s\|_2^3 \leq$$
$$f(x+s) \leq f(x) + s^T g(x) + \tfrac{1}{2} s^T H(x) s + \tfrac{1}{3}\sigma\|s\|_2^3 \equiv c^{\mathrm{U}}(s)$$

  where g/H are the gradient/Hessian of f & σ is the Lipschitz constant

- global mimima of c^{L} and c^{U} over $\mathcal{S} = \{s : \|s\|_2 \leq \Delta\}$ provide lower and upper bounds for global minimum of $f(x+s)$ over $\mathcal{S}$
- partition "space of interest" for f into overlapping hyper-spheres and apply branch and bound to find global minimizer in this space
- global mimima of c^{L} and c^{U} over $\mathcal{S}$ each has associated secular equation $\Longrightarrow$ tractable (Cartis, Farmer, Fowkes, G.,2012)

Simplification of secular equations

- without loss of generality consider

$$(A \pm \lambda I)x(\lambda) = b \ \text{ and } \ \theta(x(\lambda)) = \tau(\lambda)$$

for real, symmetric A, where for simplicity

$$\theta(x) = b^T x \ \text{ or } \ \theta(x) = \|x\|_2$$

Simplification of secular equations

- without loss of generality consider

$$(A \pm \lambda I)x(\lambda) = b \;\text{ and }\; \theta(x(\lambda)) = \tau(\lambda)$$

for real, symmetric A, where for simplicity

$$\theta(x) = b^T x \;\text{ or }\; \theta(x) = \|x\|_2$$

- suppose that eigen-decomposition $A = Q \Lambda Q^T$ is known, with $\Lambda = \text{Diag}(\lambda_i)$: $\lambda_i \leq \lambda_{i+1}$ & orthogonal Q, and that $\bar{b} = Q^T b$ $\Longrightarrow x(\lambda) = Q\bar{x}(\lambda)$:

$$(\Lambda \pm \lambda I)\bar{x}(\lambda) = \bar{b} \;\text{ and }\; \bar{\theta}(\bar{x}(\lambda)) = \tau(\lambda)$$

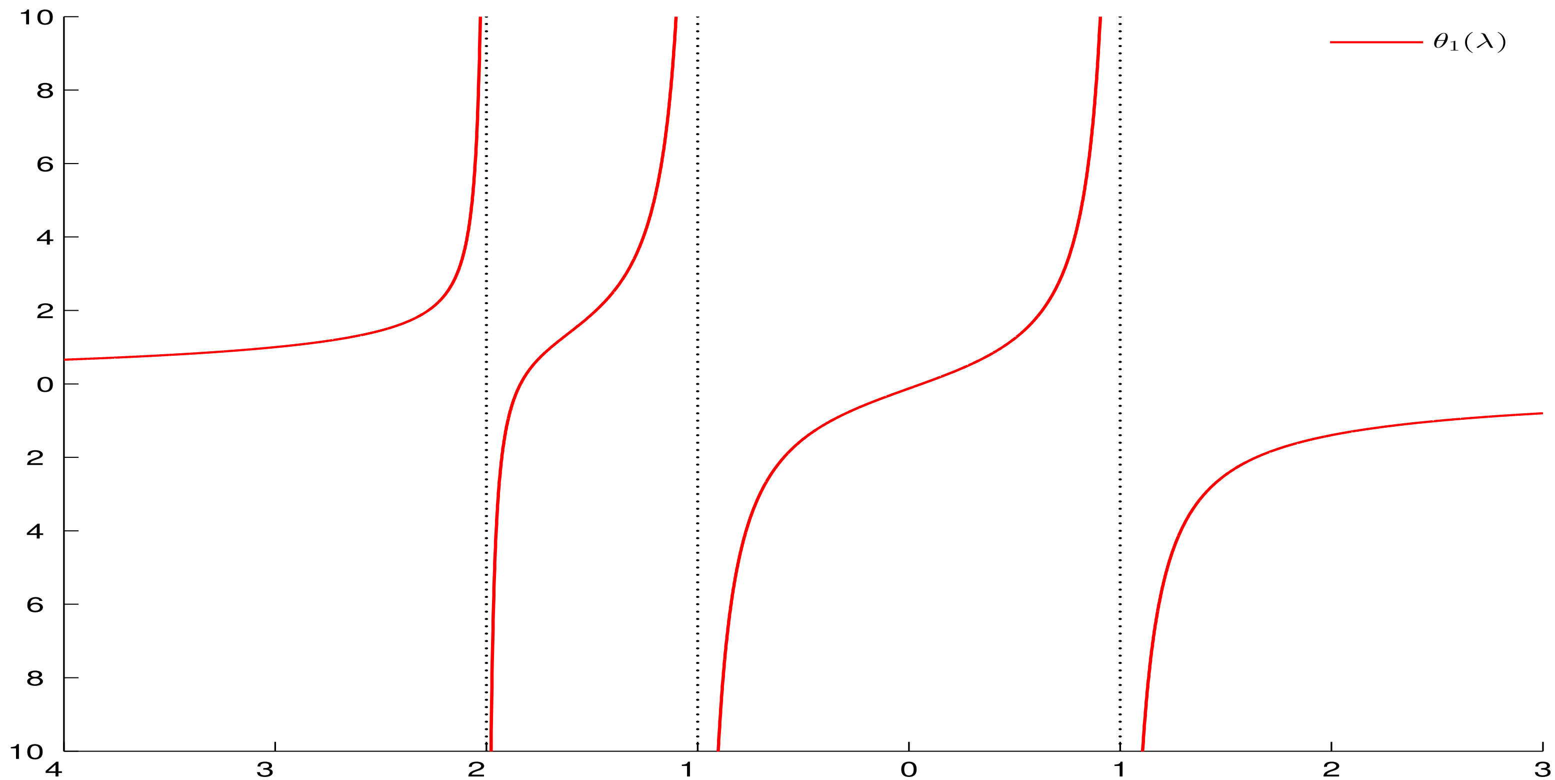
and, e.g.,

$$\bar{\theta}(\bar{x}) = \bar{b}^T \bar{x} \;\text{ or }\; \bar{\theta}(\bar{x}) = \|\bar{x}\|_2$$

Type-I secular equations

$$(\Lambda - \lambda I)\bar{x}(\lambda) = \bar{b} \quad \text{and} \quad \theta_1(\lambda) \equiv \bar{b}^T \bar{x}(\lambda) = \tau(\lambda) \implies$$

$$\theta_1(\lambda) = \sum_{i=1}^{n} \frac{\bar{b}_i^2}{\lambda_i - \lambda}$$

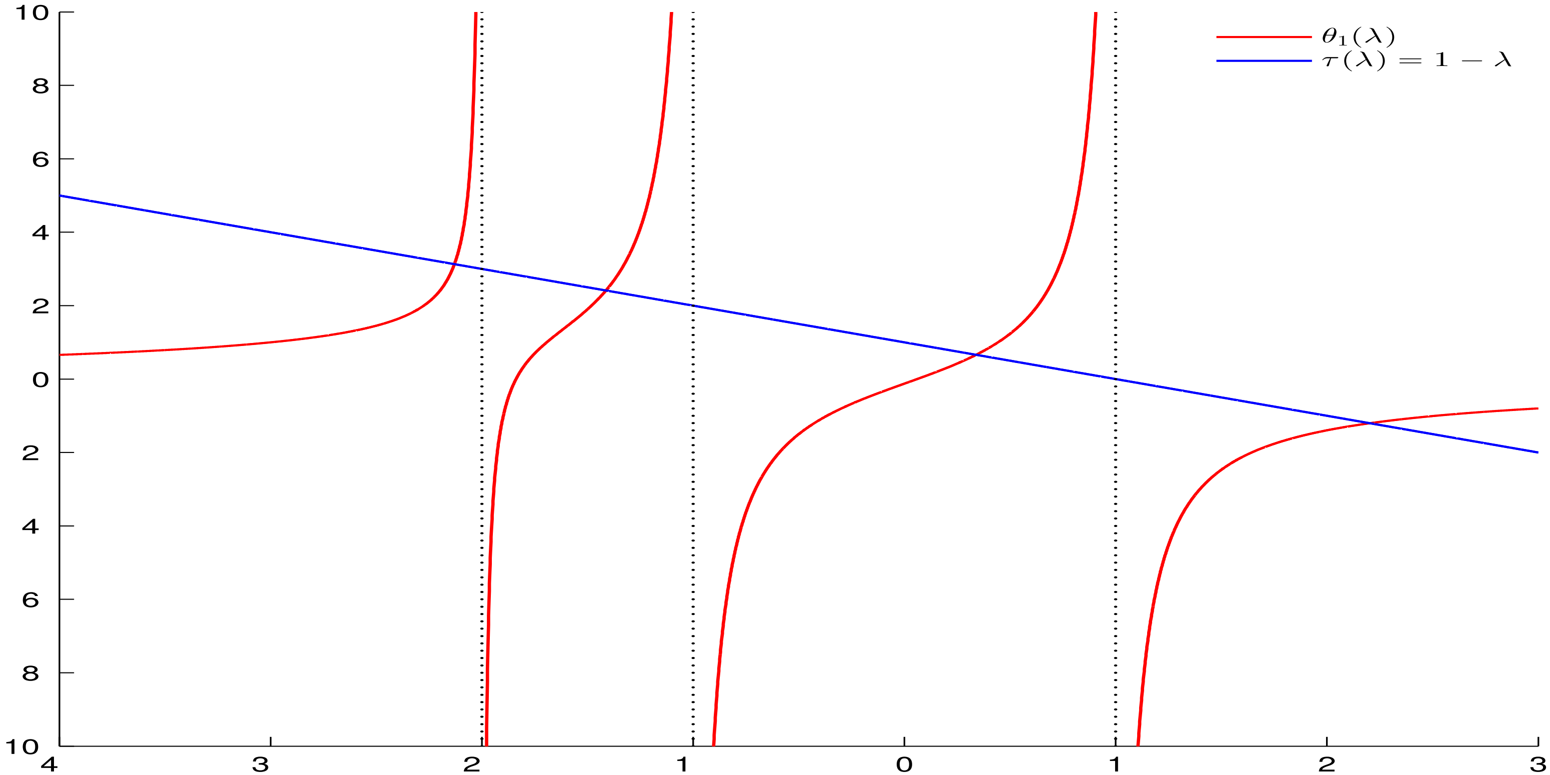


The type-I secular function $\theta_1(\lambda)$

Type-I secular equations

$$(\Lambda - \lambda I)\bar{x}(\lambda) = \bar{b} \quad \text{and} \quad \theta_1(\lambda) \equiv \bar{b}^T \bar{x}(\lambda) = \tau(\lambda) \implies$$

$$\theta_1(\lambda) = \sum_{i=1}^{n} \frac{\bar{b}_i^2}{\lambda_i - \lambda}$$

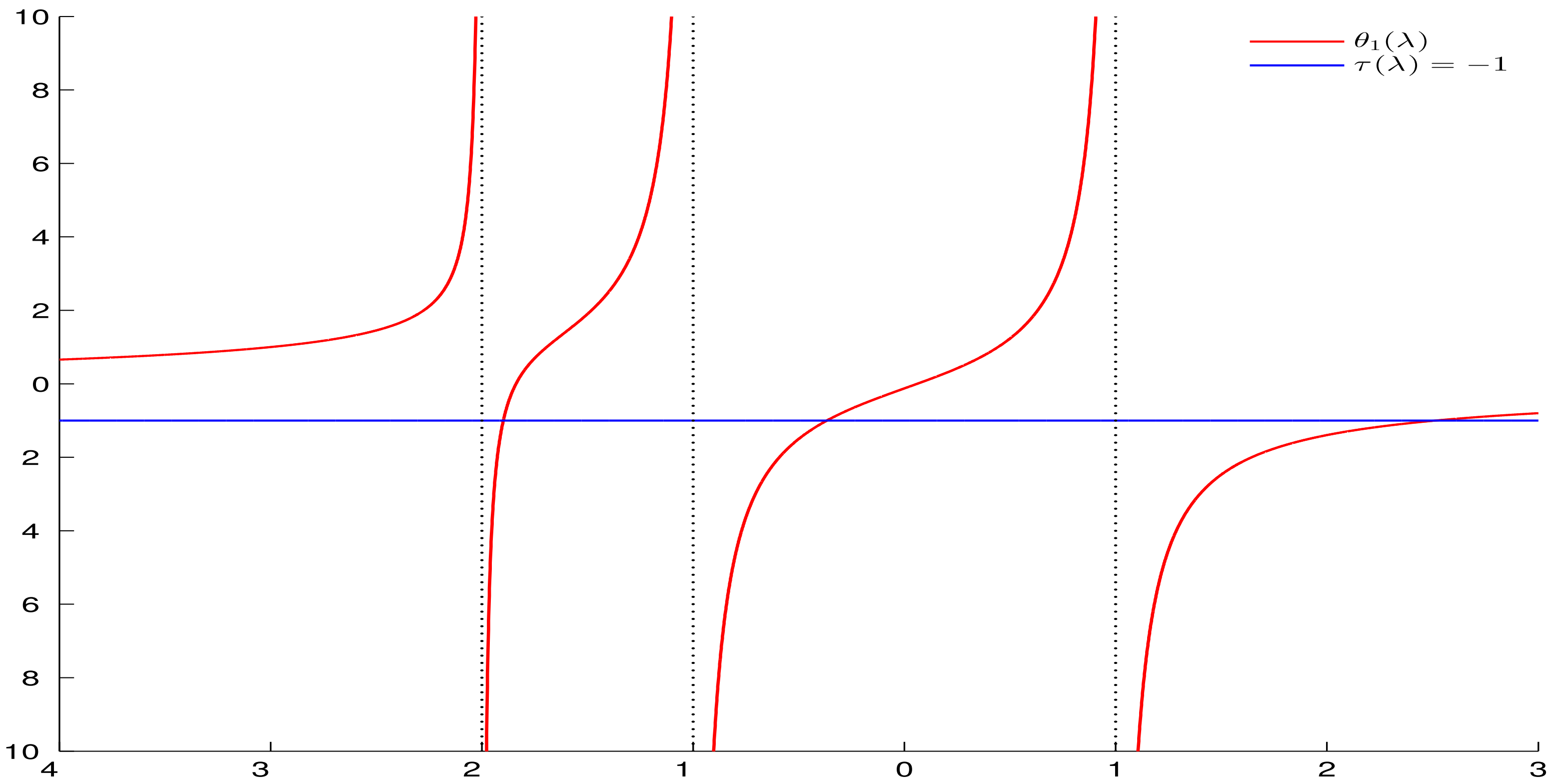


The type-I secular equation for the eigenvalues of a symmetric bordered matrix

Type-I secular equations

$$(\boldsymbol{\Lambda} - \boldsymbol{\lambda I})\bar{\boldsymbol{x}}(\boldsymbol{\lambda}) = \bar{\boldsymbol{b}} \quad \text{and} \quad \boldsymbol{\theta}_1(\boldsymbol{\lambda}) \equiv \bar{\boldsymbol{b}}^T \bar{\boldsymbol{x}}(\boldsymbol{\lambda}) = \boldsymbol{\tau}(\boldsymbol{\lambda}) \implies$$

$$\boldsymbol{\theta}_1(\boldsymbol{\lambda}) = \sum_{i=1}^{n} \frac{\bar{\boldsymbol{b}}_i^2}{\boldsymbol{\lambda}_i - \boldsymbol{\lambda}}$$

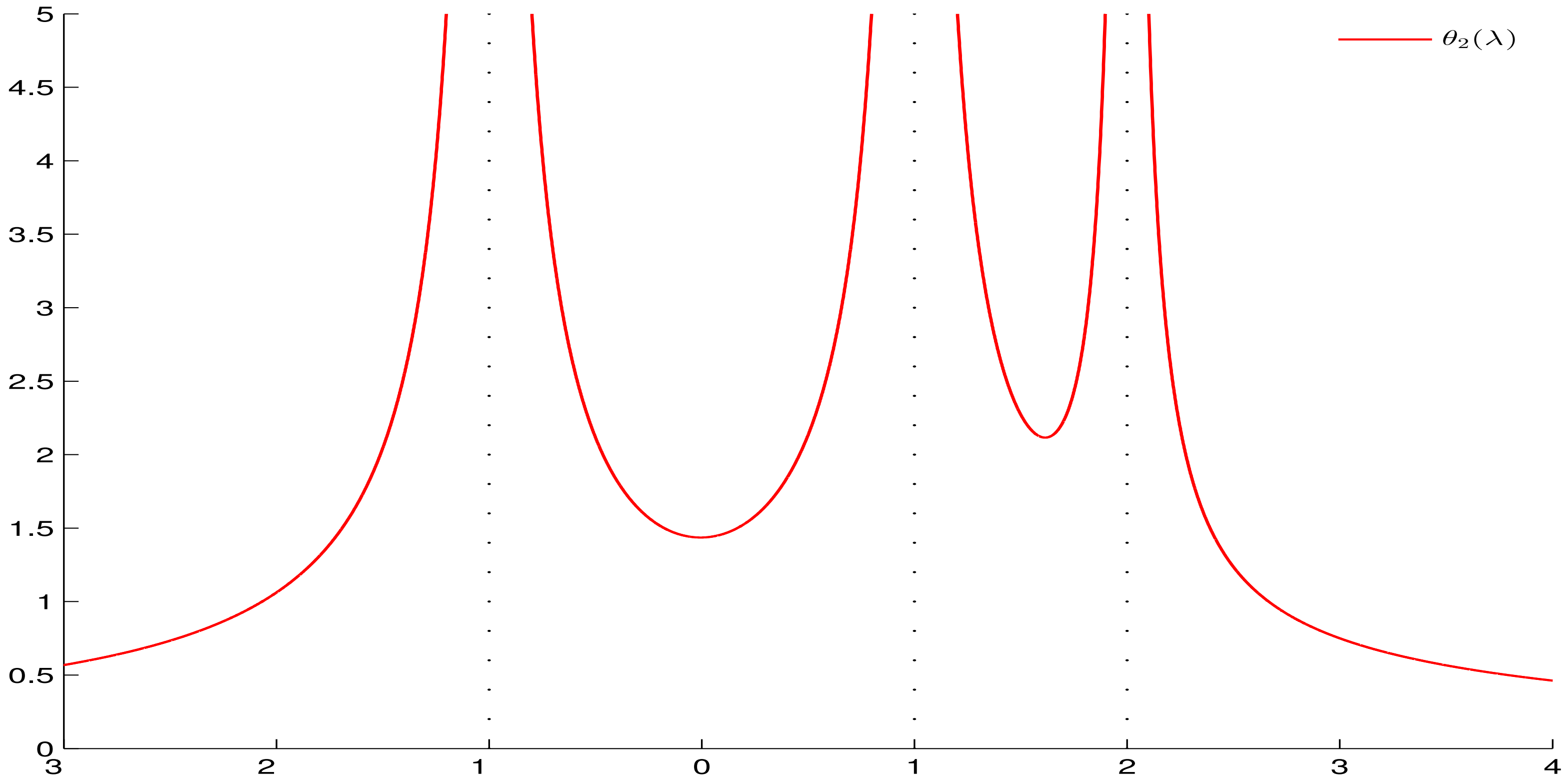


The type-I secular equation for the eigenvalues following a symmetric rank-one change

Type-II secular equations

$$(\Lambda + \lambda I)\bar{x}(\lambda) = \bar{b} \quad \text{and} \quad \theta_2(\lambda) \equiv \|\bar{x}(\lambda)\|_2 = \tau(\lambda) \implies$$

$$\theta_2(\lambda) = \sqrt{\sum_{i=1}^{n} \left(\frac{\bar{b}_i}{\lambda_i + \lambda} \right)^2}$$

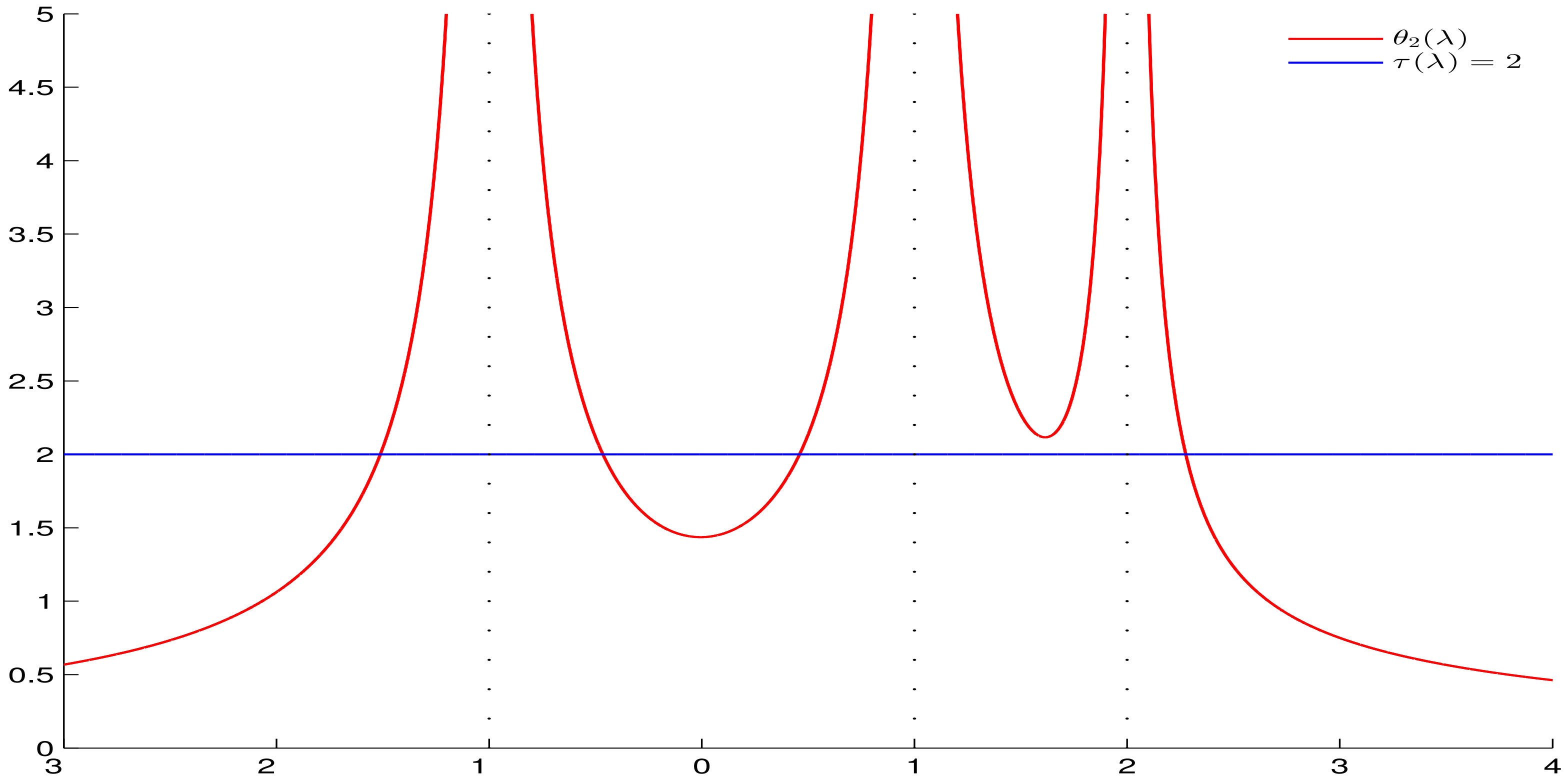


The type-II secular function $\theta_2(\lambda)$

Type-II secular equations

$$(\Lambda + \lambda I)\bar{x}(\lambda) = \bar{b} \ \text{ and } \ \theta_2(\lambda) \equiv \|\bar{x}(\lambda)\|_2 = \tau(\lambda) \implies$$

$$\theta_2(\lambda) = \sqrt{\sum_{i=1}^{n} \left(\frac{\bar{b}_i}{\lambda_i + \lambda} \right)^2}$$

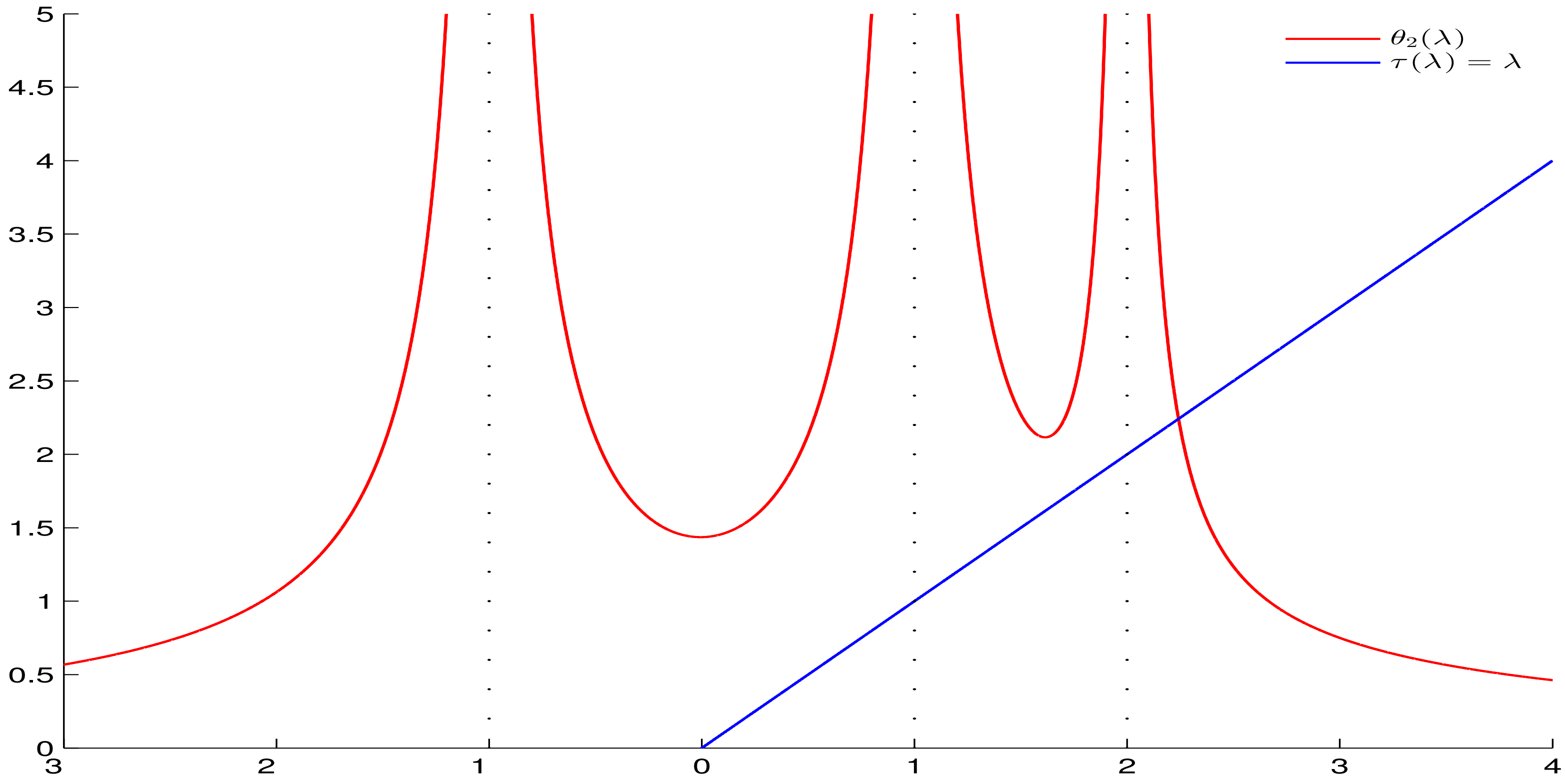


The type-II secular equation for the trust-region subproblem

Type-II secular equations

$$(\Lambda + \lambda I)\bar{x}(\lambda) = \bar{b} \;\; \text{and} \;\; \theta_2(\lambda) \equiv \|\bar{x}(\lambda)\|_2 = \tau(\lambda) \implies$$

$$\theta_2(\lambda) = \sqrt{\sum_{i=1}^{n} \left(\frac{\bar{b}_i}{\lambda_i + \lambda} \right)^2}$$



The type-II secular equation for cubic regularization of quadratic minimization

Explicit vs Implicit

In general most secular equations occur as one of two types:

- **Type I** secular equation:

  $$(A - \lambda I)x(\lambda) = b \;\; \text{and} \;\; \theta_1(\lambda) \equiv b^T x(\lambda) = \tau(\lambda)$$

  - usually more than one—maybe all—roots required
  - often eigenvalues/vectors known $\Longrightarrow$ reduces to explicit solution of

$$\sum_{i=1}^{n} \frac{d_i}{\lambda_i - \lambda} = \tau(\lambda) \;\; \text{with} \;\; d_i > 0$$

Explicit vs Implicit

In general most secular equations occur as one of two types:

- **Type I** secular equation:

$$(A - \lambda I)x(\lambda) = b \text{ and } \theta_1(\lambda) \equiv b^T x(\lambda) = \tau(\lambda)$$

  - usually more than one—maybe all—roots required
  - often eigenvalues/vectors known $\Longrightarrow$ reduces to explicit solution of

$$\sum_{i=1}^{n} \frac{d_i}{\lambda_i - \lambda} = \tau(\lambda) \text{ with } d_i > 0$$

- **Type II** secular equation:

$$(A + \lambda I)x(\lambda) = b \text{ and } \theta_2(\lambda) \equiv \|x(\lambda)\|_2 = \tau(\lambda)$$

  - usually just one—the largest—root required
  - usually eigenvalues/vectors unknown

Explicit vs Implicit

In general most secular equations occur as one of two types:

- **Type I** secular equation:

$$(A - \lambda I)x(\lambda) = b \quad \text{and} \quad \theta_1(\lambda) \equiv b^T x(\lambda) = \tau(\lambda)$$

  - usually more than one—maybe all—roots required
  - often eigenvalues/vectors known $\Longrightarrow$ reduces to explicit solution of

$$\sum_{i=1}^{n} \frac{d_i}{\lambda_i - \lambda} = \tau(\lambda) \quad \text{with} \quad d_i > 0$$

- **Type II** secular equation:

$$(A + \lambda I)x(\lambda) = b \quad \text{and} \quad \theta_2(\lambda) \equiv \|x(\lambda)\|_2 = \tau(\lambda)$$

  - usually just one—the largest—root required
  - usually eigenvalues/vectors unknown

$\tau(\lambda)$ usually a low-order (constant or linear) polynomial

Solving the explicit Type-I problem

Possible ways to find root $\in (\lambda_j, \lambda_{j+1})$ of

$$p_1(\lambda) + p_2(\lambda) \equiv \sum_{i=1}^{j} \frac{d_i}{\lambda_i - \lambda} + \sum_{i=j+1}^{n} \frac{d_i}{\lambda_i - \lambda} = \tau(\lambda) \quad \text{with} \quad d_i > 0$$

- construct low-order rational approximations $r_1^{(k)}$ and $r_2^{(k)}$ to p_1 and p_2 with poles λ_j and λ_{j+1}
- find appropriate root $\lambda^{(k)}$ of low-order polynomial equation

$$r_1^{(k)}(\lambda) + r_2^{(k)}(\lambda) = \tau(\lambda)$$

- appropriate approximations lead to global quadratic convergence of $\lambda^{(k)}$ from left or right of root, or locally from either side

(Bunch, Nielsen, Sorensen, 1978, Li, 1994)

Solving the explicit Type-I problem (continued)

Possible ways to find root $\in (\lambda_j, \lambda_{j+1})$ of

$$\sum_{i=1}^{n} \frac{d_i}{\lambda_i - \lambda} = \tau(\lambda) \quad \text{with} \quad d_i > 0$$

- use nonlinear change of variable $\lambda = \lambda_j + 1/\mu$
- apply standard fast method (Newton, Gragg, Halley, ...) to problem in transformed variable (better-behaved problem)
- superlinear to cubic global convergence established (Melman, 1995-97)

Solving the explicit Type-I problem (continued)

Possible ways to find root $\in (\lambda_j, \lambda_{j+1})$ of

$$\sum_{i=1}^{n} \frac{d_i}{\lambda_i - \lambda} = \tau(\lambda) \quad \text{with} \quad d_i > 0$$

- use nonlinear change of variable $\lambda = \lambda_j + 1/\mu$
- apply standard fast method (Newton, Gragg, Halley, ...) to problem in transformed variable (better-behaved problem)
- superlinear to cubic global convergence established (Melman, 1995-97)
- clever methods used to get good starting points for all methods
- accumulated evidence of good performance in practice (3-5 iterations for high accuracy)

Solving the explicit Type-I problem (continued)

Possible ways to find root $\in (\lambda_j, \lambda_{j+1})$ of

$$\sum_{i=1}^{n} \frac{d_i}{\lambda_i - \lambda} = \tau(\lambda) \quad \text{with} \quad d_i > 0$$

- use nonlinear change of variable $\lambda = \lambda_j + 1/\mu$
- apply standard fast method (Newton, Gragg, Halley, ...) to problem in transformed variable (better-behaved problem)
- superlinear to cubic global convergence established (Melman, 1995-97)
- clever methods used to get good starting points for all methods
- accumulated evidence of good performance in practice (3-5 iterations for high accuracy)
- $O(n)$ approximation methods for **all** roots related to multi-level integral transform evaluation (Livne, Brandt, 2002)

Solving the implicit Type-II problem

Find the largest root $\geq \max(0, -\lambda_1)$ of

$$\theta_2(\lambda) \equiv \|x(\lambda)\|_2 = \tau(\lambda) \quad \text{where} \quad (A + \lambda I)x(\lambda) = b$$

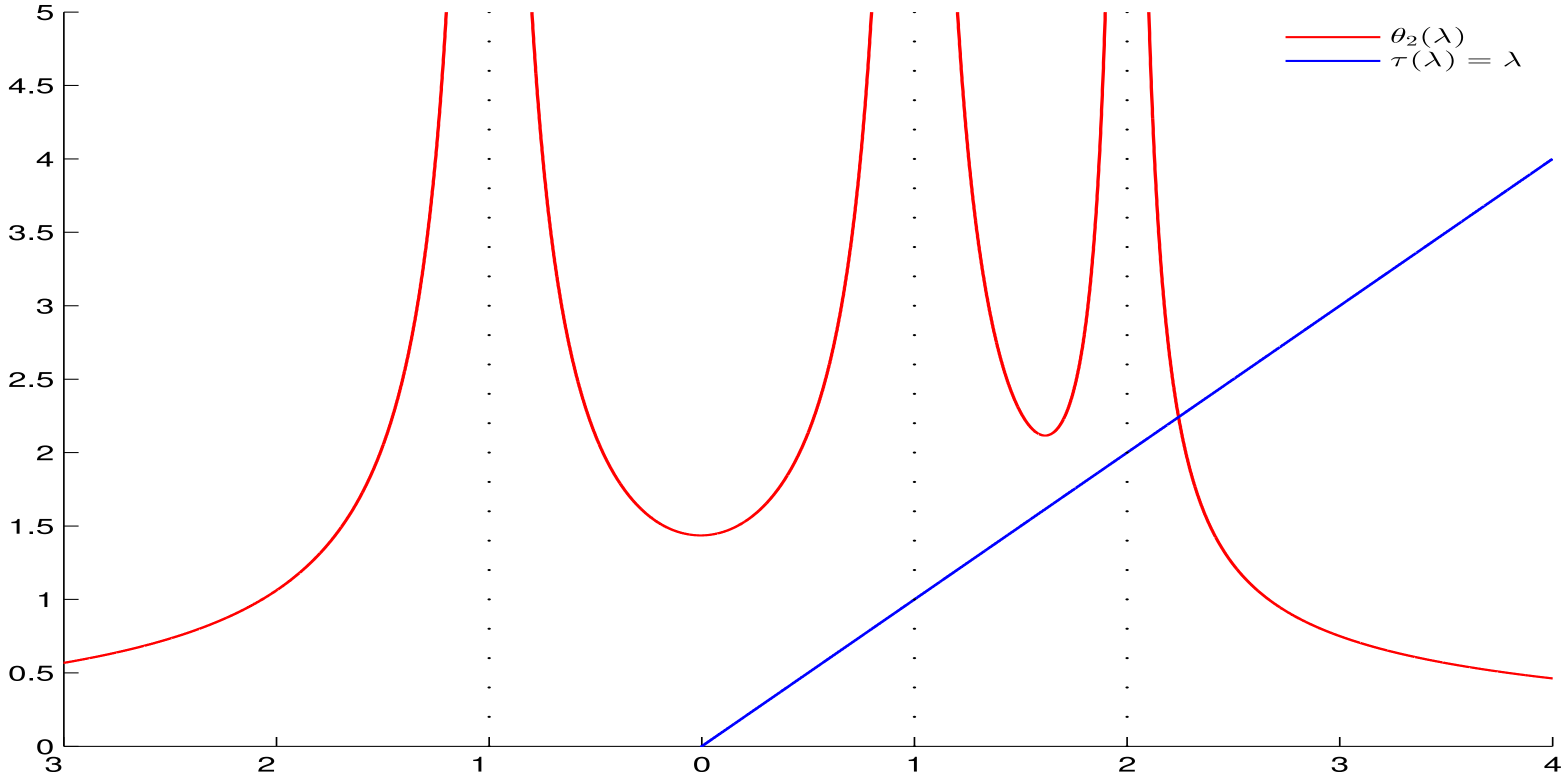
- usually $\tau(\lambda)$ is convex, increasing for positive λ
  $\Longrightarrow$ unique root in $[\max(0, -\lambda_1), \infty)$, but generally λ_1 unknown
- need to use effective rootfinder that "understands" structure of $\theta_2(\lambda)$
- need to evaluate $\theta_2(\lambda)$ and its derivatives

Solving the implicit Type-II problem

Find the largest root $\geq \max(0, -\lambda_1)$ of

$$\theta_2(\lambda) \equiv \|x(\lambda)\|_2 = \tau(\lambda) \quad \text{where} \quad (A + \lambda I)x(\lambda) = b$$

- “degenerate” **hard** case if $b \perp$ eigenvector(s) q_1 for λ_1



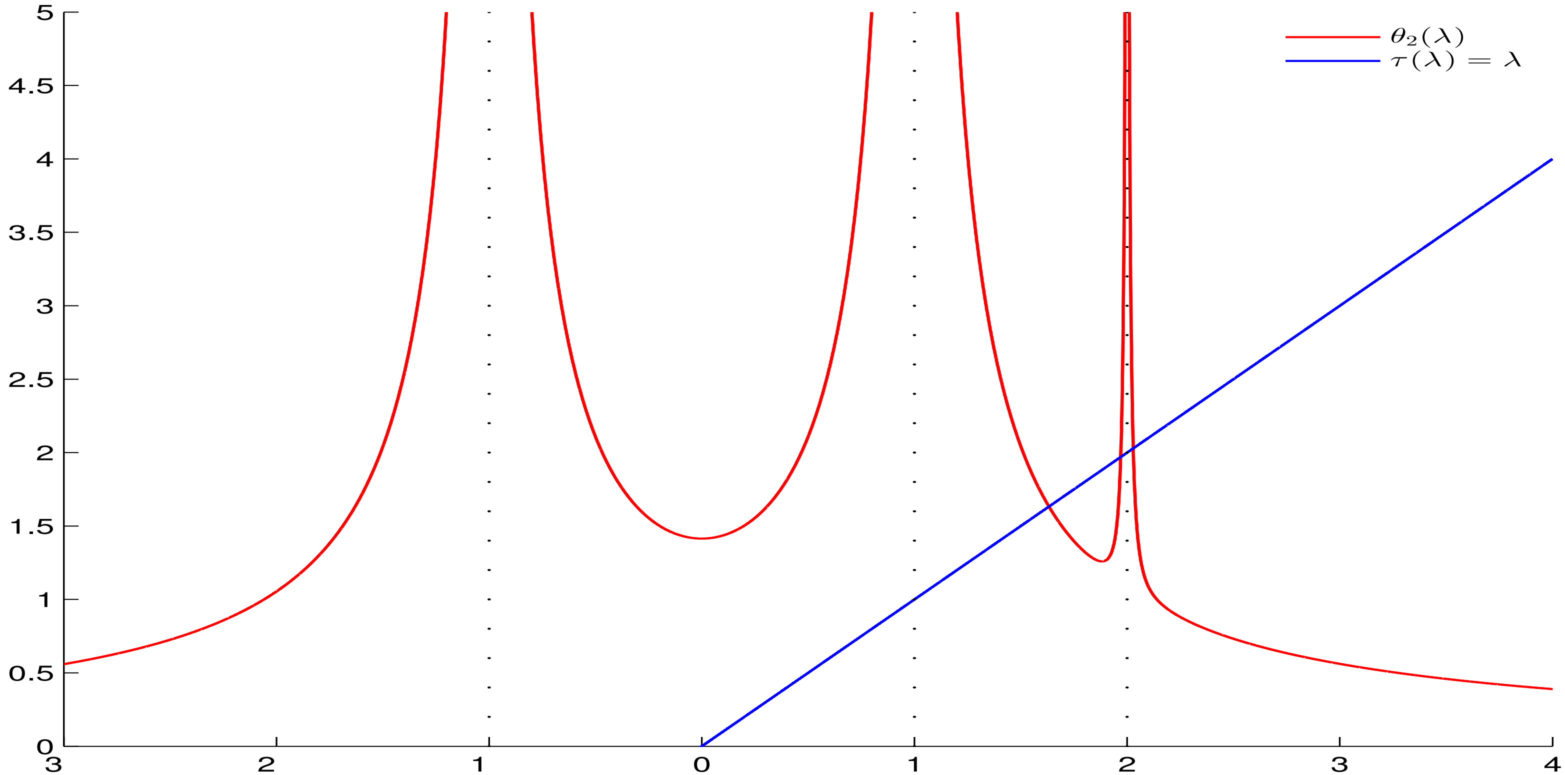
The type-II secular equation for cubic regularization of quadratic minimization - easy case

Solving the implicit Type-II problem

Find the largest root $\geq \max(0, -\lambda_1)$ of

$$\theta_2(\lambda) \equiv \|x(\lambda)\|_2 = \tau(\lambda) \quad \text{where} \quad (A + \lambda I)x(\lambda) = b$$

- “degenerate” **hard** case if $b \perp$ eigenvector(s) q_1 for λ_1



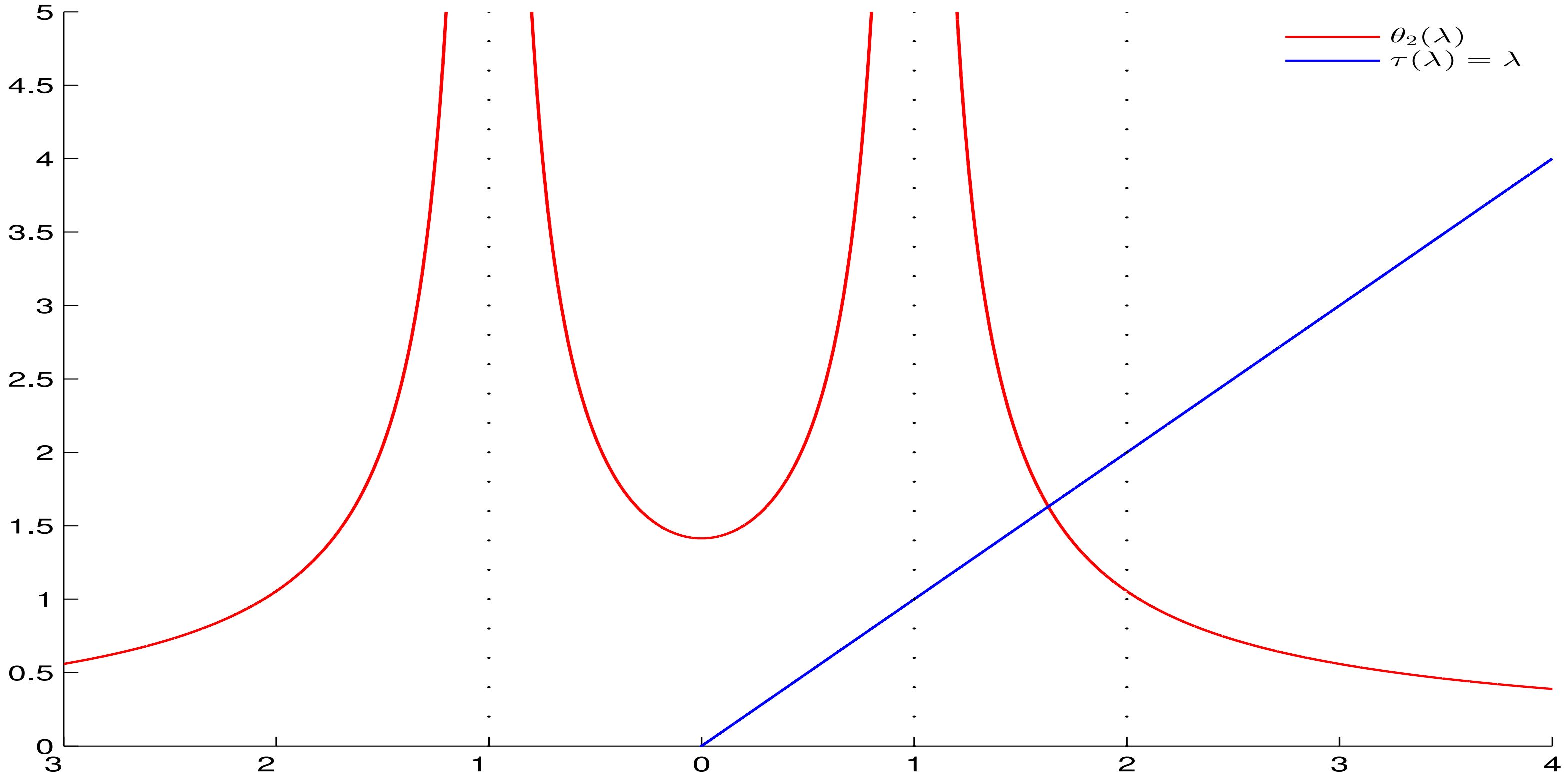
The type-II secular equation for cubic regularization of quadratic minimization - moderate case

Solving the implicit Type-II problem

Find the largest root $\geq \max(0, -\lambda_1)$ of

$$\theta_2(\lambda) \equiv \|x(\lambda)\|_2 = \tau(\lambda) \quad \text{where} \quad (A + \lambda I)x(\lambda) = b$$

- "degenerate" **hard** case if $b \perp$ eigenvector(s) q_1 for λ_1



The type-II secular equation for cubic regularization of quadratic minimization - hard case

Solving the implicit Type-II problem

Find the largest root $\geq \max(0, -\lambda_1)$ of

$$\theta_2(\lambda) \equiv \|x(\lambda)\|_2 = \tau(\lambda) \quad \text{where} \quad (A + \lambda I)x(\lambda) = b$$

- "degenerate" **hard** case if $b \perp$ eigenvector(s) q_1 for λ_1
  - solution in hard case is $x(-\lambda_1) + \alpha q_1$ where α satisfies

$$\|x(-\lambda_1) + \alpha q_1\|_2 = \tau(-\lambda_1)$$

Solving the implicit Type-II problem

Find the largest root $\geq \max(0, -\lambda_1)$ of

$$\theta_2(\lambda) \equiv \|x(\lambda)\|_2 = \tau(\lambda) \quad \text{where} \quad (A + \lambda I)x(\lambda) = b$$

- may equivalently find the largest root $\geq \max(0, -\lambda_1)$ of

$$\pi(\lambda) \equiv \|x(\lambda)\|_2^2 = \tau^2(\lambda) \equiv \rho(\lambda) \quad \text{where} \quad (A + \lambda I)x(\lambda) = b$$

  since derivatives of $\pi(\lambda)$ are easy to find

Calculating the secular function and its derivatives

Need to compute

$$\pi(\lambda) \equiv \|x(\lambda)\|_2^2 = x^T(\lambda)x(\lambda) \quad \text{where} \quad (A + \lambda I)x(\lambda) = b$$

and its derivatives

- derivatives $x^{(k)}(\lambda)$ of $x(\lambda)$

$$(A + \lambda I)x^{(1)} = -x, \quad (A + \lambda I)x^{(2)} = -2x^{(1)}$$

Calculating the secular function and its derivatives

Need to compute

$$\pi(\lambda) \equiv \|x(\lambda)\|_2^2 = x^T(\lambda)x(\lambda) \quad \text{where} \quad (A + \lambda I)x(\lambda) = b$$

and its derivatives

- derivatives $x^{(k)}(\lambda)$ of $x(\lambda)$

$$(A + \lambda I)x^{(1)} = -x, \quad (A + \lambda I)x^{(2)} = -2x^{(1)}$$
$$\ldots \; (A + \lambda I)x^{(k)} = -kx^{(k-1)}$$

Calculating the secular function and its derivatives

Need to compute

$$\pi(\lambda) \equiv \|x(\lambda)\|_2^2 = x^T(\lambda)x(\lambda) \quad \text{where} \quad (A + \lambda I)x(\lambda) = b$$

and its derivatives

- derivatives $x^{(k)}(\lambda)$ of $x(\lambda)$

$$(A + \lambda I)x^{(1)} = -x, \quad (A + \lambda I)x^{(2)} = -2x^{(1)}$$

- derivatives $\pi^{(k)}(\lambda)$ of $\pi(\lambda)$

$$\pi^{(1)} = 2x^{(1)T}x, \quad \pi^{(2)} = 2x^{(2)T}x + 2x^{(1)T}x^{(1)} = 6x^{(1)T}x^{(1)}$$

Calculating the secular function and its derivatives

Need to compute

$$\pi(\lambda) \equiv \|x(\lambda)\|_2^2 = x^T(\lambda)x(\lambda) \quad \text{where} \quad (A + \lambda I)x(\lambda) = b$$

and its derivatives

- derivatives $x^{(k)}(\lambda)$ of $x(\lambda)$

$$(A + \lambda I)x^{(k)} = -kx^{(k-1)}$$

- derivatives $\pi^{(k)}(\lambda)$ of $\pi(\lambda)$

$$\pi^{(1)} = 2x^{(1)T}x, \quad \pi^{(2)} = 6x^{(1)T}x^{(1)}$$
$$\ldots \ \pi^{(2k+1)} = 2\alpha_k x^{(k)T}x^{(k+1)}, \quad \pi^{(2k+2)} = \alpha_{k+1}x^{(k+1)T}x^{(k+1)}$$
$$\text{where} \quad \alpha_{k+1} = 2(2k+3)\alpha_k/(k+1)$$

Science & Technology Facilities Council

Calculating the secular function and its derivatives

Need to compute

$$\pi(\lambda) \equiv \|x(\lambda)\|_2^2 = x^T(\lambda)x(\lambda) \quad \text{where} \quad (A+\lambda I)x(\lambda) = b$$

and its derivatives

- derivatives $x^{(k)}(\lambda)$ of $x(\lambda)$

$$(A+\lambda I)x^{(1)} = -x$$

- derivatives $\pi^{(k)}(\lambda)$ of $\pi(\lambda)$

$$\pi^{(1)} = 2x^{(1)T}x, \quad \pi^{(2)} = 6x^{(1)T}x^{(1)}$$

- given the Cholesky factorization $A+\lambda I = LL^T$ and

$$Ly^{(1)} = -x, \quad L^T x^{(1)} = y^{(1)} \quad \Longrightarrow$$
$$\pi^{(1)} = -2\|y^{(1)}\|_2^2, \quad \pi^{(2)} = 6\|x^{(1)}\|_2^2$$

Calculating the secular function and its derivatives

Need to compute

$$\pi(\lambda) \equiv \|x(\lambda)\|_2^2 = x^T(\lambda)x(\lambda) \quad \text{where} \quad (A + \lambda I)x(\lambda) = b$$

and its derivatives

- derivatives $x^{(k)}(\lambda)$ of $x(\lambda)$

$$(A + \lambda I)x^{(k)} = -kx^{(k-1)}$$

- derivatives $\pi^{(k)}(\lambda)$ of $\pi(\lambda)$

$$\pi^{(2k+1)} = 2\alpha_k x^{(k)T} x^{(k+1)}, \quad \pi^{(2k+2)} = \alpha_{k+1} x^{(k+1)T} x^{(k+1)}$$
$$\text{where} \quad \alpha_{k+1} = 2(2k+3)\alpha_k/(k+1)$$

- given the Cholesky factorization $A + \lambda I = LL^T$ and

$$Ly^{(k)} = -kx^{(k-1)}, \quad L^T x^{(k)} = y^{(k)} \implies$$
$$\pi^{(2k+1)} = -2\beta_k \|y^{(k)}\|_2^2, \quad \pi^{(2k+2)} = \alpha_{k+1} \|x^{(k)}\|_2^2$$
$$\text{where} \quad \beta_k = 2\alpha_k/(k+1) \quad \text{and} \quad \alpha_{k+1} = (2k+3)\beta_k$$

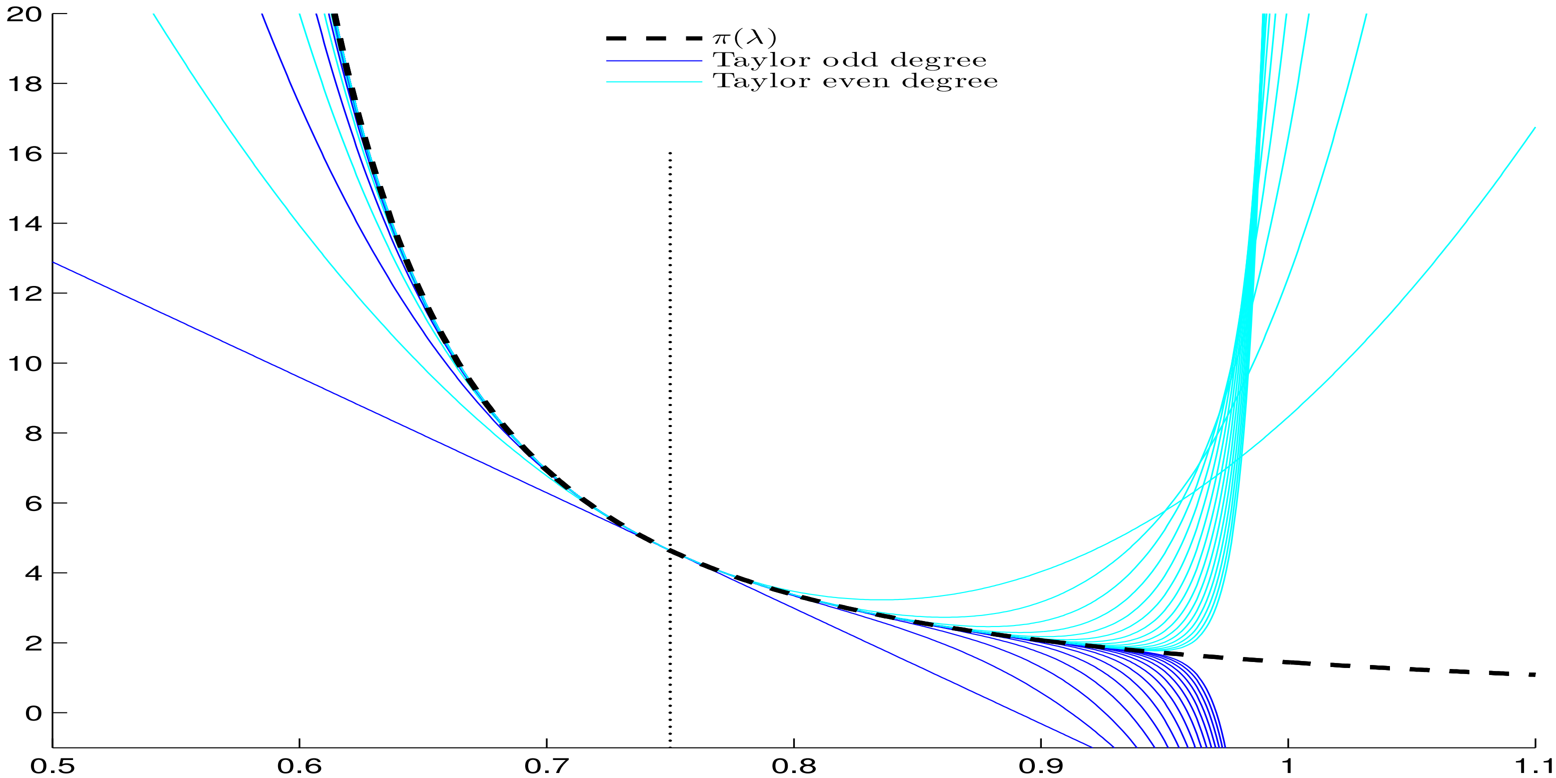
- one triangular solve per order of derivative

Taylor-series approximations

Let $p_i(\lambda)$ by the ith order Taylor approximation to $\pi(\lambda)$ at $\lambda_c > -\lambda_1 \Longrightarrow$

- $p_{2k+1}(\lambda)$ underestimates $\pi(\lambda)$ for $\lambda > -\lambda_1$ for all k
- $p_k(\lambda)$ underestimates $\pi(\lambda)$ for $-\lambda_1 < \lambda < \lambda_c$ for all k

(G., Robinson, Thorne, 2011)



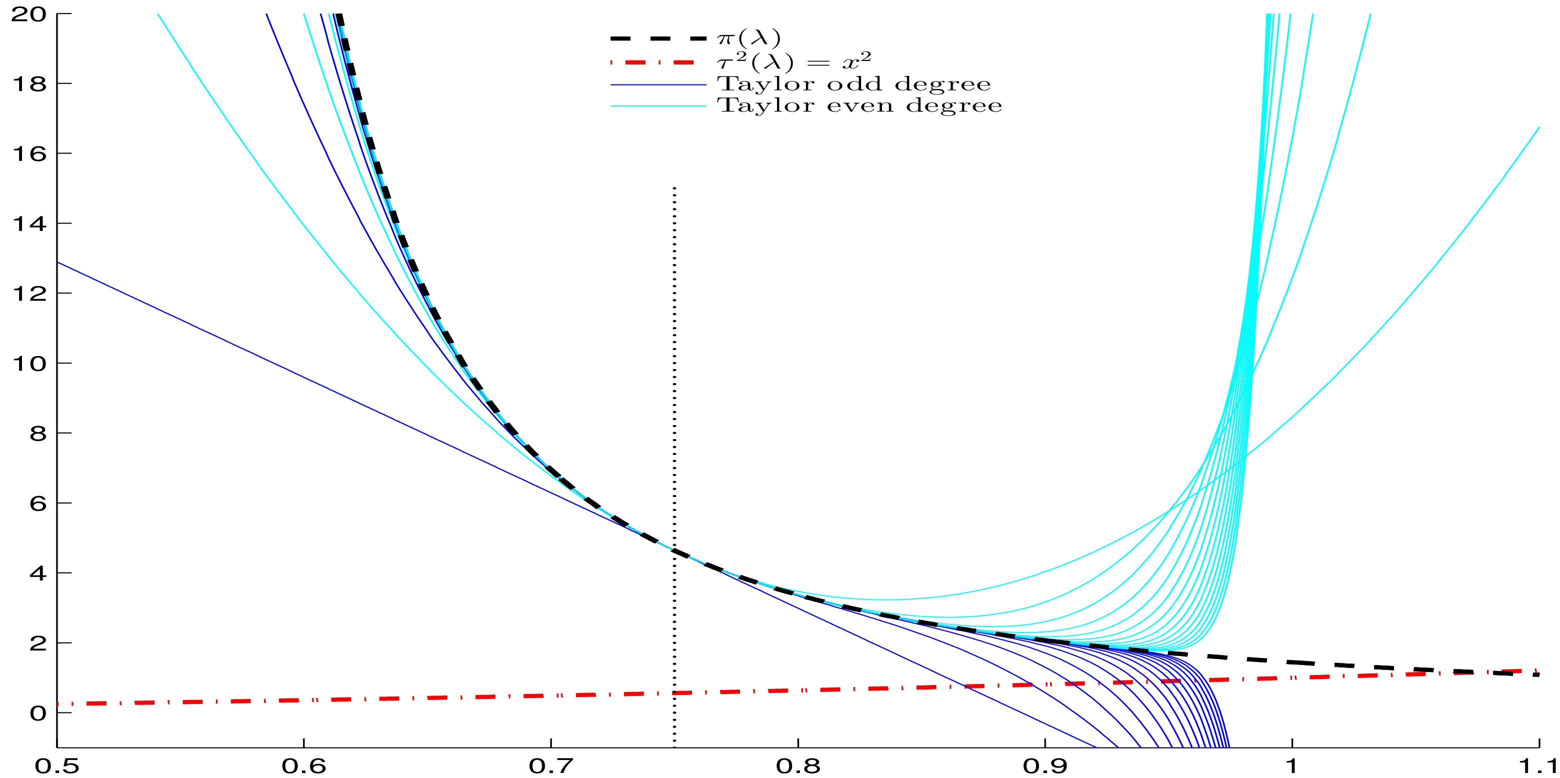
$\pi(\lambda)$ and its first 30 Taylor-series approximations $p_i(\lambda)$ at $\lambda_c = 0.75$

Taylor-series approximations

Let $p_i(\lambda)$ by the ith order Taylor approximation to $\pi(\lambda)$ at $\lambda_c > -\lambda_1 \Longrightarrow$

- $p_{2k+1}(\lambda)$ underestimates $\pi(\lambda)$ for $\lambda > -\lambda_1$ for all k
- $p_k(\lambda)$ underestimates $\pi(\lambda)$ for $-\lambda_1 < \lambda < \lambda_c$ for all k

(G., Robinson, Thorne, 2011)



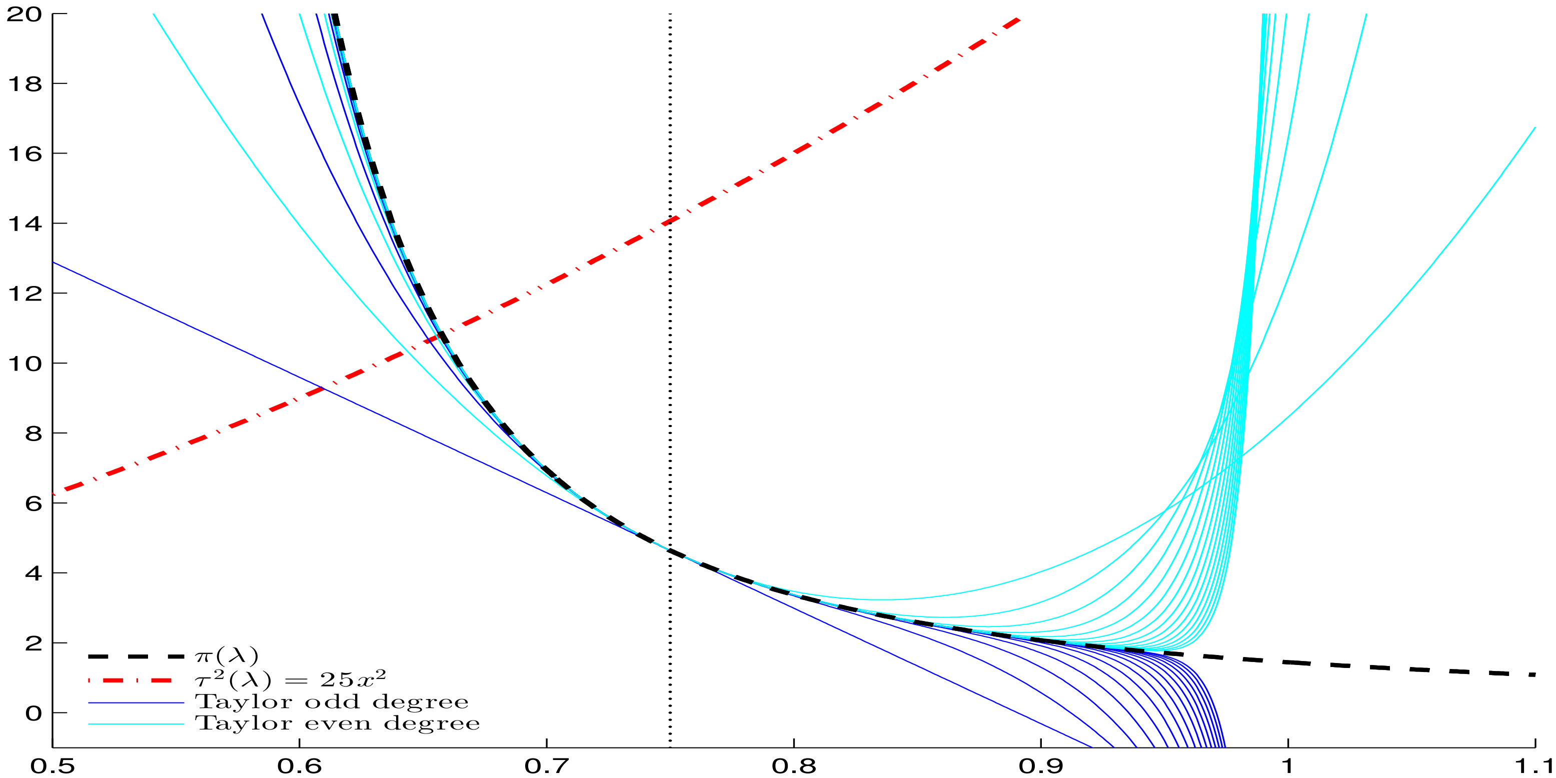
Root of $p_{2k+1}(\lambda) = \rho(\lambda) <$ root λ_* of $\pi(\lambda) = \rho(\lambda)$ for $\rho(\lambda) \nearrow$ when $\lambda_c < \lambda_*$

Taylor-series approximations

Let $p_i(\lambda)$ by the ith order Taylor approximation to $\pi(\lambda)$ at $\lambda_c > -\lambda_1 \Longrightarrow$

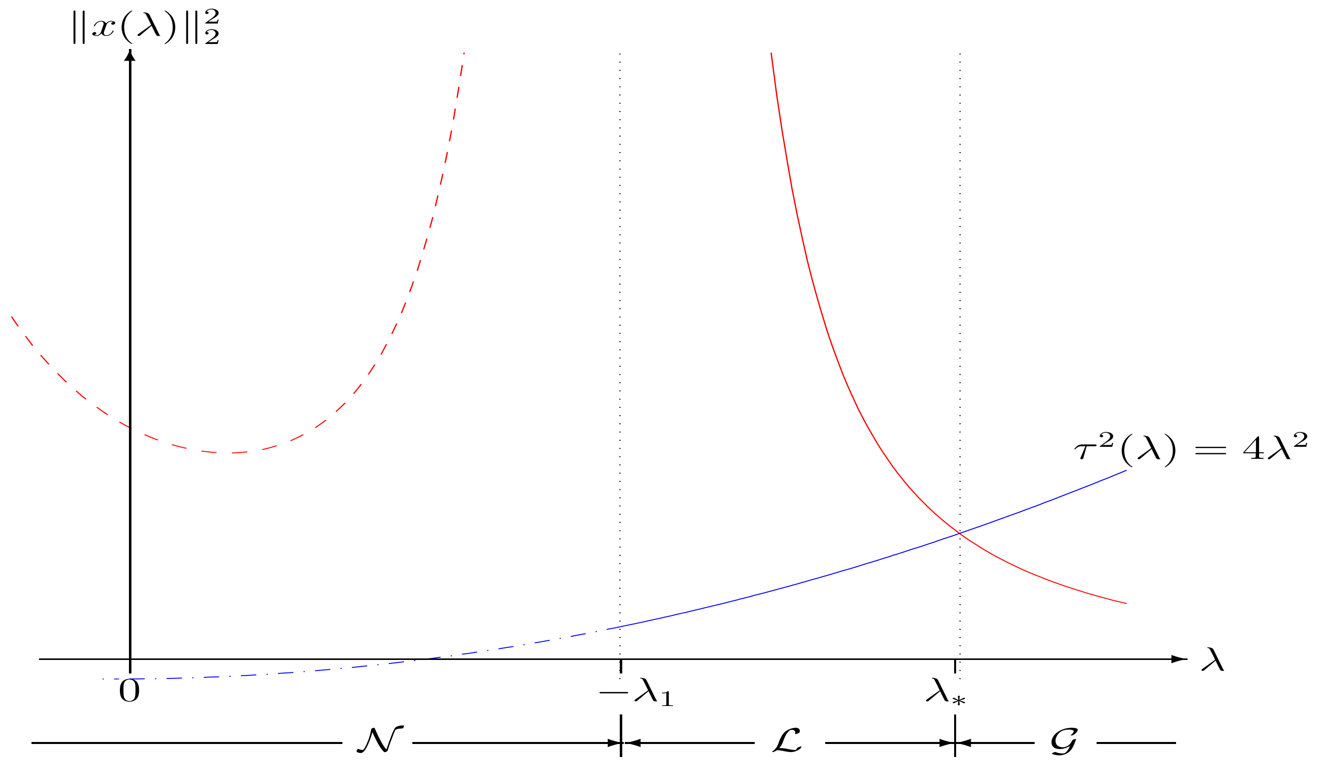
- $p_{2k+1}(\lambda)$ underestimates $\pi(\lambda)$ for $\lambda > -\lambda_1$ for all k
- $p_k(\lambda)$ underestimates $\pi(\lambda)$ for $-\lambda_1 < \lambda < \lambda_c$ for all k

(G., Robinson, Thorne, 2011)



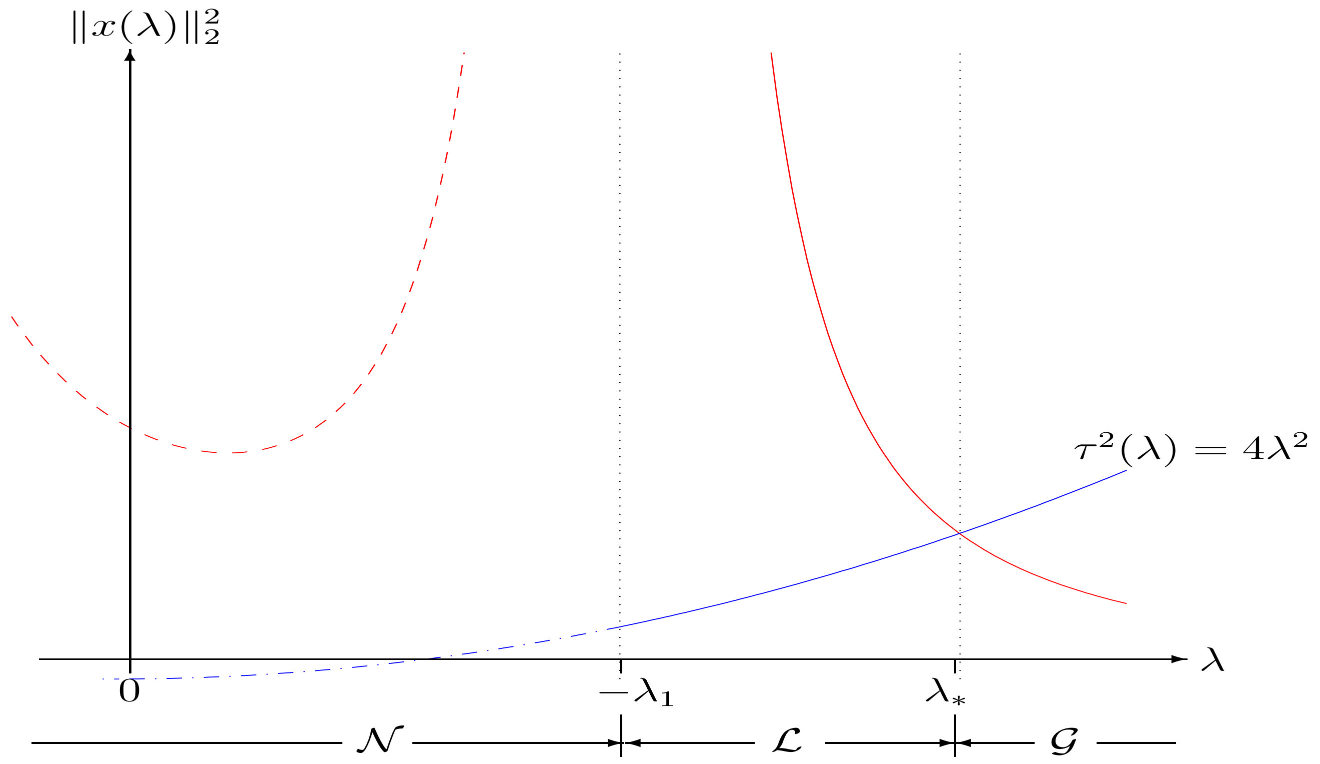
Root of $p_k(\lambda) = \rho(\lambda) <$ root λ_* of $\pi(\lambda) = \rho(\lambda)$ for $\rho(\lambda) \nearrow$ when $\lambda_* < \lambda_c$

Algorithm overview



- iterate root of $p_{2k+1}(\lambda) = \tau^2(\lambda)$ within $\mathcal{L} \Longrightarrow$ globally convergent with Q-order $2k+2$ (G., Robinson, Thorne, 2011)

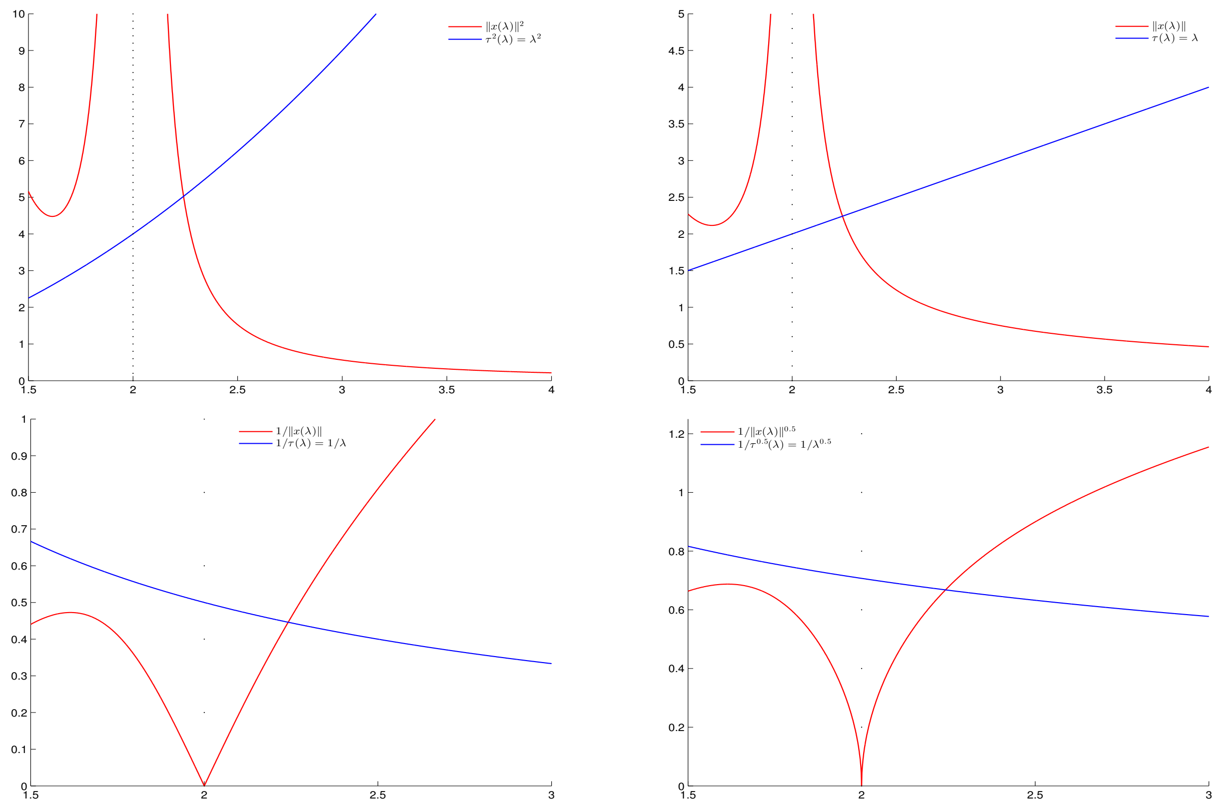
Algorithm overview



- iterate root of $p_{2k+1}(\lambda) = \tau^2(\lambda)$ within $\mathcal{L} \Longrightarrow$ globally convergent with Q-order $2k+2$ (G., Robinson, Thorne, 2011)
- safeguarded bisection to locate $\mathcal{L}$
  - $\lambda \in \mathcal{N} \iff H + \lambda I \not\succ 0$
  - $\lambda \in \mathcal{G}$ when $\pi(\lambda) < \tau^2(\lambda) \Longrightarrow$ root of $p_k(\lambda) = \tau^2(\lambda) \notin \mathcal{G}$

Nonlinear transformation

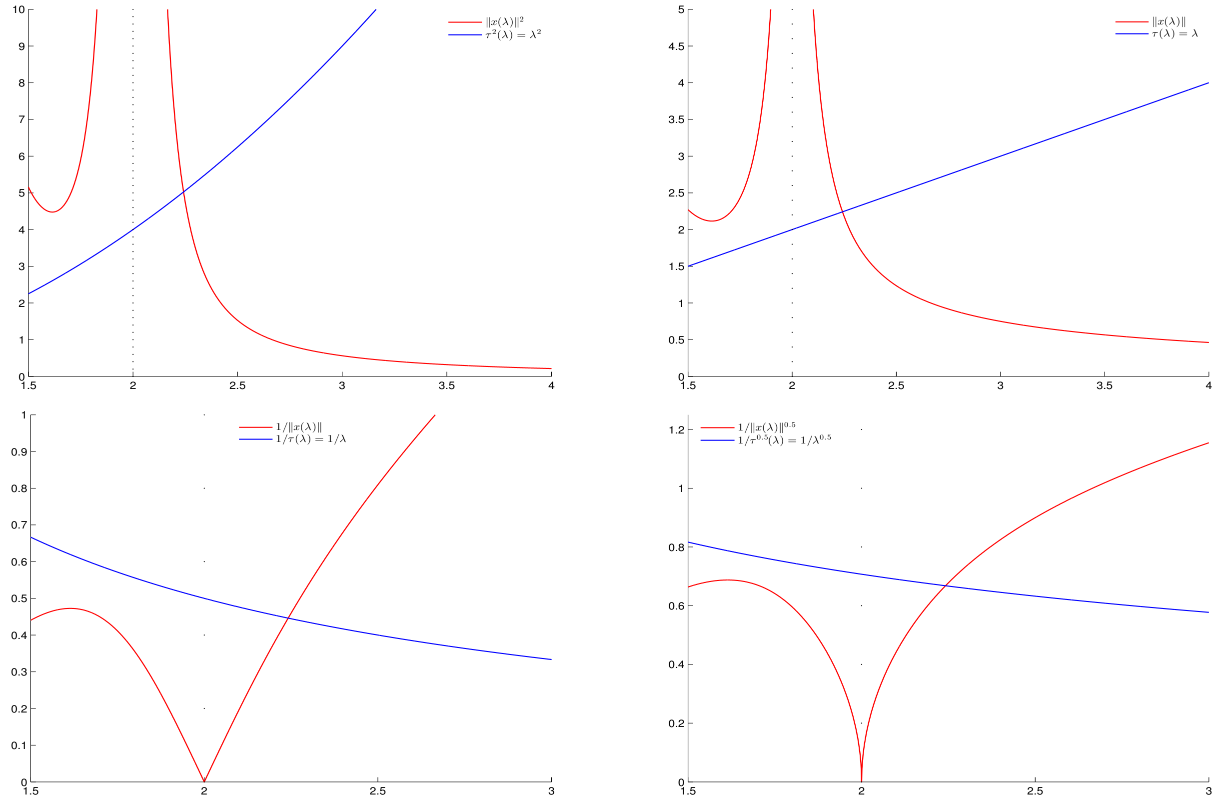
- for $\lambda > -\lambda_1$: $\|x(\lambda)\|_2^\beta$ convex for $\beta > 0$, concave for $\beta \in [-1, 0)$
- solve instead $\|x(\lambda)\|_2^\beta = \tau^\beta(\lambda)$ (Reinsch, 1971)



Clockwise from top left: $\beta = 2, 1, -0.5, -1$

Nonlinear transformation

- for $\lambda > -\lambda_1$: $\|x(\lambda)\|_2^{\beta}$ convex for $\beta > 0$, concave for $\beta \in [-1, 0)$
- solve instead $\|x(\lambda)\|_2^{\beta} = \tau^{\beta}(\lambda)$ (Reinsch, 1971)



Clockwise from top left: $\beta = 2, 1, -0.5, -1$

- Taylor approximation $p_1(\lambda)$ to $\|x(\lambda)\|_2^{\beta}$ for $\beta \geq -1$ underestimates root for $\lambda \in (-\lambda_1, \lambda_*)$ (G., Robinson, Thorne, 2011)

Solving the implicit problem by iteration

What if factorization of $H + \lambda I$ is not possible?

- pick $n \times k$ matrices Q_k with orthonormal columns
- look for approximation to the root $x = Q_k x_k$, where

$$\begin{array}{l} (H + \lambda I) Q_k x_k(\lambda) = b \ \text{ and } \ \|Q_k x_k(\lambda)\|_2 = \tau(\lambda) \\ \Longrightarrow \\ (Q_k^T H Q_k + \lambda I) x_k(\lambda) = Q_k^T b \ \text{ and } \ \|x_k(\lambda)\|_2 = \tau(\lambda) \end{array}$$

Solving the implicit problem by iteration

What if factorization of $H + \lambda I$ is not possible?

- pick $n \times k$ matrices Q_k with orthonormal columns
- look for approximation to the root $x = Q_k x_k$, where

$$(H + \lambda I) Q_k x_k(\lambda) = b \quad \text{and} \quad \|Q_k x_k(\lambda)\|_2 = \tau(\lambda)$$

$$\Longrightarrow$$

$$(Q_k^T H Q_k + \lambda I) x_k(\lambda) = Q_k^T b \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$$

- if Q_k is generated by the **Lanczos** method (G., Lucidi, Roma, Toint, 1999)
  - $Q_k^T H Q_k =$ tridiagonal T_k and $Q_k^T b = \|b\|_2 e_1$
  - $(T_k + \lambda I) x_k(\lambda) = \|b\|_2 e_1 \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$

Solving the implicit problem by iteration

What if factorization of $H + \lambda I$ is not possible?

- pick $n \times k$ matrices Q_k with orthonormal columns
- look for approximation to the root $x = Q_k x_k$, where

$$(H + \lambda I) Q_k x_k(\lambda) = b \quad \text{and} \quad \|Q_k x_k(\lambda)\|_2 = \tau(\lambda)$$

$$\Longrightarrow$$

$$(Q_k^T H Q_k + \lambda I) x_k(\lambda) = Q_k^T b \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$$

- if Q_k is generated by the **Lanczos** method (G., Lucidi, Roma, Toint, 1999)
  - $Q_k^T H Q_k =$ tridiagonal T_k and $Q_k^T b = \|b\|_2 e_1$
  - $(T_k + \lambda I) x_k(\lambda) = \|b\|_2 e_1 \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$
  - Cholesky factorization of $T_k + \lambda I$ possible in $O(k)$ flops

Solving the implicit problem by iteration

What if factorization of $H + \lambda I$ is not possible?

- pick $n \times k$ matrices Q_k with orthonormal columns
- look for approximation to the root $x = Q_k x_k$, where

$$(H + \lambda I) Q_k x_k(\lambda) = b \quad \text{and} \quad \|Q_k x_k(\lambda)\|_2 = \tau(\lambda)$$

$$\Longrightarrow$$

$$(Q_k^T H Q_k + \lambda I) x_k(\lambda) = Q_k^T b \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$$

- if Q_k is generated by the **Lanczos** method (G., Lucidi, Roma, Toint, 1999)
  - $Q_k^T H Q_k =$ tridiagonal T_k and $Q_k^T b = \|b\|_2 e_1$
  - $(T_k + \lambda I) x_k(\lambda) = \|b\|_2 e_1$ and $\|x_k(\lambda)\|_2 = \tau(\lambda)$
  - Cholesky factorization of $T_k + \lambda I$ possible in $O(k)$ flops
  - $\|(H + \lambda_k I) Q_k x_k - b\|_2 = \gamma_k e_k^T x_k$, where $\gamma_k = (T_k)_{k,k-1}$

Solving the implicit problem by iteration

What if factorization of $H + \lambda I$ is not possible?

- pick $n \times k$ matrices Q_k with orthonormal columns
- look for approximation to the root $x = Q_k x_k$, where

$$(H + \lambda I) Q_k x_k(\lambda) = b \quad \text{and} \quad \|Q_k x_k(\lambda)\|_2 = \tau(\lambda)$$

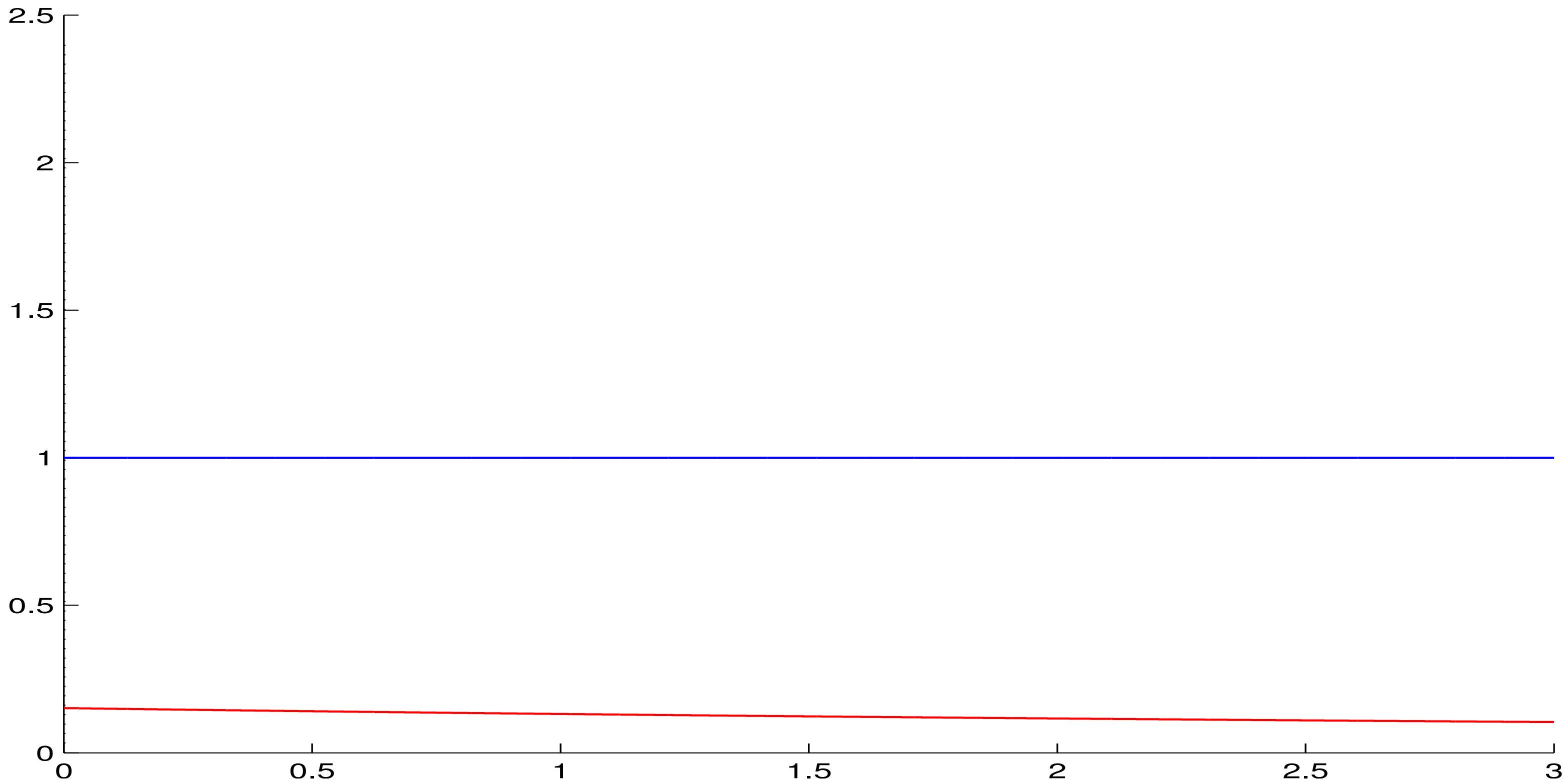
$$\Longrightarrow$$

$$(Q_k^T H Q_k + \lambda I) x_k(\lambda) = Q_k^T b \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$$

- if Q_k is generated by the **Lanczos** method (G., Lucidi, Roma, Toint, 1999)
  - $Q_k^T H Q_k =$ tridiagonal T_k and $Q_k^T b = \|b\|_2 e_1$
  - $(T_k + \lambda I) x_k(\lambda) = \|b\|_2 e_1 \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$
  - Cholesky factorization of $T_k + \lambda I$ possible in $O(k)$ flops
  - $\|(H + \lambda_k I) Q_k x_k - b\|_2 = \gamma_k e_k^T x_k$, where $\gamma_k = (T_k)_{k,k-1}$
  - good starting guess for λ_{k+1} from kth problem

How the secular equation evolves

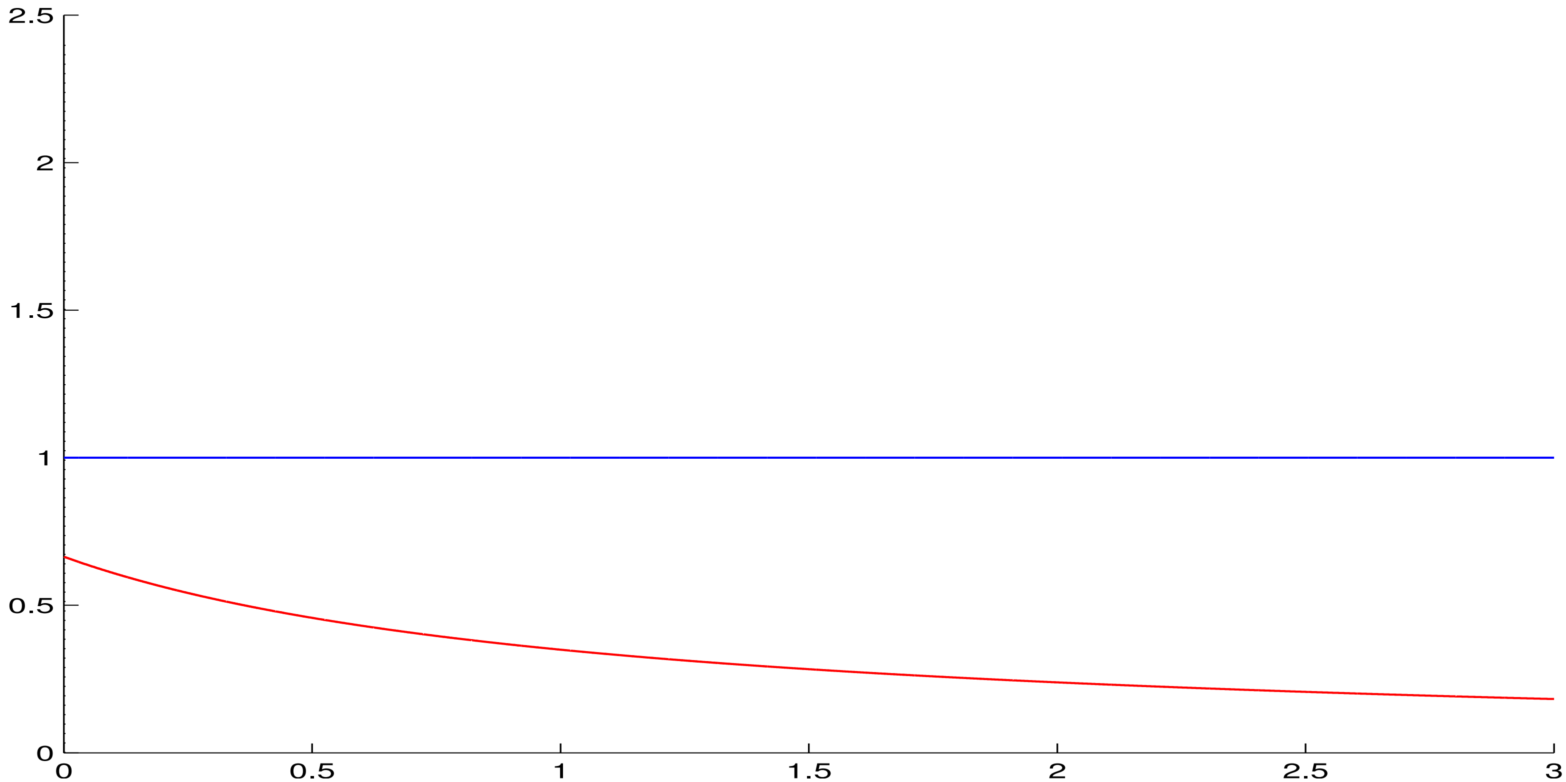
$$(T_k + \lambda I)x_k(\lambda) = \|b\|_2 e_1 \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$$



Iteration k = 1 for trust-region problem ($\tau(\lambda) = 1$)

How the secular equation evolves

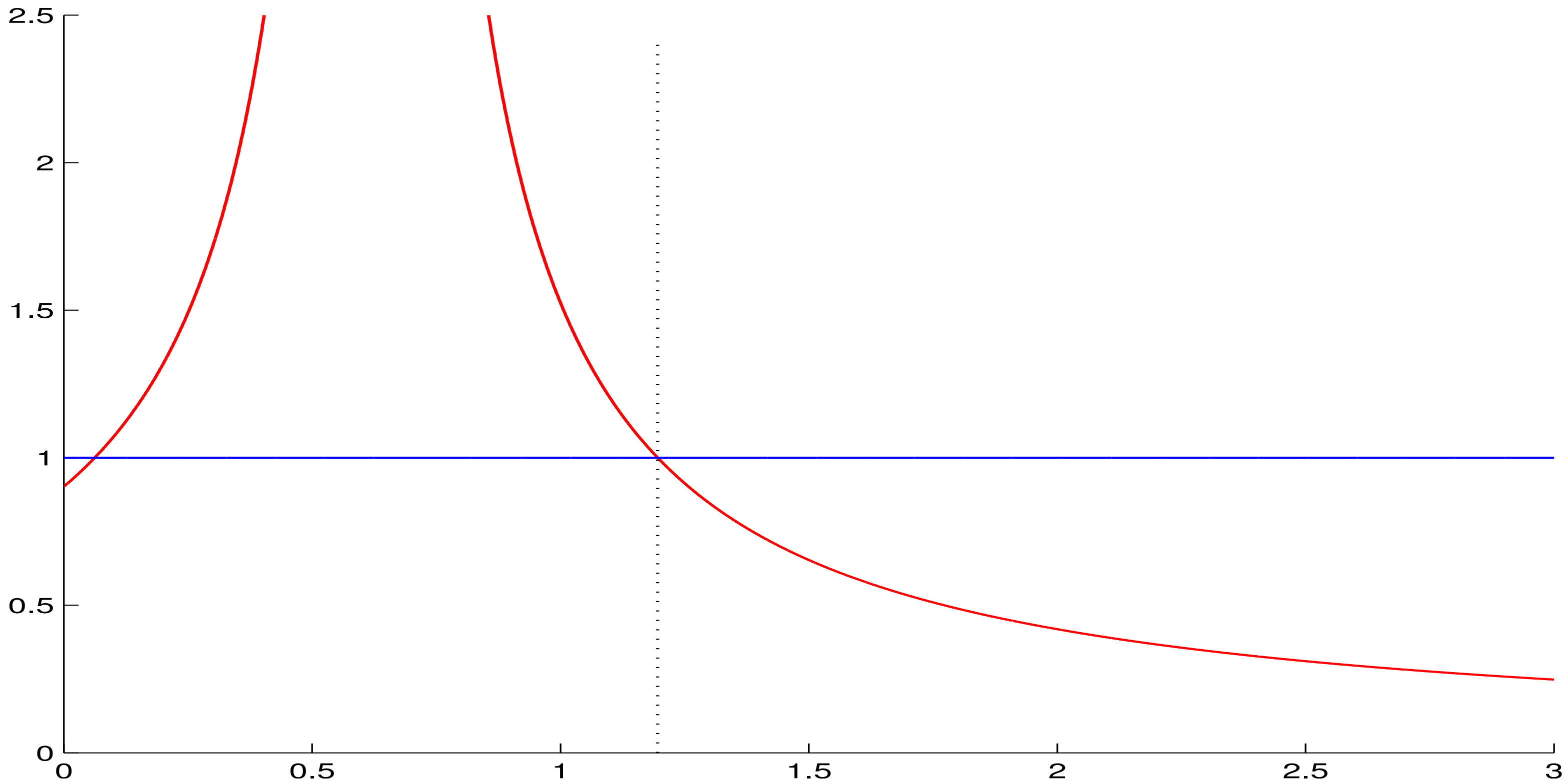
$$(T_k + \lambda I)x_k(\lambda) = \|b\|_2 e_1 \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$$



Iteration k = 2 for trust-region problem ($\tau(\lambda) = 1$)

How the secular equation evolves

$$(T_k + \lambda I)x_k(\lambda) = \|b\|_2 e_1 \text{ and } \|x_k(\lambda)\|_2 = \tau(\lambda)$$

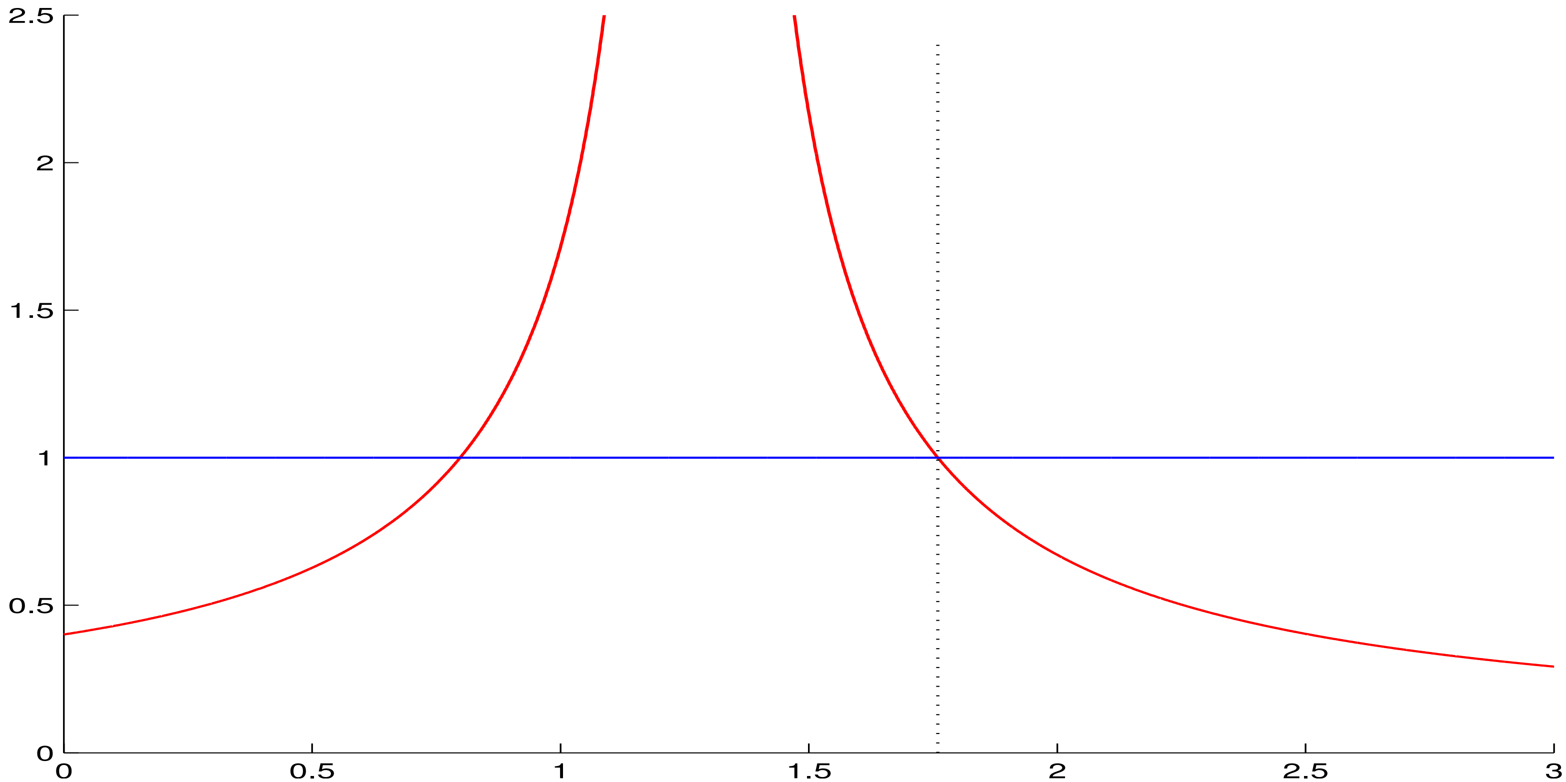


Iteration k = 3 for trust-region problem ($\tau(\lambda) = 1$)

Science & Technology Facilities Council

How the secular equation evolves

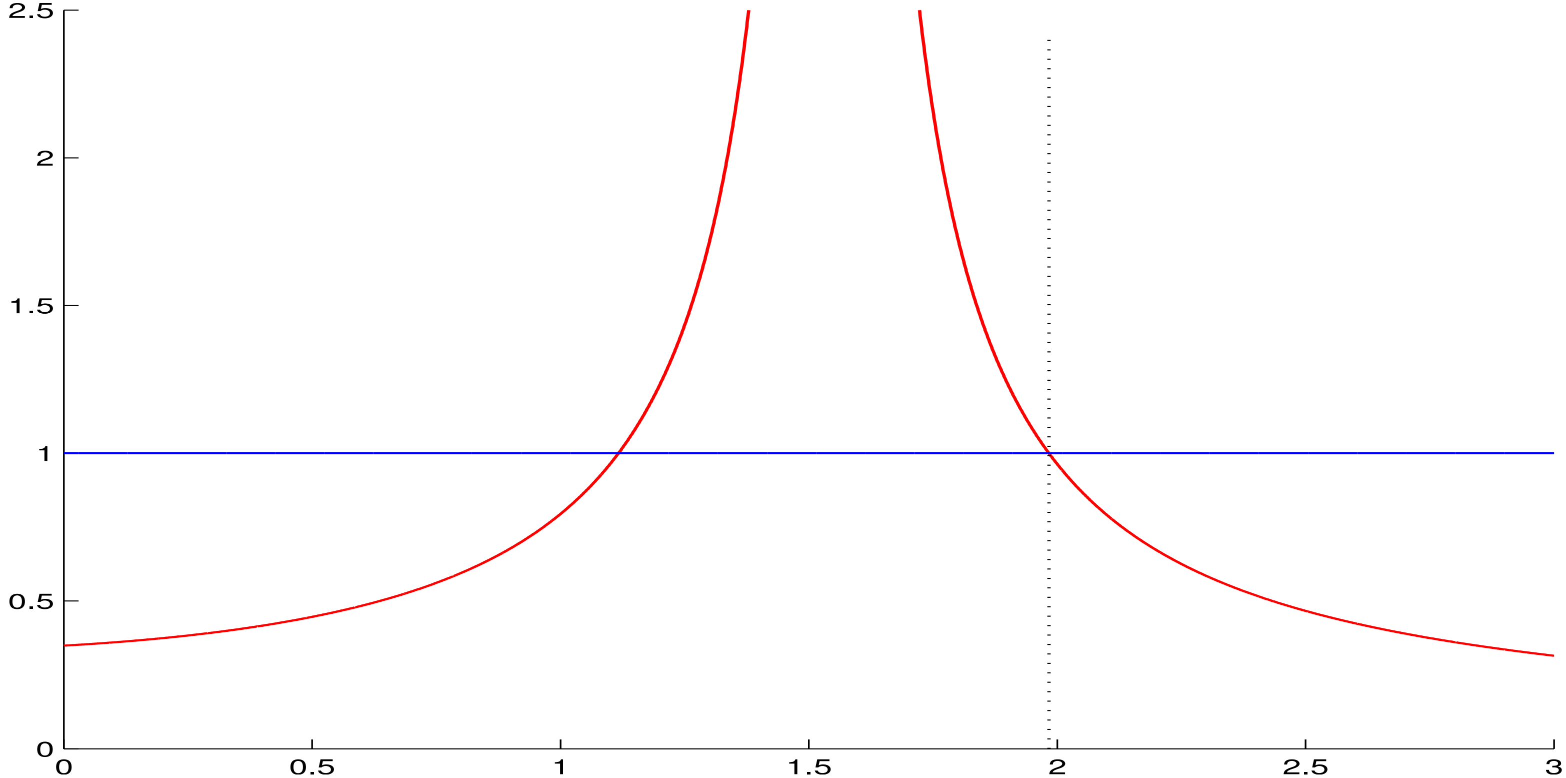
$$(T_k + \lambda I)x_k(\lambda) = \|b\|_2 e_1 \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$$



Iteration k = 4 for trust-region problem ($\tau(\lambda) = 1$)

How the secular equation evolves

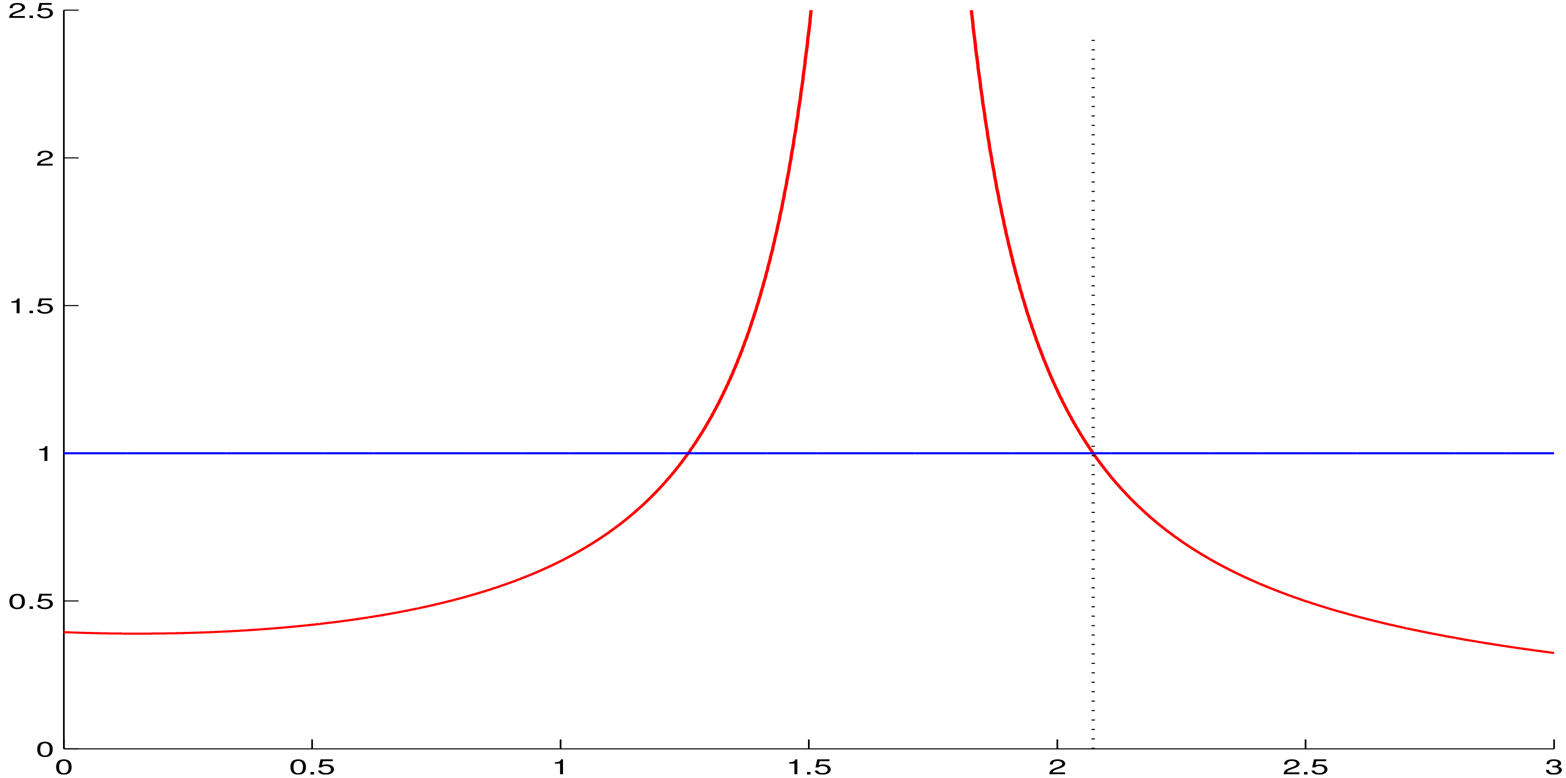
$$(T_k + \lambda I)x_k(\lambda) = \|b\|_2 e_1 \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$$



Iteration k = 5 for trust-region problem ($\tau(\lambda) = 1$)

How the secular equation evolves

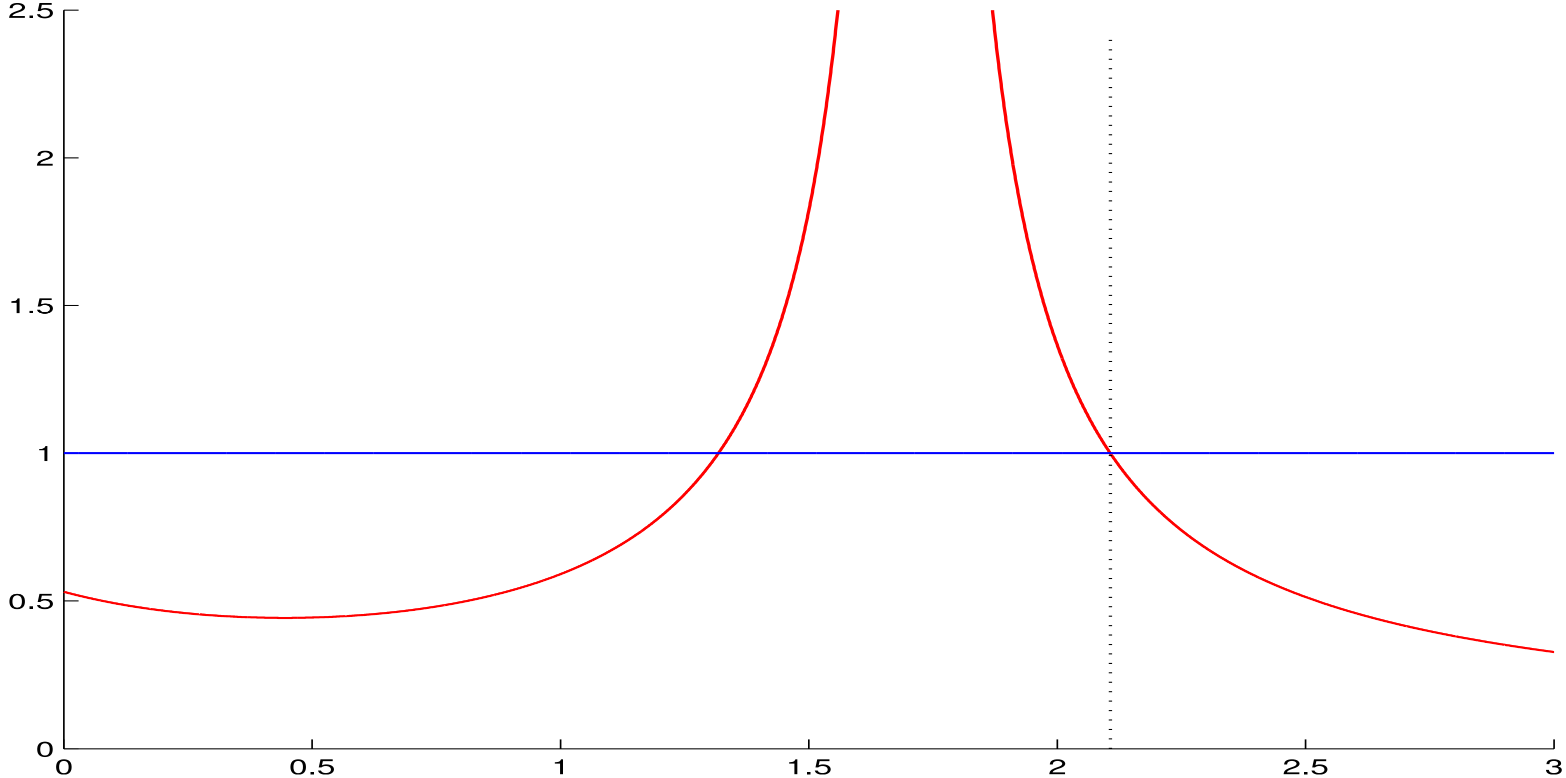
$$(T_k + \lambda I)x_k(\lambda) = \|b\|_2 e_1 \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$$



Iteration k = 6 for trust-region problem ($\tau(\lambda) = 1$)

How the secular equation evolves

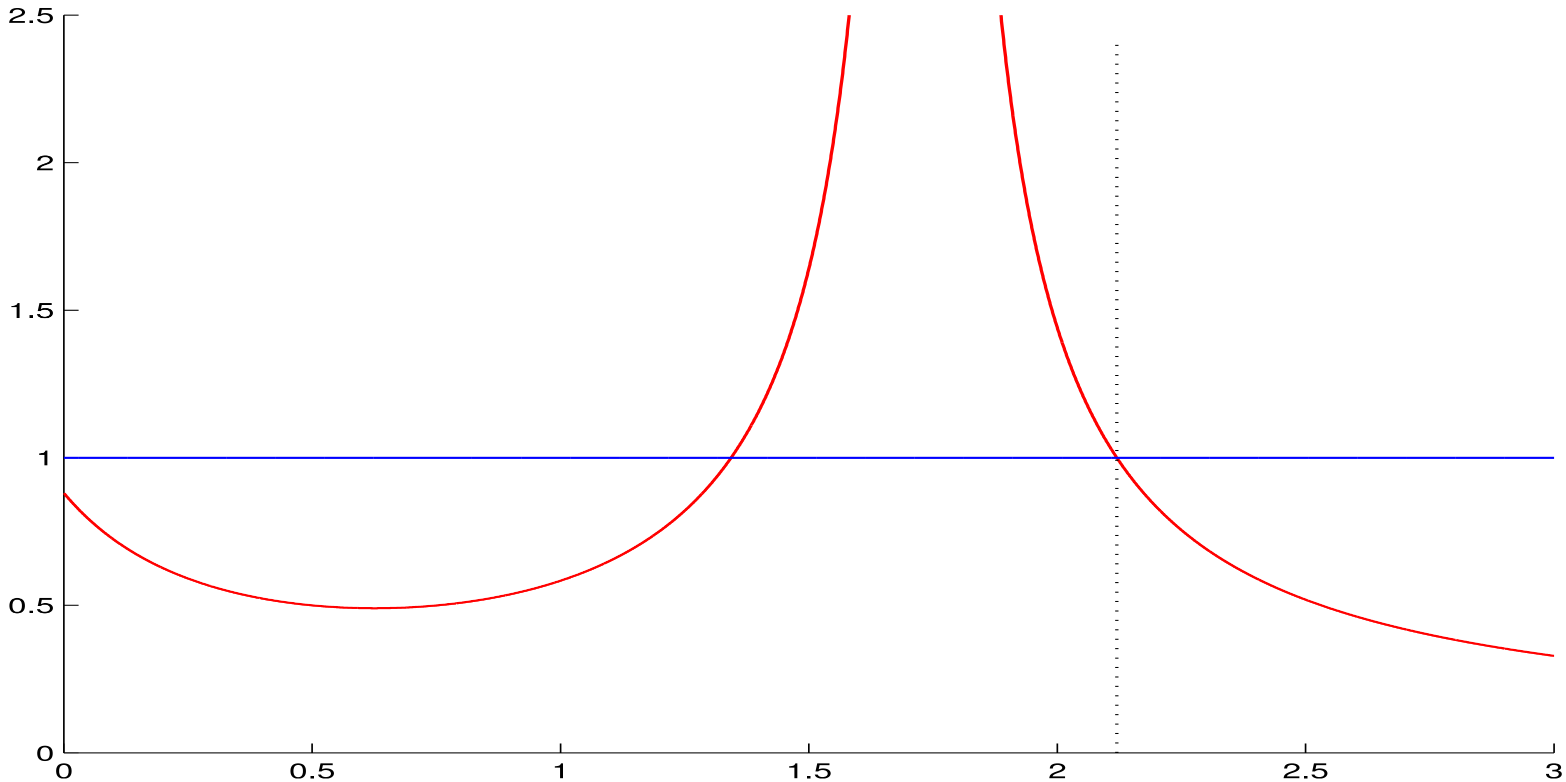
$$(T_k + \lambda I)x_k(\lambda) = \|b\|_2 e_1 \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$$



Iteration k = 7 for trust-region problem ($\tau(\lambda) = 1$)

How the secular equation evolves

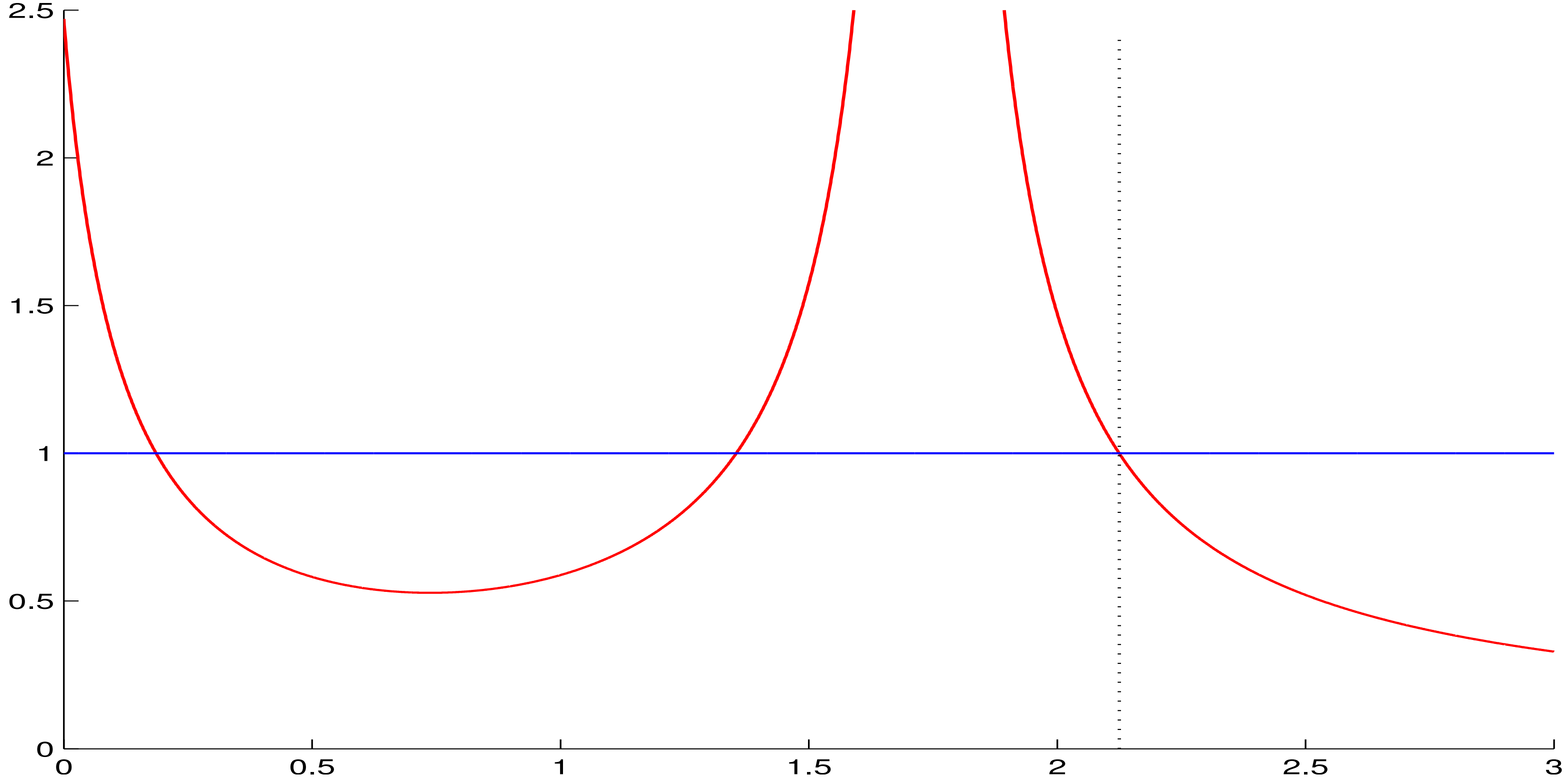
$$(T_k + \lambda I)x_k(\lambda) = \|b\|_2 e_1 \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$$



Iteration k = 8 for trust-region problem ($\tau(\lambda) = 1$)

How the secular equation evolves

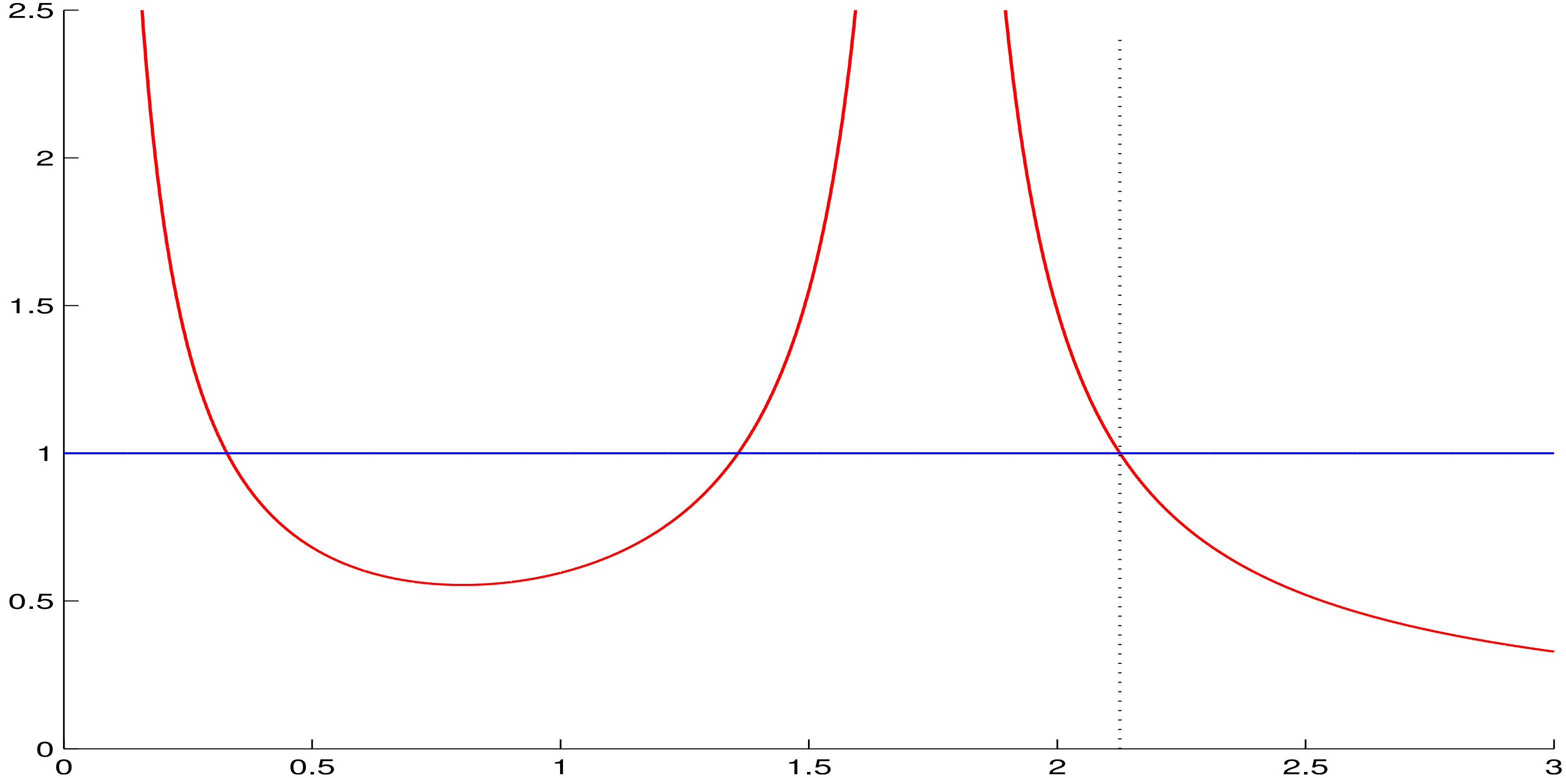
$$(T_k + \lambda I)x_k(\lambda) = \|b\|_2 e_1 \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$$



Iteration k = 9 for trust-region problem ($\tau(\lambda) = 1$)

How the secular equation evolves

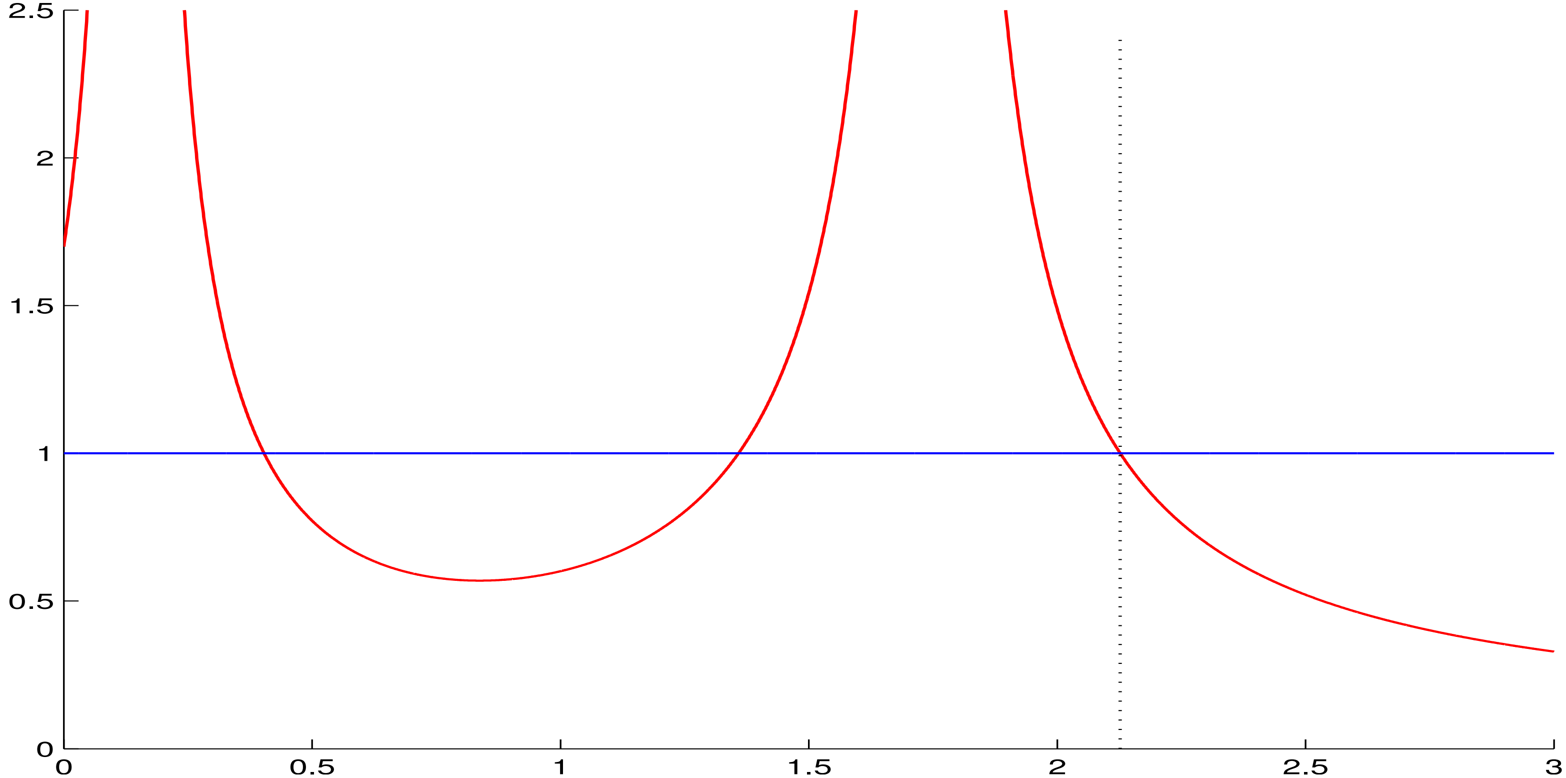
$$(T_k + \lambda I)x_k(\lambda) = \|b\|_2 e_1 \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$$



Iteration k = 10 for trust-region problem ($\tau(\lambda) = 1$)

How the secular equation evolves

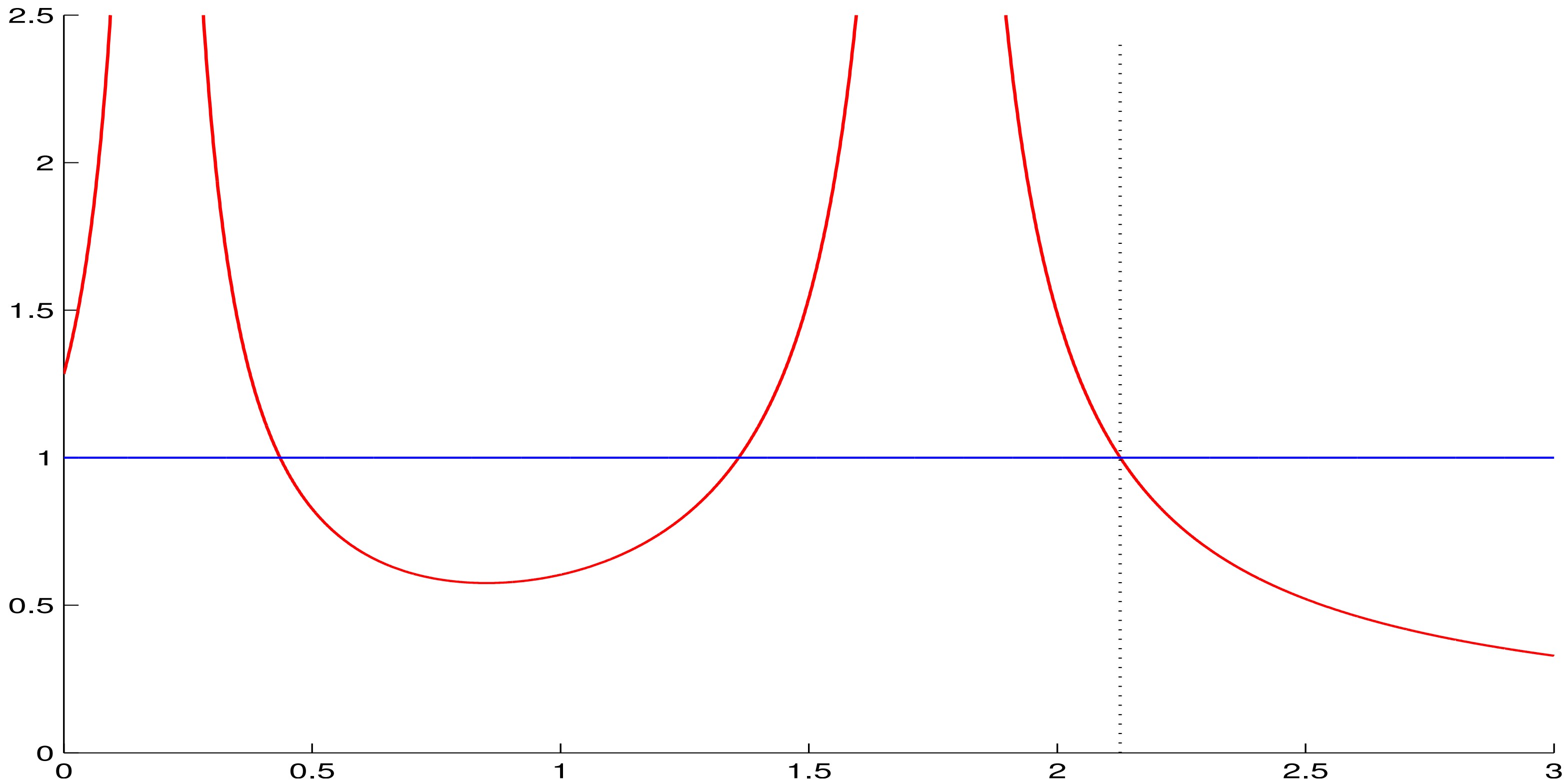
$$(T_k + \lambda I)x_k(\lambda) = \|b\|_2 e_1 \text{ and } \|x_k(\lambda)\|_2 = \tau(\lambda)$$



Iteration k = 11 for trust-region problem ($\tau(\lambda) = 1$)

How the secular equation evolves

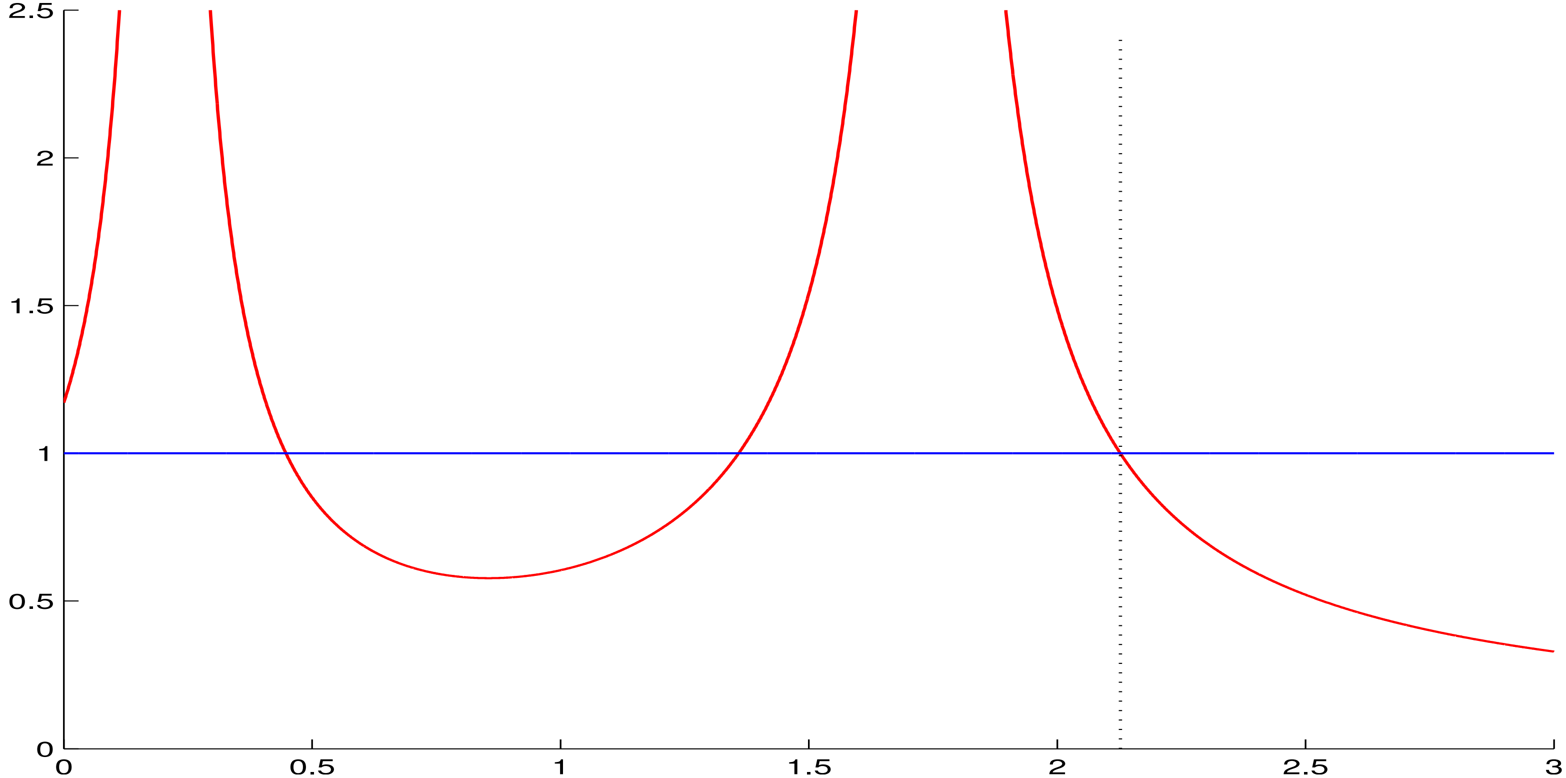
$$(T_k + \lambda I)x_k(\lambda) = \|b\|_2 e_1 \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$$



Iteration k = 12 for trust-region problem ($\tau(\lambda) = 1$)

How the secular equation evolves

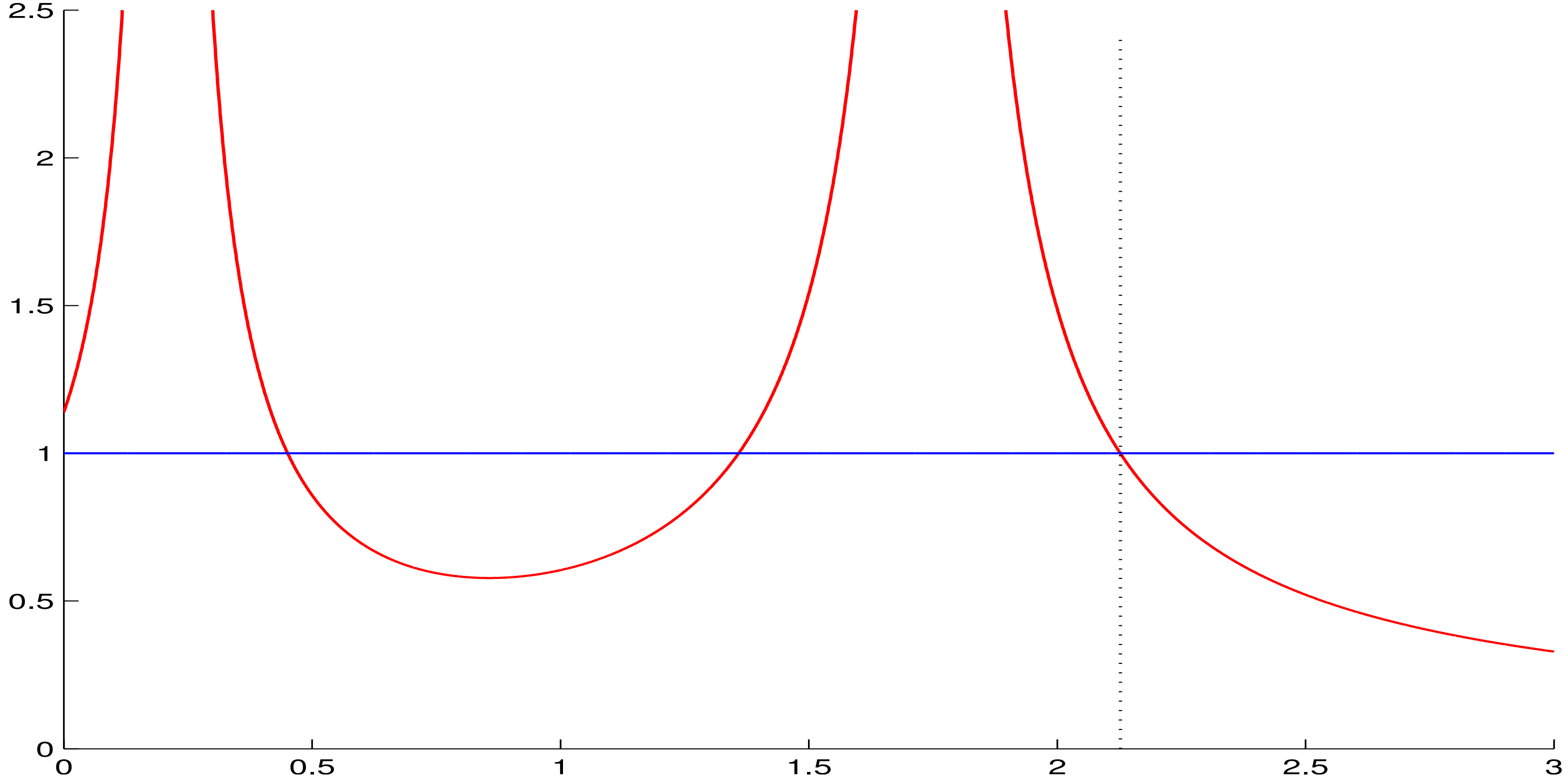
$$(T_k + \lambda I)x_k(\lambda) = \|b\|_2 e_1 \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$$



Iteration k = 13 for trust-region problem ($\tau(\lambda) = 1$)

How the secular equation evolves

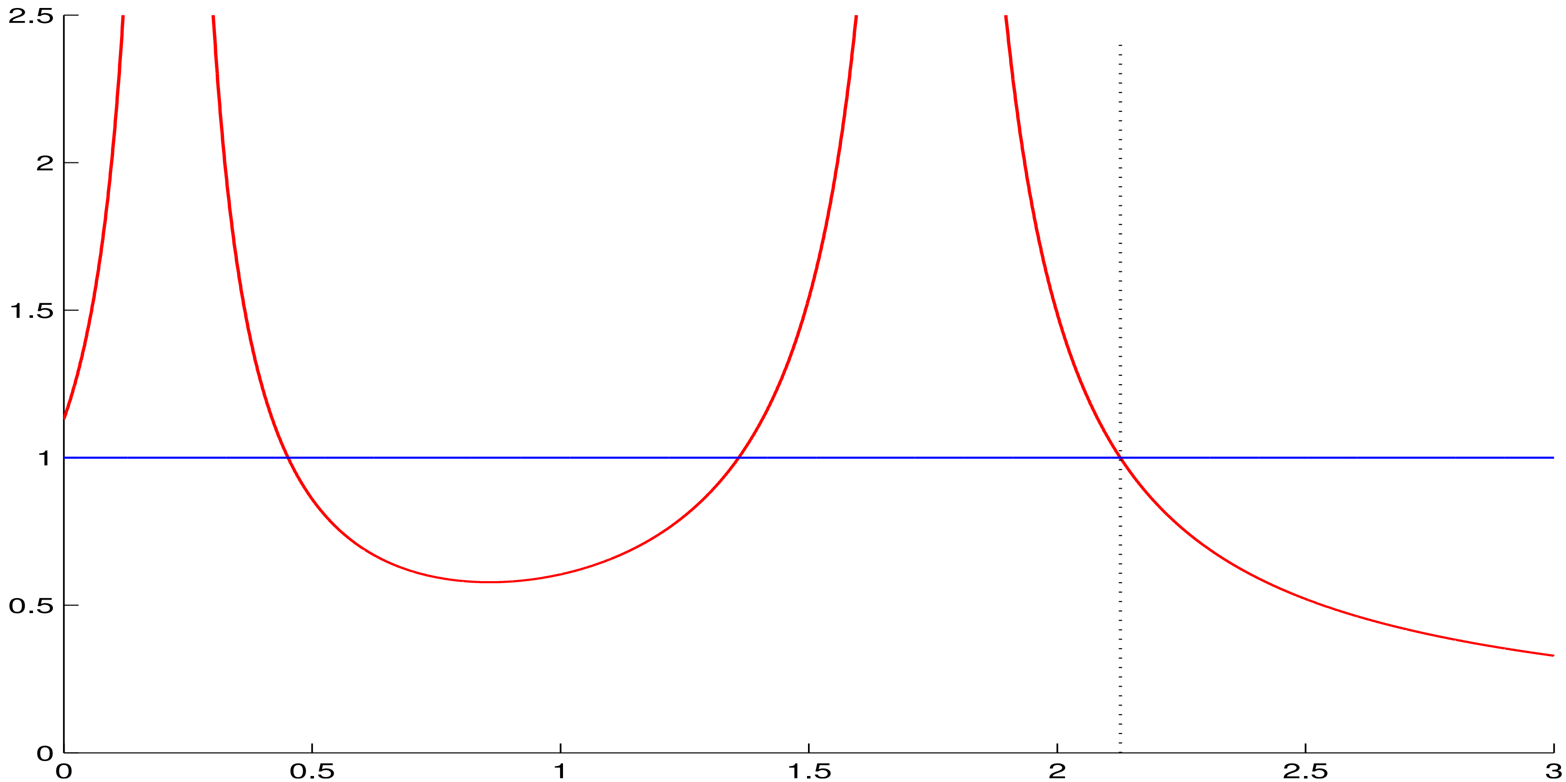
$$(T_k + \lambda I)x_k(\lambda) = \|b\|_2 e_1 \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$$



Iteration k = 14 for trust-region problem ($\tau(\lambda) = 1$)

How the secular equation evolves

$$(T_k + \lambda I)x_k(\lambda) = \|b\|_2 e_1 \quad \text{and} \quad \|x_k(\lambda)\|_2 = \tau(\lambda)$$



Iteration k = 15 for trust-region problem ($\tau(\lambda) = 1$)

Conclusions

- many important problems involve secular equations
- nonlinear transformation of variables and/or equations very useful
- high-order methods often pay off
- can be solved by either factorization or subspace iteration
- future advances to both large-scale linear equation and eigen solvers will be beneficial

Conclusions

- many important problems involve secular equations
- nonlinear transformation of variables and/or equations very useful
- high-order methods often pay off
- can be solved by either factorization or subspace iteration
- future advances to both large-scale linear equation and eigen solvers will be beneficial
- freely available optimization software as part of GALAHAD
  - `TRS`/`RQS` for direct solution of trust-region/regularization subproblems (G., Robinson, Thorne, 2011)
  - `GLTR`/`GLRT` for iterative solution of trust-region/regularization subproblems (G., Lucidi, Roma, Toint, 1999, Cartis, G., Toint, 2011)
  - `LSTR`/`LSRT`/`L2RT` for iterative solution of trust-region/regularization of least-squares subproblems (Cartis, G., Toint, 2010)