Numerical Methods for Ill-Posed Problems II

Lothar Reichel

Summer School on Applied Analysis
TU Chemnitz, Oct. 4-8, 2010

Outline of Lecture 2:

- Review of Tikhonov regularization
- Orthogonal polynomials and Gauss quadrature
- Matrix functionals, Lanczos tridiagonalization, and orthogonal polynomials
- Tikhonov regularization of large-scale problems using the L- and curvature-ribbons

Tikhonov regularization

Solve the penalized least-squares problem

$$\min_x \{\|Ax - b\|^2 + \mu\|x\|^2\}, \qquad (1)$$

where $\mu > 0$ is a regularization parameter. It is important to determine a suitable value of μ.

The normal equations

$$(A^T A + \mu I)x = A^T b$$

show that (1) has the solution

$$x_\mu := (A^T A + \mu I)^{-1} A^T b, \qquad \mu > 0.$$

The L-curve is an aid for determining a suitable value of $\mu > 0$. It is the graph

$$(\log_{10} \|x_\mu\|, \log_{10} \|b - Ax_\mu\|), \quad \mu > 0$$

Choose the value of μ, denoted μ_L, that corresponds to the *vertex* of the L-curve.

How can one determine μ_L efficiently for large-scale problems?

Orthogonal polynomials and Gauss quadrature

Let $d\nu$ be a positive measure with support on the real axis such that all moments

$$\int_{-\infty}^{\infty} t^j d\nu(t), \qquad j = 0, 1, 2, \dots \ ,$$

exist. Define the inner product

$$(\phi, \psi) := \int_{-\infty}^{\infty} \phi(t)\psi(t)d\nu(t)$$

and the associated norm

$$\|\phi\| := (\phi, \phi)^{1/2}.$$

Some properties of orthogonal polynomials

Let $\pi_0, \pi_1, \pi_2, \ldots$ be a family of monic orthogonal polynomials associated with $d\nu$. Thus,

$$(\pi_j, \pi_k) = 0, \qquad j \neq k,$$

and $(\pi_j, \pi_j) > 0$.

The π_j satisfy a 3-term recurrence relation: For $j = 0, 1, 2, \dots$,

$$\pi_{j+1}(t) = (t - \delta_{j+1})\pi_j(t) - \gamma_{j+1}^2 \pi_{j-1}(t),$$

where $\pi_0(t) := 1$, $\pi_{-1}(t) := 0$, and

$$\begin{aligned} \delta_{j+1} &= (t\pi_j, \pi_j)/(\pi_j, \pi_j), \qquad j \geq 0, \\ \gamma_{j+1}^2 &= \begin{cases} 0, & j = 0, \\ (\pi_j.\pi_j)/(\pi_{j-1}, \pi_{j-1}), & j \geq 1. \end{cases} \end{aligned}$$

Proof by induction.

The zeros of each π_j are simple and real, and live in the convex hull of the support of $d\nu$.

Proof by contradiction.

An ℓ-point quadrature rule

$$G_\ell(\phi) = \sum_{j=1}^{\ell} \phi(\theta_j) w_j$$

is said to be a Gauss rule with respect to the measure $d\nu$ if

$$\int_{-\infty}^{\infty} p(t) d\nu(t) = G_\ell(p), \quad \forall p \in \Pi_{2\ell-1}.$$

G_ℓ has the following properties:

- The nodes $\theta_1, \theta_2, \ldots, \theta_\ell$ are the zeros of π_ℓ.
- The weights $w_1, w_2, \ldots, w_\ell$ are positive and uniquely determined.

- Let ϕ have 2ℓ continuous derivatives on the convex hull of the support of $d\nu$. Then the quadrature error can be expressed as

$$E_{G_\ell}(\phi) := \int_{-\infty}^{\infty} \phi(t)d\nu(t) - G_\ell(\phi)$$
$$= \frac{\phi^{(2\ell)}(\xi)}{(2\ell)!} \int_{-\infty}^{\infty} \prod_{j=1}^{\ell} (t - \theta_j)^2 d\nu(t),$$

where ξ lives in the convex hull of supp(dν),

$$\phi^{(2\ell)} \geq 0 \Longrightarrow G_\ell(\phi) \leq \int_{-\infty}^{\infty} \phi(t)d\nu(t).$$

The nodes θ_j are the eigenvalues and the w_j are the square of the first component of normalized eigenvectors of the symmetric tridiagonal matrix

$$T_\ell = \begin{bmatrix} \delta_1 & \gamma_2 & & & \\ \gamma_2 & \delta_2 & \gamma_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \gamma_{\ell-1} & \delta_{\ell-1} & \gamma_\ell \\ & & & \gamma_\ell & \delta_\ell \end{bmatrix}.$$

Compute all θ_j and w_j in $O(\ell^2)$ flops by the QR algorithm.

Assume that supp($d\nu$) lives on the nonnegative real axis. The $(\ell+1)$-point Gauss-Radau rule with respect to $d\nu$ and a fixed node $\tilde{\theta}_0 := 0$,

$$R_{\ell+1}(\phi) = \sum_{j=0}^{\ell} \phi(\tilde{\theta}_j)\tilde{w}_j,$$

satisfies

$$\int_{-\infty}^{\infty} p(t)d\nu(t) = R_{\ell+1}(p), \quad \forall p \in \Pi_{2\ell}.$$

$R_{\ell+1}$ has the following properties:

- Let ϕ have $2\ell + 1$ continuous derivatives on the convex hull of the support of $d\nu$. Then the quadrature error can be expressed as

$$E_{R_{\ell+1}}(\phi) := \int_{-\infty}^{\infty} \phi(t) d\nu(t) - R_{\ell+1}(\phi)$$
$$= \frac{\phi^{(2\ell+1)}(\xi)}{(2\ell+1)!} \int_{-\infty}^{\infty} t \prod_{j=1}^{\ell} (t - \tilde{\theta}_j)^2 d\nu(t),$$

where ξ lives in the convex hull of supp$(d\nu)$. In particular,

$$\phi^{(2\ell+1)} \leq 0 \Longrightarrow R_{\ell+1}(\phi) \geq \int_{-\infty}^{\infty} \phi(t) d\nu(t).$$

- The nodes $\tilde{\theta}_0, \tilde{\theta}_1, \ldots, \tilde{\theta}_\ell$ are the eigenvalues of

$$\tilde{T}_{\ell+1} = \begin{bmatrix} \delta_1 & \gamma_2 & & & \\ \gamma_2 & \delta_2 & \gamma_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \gamma_\ell & \delta_\ell & \gamma_{\ell+1} \\ & & & \gamma_{\ell+1} & \tilde{\delta}_{\ell+1} \end{bmatrix},$$

  where $\tilde{\delta}_{\ell+1} := \gamma_{\ell+1}^2 e_\ell^T T_\ell^{-1} e_\ell$.

- The weights $\tilde{w}_0, \tilde{w}_1, \ldots, \tilde{w}_\ell$ are the square of the first components of normalized eigenvectors of $\tilde{T}_{\ell+1}$.

Matrix functionals, Lanczos tridiagonalization, and orthogonal polynomials

We consider the computation of upper and lower bounds for matrix functionals

$$\Phi(A) := v^T \phi(A) v,$$

where $A \in \mathrm{R}^{n \times n}$ is a large symmetric matrix and v is a given unit vector.

For Tikhonov regularization

$$\phi(t) = (t + \mu)^{-2}.$$

Spectral factorization

$$A = S\Lambda S^T, \quad \Lambda = \mathrm{diag}[\lambda_1, \lambda_2, \ldots, \lambda_n],$$

with S orthogonal. Define

$$[\nu_1, \nu_2, \ldots, \nu_n] := v^T S$$

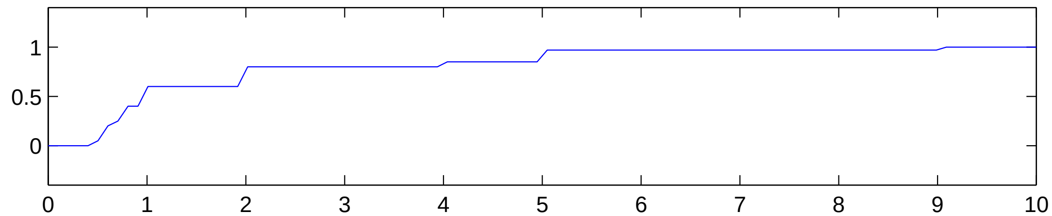
Then

$$\Phi(A) = v^T S \phi(\Lambda) S^T v = \sum_{j=1}^{n} \phi(\lambda_j) \nu_j^2.$$

The right-hand side is a Stieltjes integral

$$\mathcal{I}(\phi) := \int_{-\infty}^{\infty} \phi(t) d\nu(t).$$

The nonnegative measure $d\nu(t)$ is such that $\nu(t)$ is a nondecreasing step function:



Define the inner product

$$(\phi, \psi) := \mathcal{I}(\phi\psi).$$

Gauss quadrature rules associated with $d\nu(t)$ can be computed by the Lanczos
process applied to the matrix A with initial vector $v_1 := v$.

ℓ steps of the Lanczos process yield

$$AV_\ell = V_\ell T_\ell + f_\ell e_\ell^T, \qquad (1)$$

where $V_\ell = [v_1, v_2, \ldots, v_\ell]$ satisfies

$$V_\ell^T V_\ell = I_\ell, \qquad V_\ell^T f_\ell = 0$$

and

$$T_\ell = \begin{bmatrix} \alpha_1 & \beta_1 & & & \\ \beta_1 & \alpha_2 & \beta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \beta_{\ell-2} & \alpha_{\ell-1} & \beta_{\ell-1} \\ & & & \beta_{\ell-1} & \alpha_\ell \end{bmatrix}.$$

Computation of a partial Lanczos decomposition

Lanczos tridiagonalization

$\beta_0 = \|b\|; \quad \tilde{v} = b;$

for $j = 1, 2, \ldots, \ell$ **do**

$\quad v_j = \tilde{v}/\beta_{j-1};$

$\quad \tilde{v} = A v_j;$

$\quad$ if $j > 1$, $\tilde{v} = \tilde{v} - \beta_{j-1} v_{j-1};$

$\quad \alpha_j = \tilde{v}^T v_j; \;\; \tilde{v} = \tilde{v} - \alpha_j v_j;$

$\quad \beta_j = \|\tilde{v}\|;$

end

$f_\ell = \tilde{v};$

It follows from the algorithm or (1) that

$$v_j = p_{j-1}(A)v_1, \qquad 1 \le j \le \ell,$$

and by the orthogonality of the v_j, that

$$\begin{aligned}
(p_{j-1}, p_{k-1}) &= \int_{-\infty}^{\infty} p_{j-1}(t)p_{k-1}(t)d\nu(t) \\
&= v^T S p_{j-1}(\Lambda) p_{k-1}(\Lambda) S^T v \\
&= v^T p_{j-1}(A) p_{k-1}(A) v \\
&= v_j^T v_k = \begin{cases} 0, & j \neq k, \\ 1, & j = k. \end{cases}
\end{aligned}$$

3-term recurrence relation, $j = 1, 2, \ldots$:

$$\beta_j p_j(t) = (t - \alpha_j) p_{j-1}(t) - \beta_{j-1} p_{j-2}(t),$$

where

$$\beta_1 p_1(t) = (t - \alpha_1) p_0(t), \qquad p_0(t) = 1.$$

Using the spectral decomposition of T_m one can show

$$G_\ell(\phi) = e_1^T \phi(T_\ell) e_1.$$

Matrix functionals for the L-ribbon

Let $A \in \mathrm{R}^{m \times n}$, $m \geq n \gg 1$, and let

$$P(\mu) := (\|x_\mu\|, \|Ax_\mu - b\|)$$

be a point on the L-curve. Then

$$\|x_\mu\|^2 = x_\mu^T x_\mu = b^T A(A^T A + \mu I)^{-2} A^T b$$

and

$$\|Ax_\mu - b\|^2 = \|(A(A^T A + \mu I)^{-1} A^T b - b\|^2$$

Using the identity

$$I - A(A^TA + \mu I)^{-1}A^T = \mu(AA^T + \mu I)^{-1}$$

yields

$$\|Ax_\mu - b\|^2 = \mu^2 b^T (AA^T + \mu I)^{-2} b.$$

Let

$$\phi(t) = (t + \mu)^{-2}$$

and define the matrix functionals

$$\begin{aligned} s &= \mu^2 b^T \phi(AA^T) b = \|Ax_\mu - b\|^2, \\ \hat{s} &= (A^T b)^T \phi(A^T A) A^T b = \|x_\mu\|^2. \end{aligned}$$

Spectral factorizations:

$$AA^T = W\Lambda W^T, \qquad A^T A = \hat{W}\hat{\Lambda}\hat{W}^T,$$

where

$$\Lambda = \mathrm{diag}[\lambda_1, \lambda_2, \ldots, \lambda_n, 0, \ldots, 0] \in \mathrm{R}^{m\times m},$$

$$\hat{\Lambda} = \mathrm{diag}[\lambda_1, \lambda_2, \ldots, \lambda_n];$$

$$W^T W = I_m; \qquad \hat{W}^T\hat{W} = I_n.$$

Let

$$h = [h_1, h_2, \ldots, h_m]^T = \mu W^T b,$$

$$\hat{h} = [\hat{h}_1, \hat{h}_2, \ldots, \hat{h}_n]^T = \hat{W}^T A^T b.$$

Then

$$\begin{aligned} s &= h^T\phi(\Lambda)h = \sum_{k=1}^{n} \phi(\lambda_k)h_k^2 + \phi(0) \sum_{k=n+1}^{m} h_k^2 \\ &= \int_{-\infty}^{\infty} \phi(t)d\omega(t), \end{aligned}$$

$$\hat{s} = \hat{h}^T\phi(\hat{\Lambda})\hat{h} = \sum_{k=1}^{n} \phi(\lambda_k)\hat{h}_k^2 = \int_{-\infty}^{\infty} \phi(t)d\hat{\omega}(t).$$

The distribution functions $\omega(t)$ and $\hat{\omega}(t)$ are non-decreasing step functions with jumps at the eigenvalues λ_k and at the origin.

The computation of s and $\hat{s}$ for many different values of the parameter μ is not feasible when A is very large.

However, upper and lower bounds for s and $\hat{s}$ can be computed efficiently by Gauss and Gauss-Radau rules.

Define inner product induced by $d\omega(t)$:

$$\begin{aligned}
(f, g) &= \int_{-\infty}^{\infty} f(t) g(t) d\omega(t) \\
&= \sum_{k=1}^{n} f(\lambda_k) g(\lambda_k) h_k^2 + f(0) g(0) \sum_{k=n+1}^{m} h_k^2 \\
&= h^T f(\Lambda) g(\Lambda) h,
\end{aligned}$$

ℓ steps of the Lanczos process applied to the matrix AA^T with initial vector b yields

$$T_\ell = \begin{bmatrix} \alpha_1 & \beta_1 & & & \\ \beta_1 & \alpha_2 & \beta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \beta_{\ell-2} & \alpha_{\ell-1} & \beta_{\ell-1} \\ & & & \beta_{\ell-1} & \alpha_\ell \end{bmatrix}.$$

Assume that ℓ is small enough that $\beta_j > 0, \ 1 \leq j < \ell$.

Symmetric tridiagonal matrices with positive subdiagonal entries are called **Jacobi matrices**.

Introduce the Cholesky factor of T_ℓ, i.e., $T_\ell = C_\ell C_\ell^T$, where

$$C_\ell = \begin{bmatrix} \rho_1 & & & & \\ \sigma_2 & \rho_2 & & & \\ & \ddots & \ddots & & \\ & & \sigma_{\ell-1} & \rho_{\ell-1} & \\ & & & \sigma_\ell & \rho_\ell \end{bmatrix} \in \mathrm{R}^{\ell\times\ell}.$$

Also define

$$\bar{C}_\ell = \begin{bmatrix} C_\ell \\ \sigma_{\ell+1} e_\ell^T \end{bmatrix} \in \mathrm{R}^{(\ell+1)\times\ell}.$$

The ℓ-point Gauss quadrature rule associated with the measure $d\omega$ is given by

$$\begin{aligned} G_\ell(\phi) &= \mu^2 \|b\|^2 e_1^T \phi(T_\ell) e_1 \\ &= \mu^2 \|b\|^2 e_1^T \phi(C_\ell C_\ell^T) e_1. \end{aligned}$$

The corresponding $(\ell + 1)$-point Gauss-Radau rule with one node at the origin can be expressed as

$$R_{\ell+1}(\phi) = \mu^2 \|b\|^2 e_1^T \phi(\bar{C}_\ell \bar{C}_\ell^T) e_1.$$

Assume that $\mu > 0$ and let

$$\phi(t) = (t + \mu)^{-2}.$$

The support of $d\omega$ is on the nonnegative real axis, and

$$\phi^{(2\ell)} > 0, \qquad \phi^{(2\ell+1)} < 0$$

on the support. Hence,

$$E_{G_\ell}(\phi) > 0, \qquad E_{R_{\ell+1}}(\phi) < 0.$$

Now consider $\hat{s}$:

Let $\hat{T}_\ell$ be the tridiagonal matrix obtained by ℓ steps of the Lanczos process applied to A^TA with initial vector A^Tb, and let $\hat{C}_\ell$ be the Cholesky factor of $\hat{T}_\ell$, i.e.,

$$\hat{T}_\ell = \hat{C}_\ell \hat{C}_\ell^T.$$

The ℓ-point quadrature rule with respect to $d\hat{\omega}$ can be written as

$$\hat{G}_\ell(\phi) = \|A^Tb\|^2 e_1^T \phi(\hat{T}_\ell) e_1 = \|A^Tb\|^2 e_1^T \phi(\hat{C}_\ell \hat{C}_\ell^T) e_1.$$

The ℓ-point Gauss-Radau rule with one node assigned at the origin can be expressed as

$$\hat{R}_\ell(\phi) = \|A^T b\|^2 e_1^T \phi(\bar{\hat{C}}_{\ell-1}(\bar{\hat{C}}_{\ell-1})^T)e_1,$$

where $\bar{\hat{C}}_{\ell-1}$ consists of the $\ell - 1$ first columns of $\hat{C}_\ell$.

Moreover,

$$E_{\hat{G}_\ell}(\phi) > 0, \qquad E_{\hat{R}_\ell}(\phi) < 0.$$

Computation of C_ℓ without forming T_ℓ

Lanczos bidiagonalization

$\sigma_1 = \|b\|; \quad u_1 = b/\sigma_1; \quad \tilde{v}_1 = A^T u_1; \quad \rho_1 = \|\tilde{v}_1\|;$

for $j = 2, 3, \ldots, \ell$ **do**

$\quad \tilde{u}_j = A v_{j-1} - \rho_{j-1} u_{j-1}; \quad \sigma_j = \|\tilde{u}_j\|; \quad u_j = \tilde{u}_j / \sigma_j;$

$\quad \tilde{v}_j = A^T u_j - \sigma_j v_{j-1}; \quad \rho_j = \|\tilde{v}_j\|; \quad v_j = \tilde{v}_j / \rho_j;$

end

$\tilde{u}_{\ell+1} = A v_\ell - \rho_\ell u_\ell; \quad \sigma_{\ell+1} = \|\tilde{u}_{\ell+1}\|; \quad u_{\ell+1} = \tilde{u}_{\ell+1} / \sigma_{\ell+1};$

The matrices

$$U_\ell = [u_1, u_2, \ldots, u_\ell], \quad V_\ell = [v_1, v_2, \ldots, v_\ell]$$

have orthonormal columns.

It follows from the algorithm that

$$AV_\ell = U_\ell C_\ell + \sigma_{\ell+1} u_{\ell+1} e_\ell^T,$$

$$A^T U_\ell = V_\ell C_\ell^T, \qquad b = \sigma_1 U_\ell e_1.$$

Therefore

$$AA^T U_\ell = U_\ell C_\ell C_\ell^T + \sigma_{\ell+1} \rho_\ell u_{\ell+1} e_\ell^T.$$

Hence, the $\{u_j\}$ are the Lanczos vectors and

$$T_\ell = C_\ell C_\ell^T$$

is the matrix obtained after ℓ steps of the Lanczos process applied to AA^T with initial vector b.

The $\{v_j\}$ are the Lanczos vectors determined by ℓ steps of the Lanczos process applied to A^TA with initial vector A^Tb. Also

$$\hat{T}_\ell = \bar{C}_\ell^T \bar{C}_\ell.$$

The Cholesky factor $\hat{C}_\ell$ of $\hat{T}_\ell$ can be determined from the QR factorization

$$\bar{C}_\ell = \bar{Q}_\ell \hat{C}_\ell^T.$$

Since $\bar{C}_\ell$ is bidiagonal only $\mathcal{O}(\ell)$ flops are needed to compute $\hat{C}_\ell$ from $\bar{C}_\ell$. (Use Givens rotations.)

The L-curve and L-ribbon

Recall that the point $P(\mu)$ on the L-curve has coordinates

$$P(\mu) = (\|x_\mu\|, \|Ax_\mu - b\|) = (\hat{s}^{1/2}, s^{1/2}).$$

After ℓ Lanczos bidiagonalization steps, we can evaluate the bounds

$$G_\ell(\phi) \leq s \leq R_{\ell+1}(\phi),$$

where

$$G_\ell(\phi) = \mu^2 \|b\|^2 e_1^T (C_\ell C_\ell^T + \mu I_\ell)^{-2} e_1$$

and

$$R_{\ell+1}(\phi) = \mu^2 \|b\|^2 e_1^T (\bar{C}_\ell \bar{C}_\ell^T + \mu I_{\ell+1})^{-2} e_1.$$

To compute $G_\ell(\phi)$, first determine the vector $y = (C_\ell C_\ell^T + \mu I)^{-1} e_1$ as the solution of

$$\min_{y \in \mathrm{R}} \left\| \begin{bmatrix} C_\ell^T \\ \mu^{1/2} I_\ell \end{bmatrix} y - \mu^{-1/2} \begin{bmatrix} 0 \\ e_1 \end{bmatrix} \right\|$$

Note: For each value of μ, evaluation of the Gauss rule requires only $\mathcal{O}(\ell)$ flops.

Similar results hold for the Gauss-Radau rule.

Bounds for $\|x_\mu\|^2$ are given by

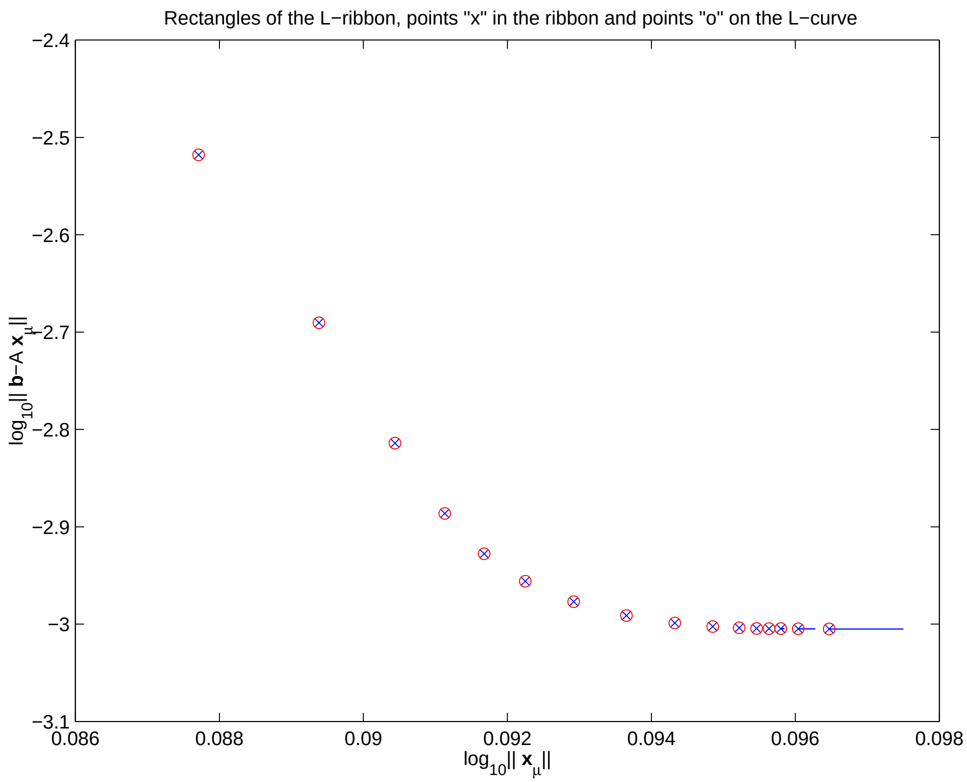
$$\hat{G}_\ell(\phi) \leq \hat{s} \leq \hat{R}_\ell(\phi).$$

Introduce the quantities

$$\begin{aligned} y^-(\mu) &:= (G_\ell(\phi))^{1/2}, \\ y^+(\mu) &:= (R_{\ell+1}(\phi))^{1/2}, \\ x^-(\mu) &:= (\hat{G}_\ell(\phi))^{1/2}, \\ x^+(\mu) &:= (\hat{R}_\ell(\phi))^{1/2}, \end{aligned}$$

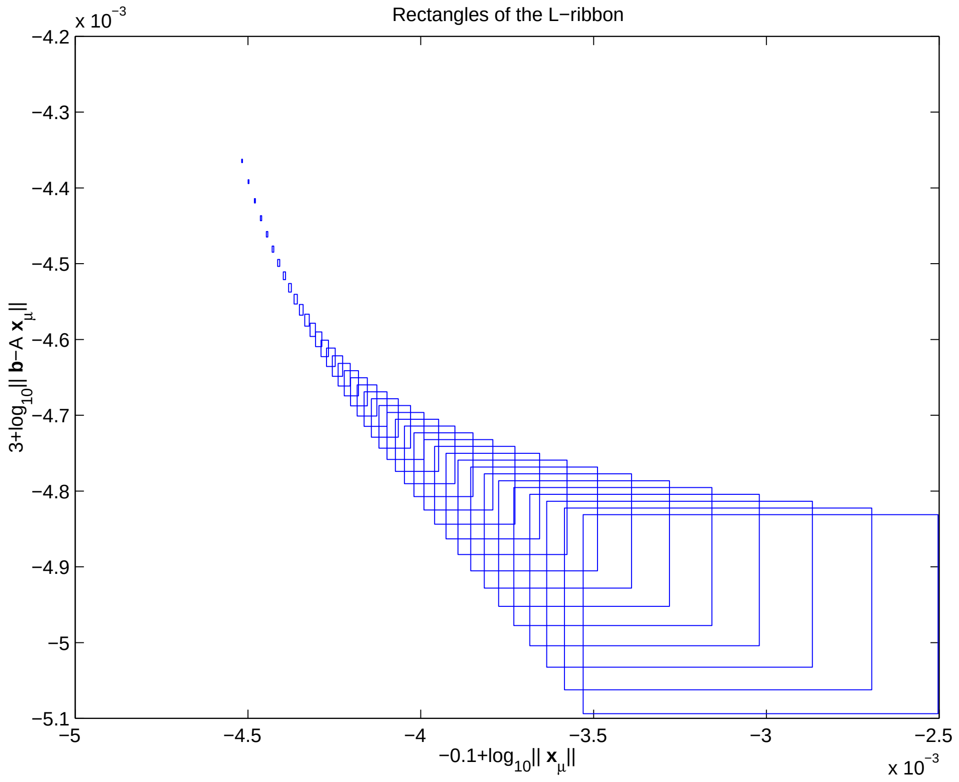
and define the L-ribbon as the union of the rectangular regions

$$\bigcup_{\mu>0} \{\{x(\mu), y(\mu)\} : x^-(\mu) \le x(\mu) \le x^+(\mu), y^-(\mu) \le y(\mu) \le y^+(\mu)\}.$$



Rectangles of the L−ribbon, points "x" in the ribbon and points "o" on the L−curve
−2.4
−2.5
−2.6
−2.7
−2.8
−2.9
−3
−3.1
0.086
0.088
0.09
0.092
0.094
0.096
0.098
log10|| b−A xμ||
log10|| xμ||

Blow-up


Rectangles of the L−ribbon
x 10^−3
−4.2
−4.3
−4.4
−4.5
−4.6
−4.7
−4.8
−4.9
−5
−5.1
3+log10|| b−A xμ||
−5
−4.5
−4
−3.5
−3
−2.5
−0.1+log10|| xμ||
x 10^−3

Computing the regularized solution

After determining μ_L, the desired approximate solution $x_{\mu_L,\ell}$ is of the form

$$x_{\mu_L,\ell} = V_\ell y_{\mu_L,\ell}$$

determined by the Galerkin equation

$$V_\ell^T(A^TA + \mu_L I_n)V_\ell y_{\mu_L,\ell} = V_\ell^T A^T b$$

where

$$(\bar{C}_\ell^T \bar{C}_\ell + \mu_L I_\ell) y_{\mu,\ell} = \sigma_1 C_\ell^T e_1.$$

Note:

$$\|x_{\mu_L,\ell}\| = (\hat{G}_\ell(\phi))^{1/2}$$

and

$$\|b - Ax_{\mu_L,\ell}\| = (R_{\ell+1}(\phi))^{1/2}.$$

Th curvature ribbon

To determine μ_L, one can

- visualize the L-ribbon and try to determine the value of μ that corresponds to the "vertex", i.e., the point of largest curvature (in magnitude)or, or
- estimate the curvature of L-curve, and pick the value of μ that corresponds to a point with max curvature.

The curvature of the L-curve at $P(\mu)$ is

$$\kappa_\mu = 2\frac{\hat{\rho}''_\mu \hat{\eta}''_\mu - \hat{\rho}'_\mu \hat{\eta}''_\mu}{((\hat{\rho}'_\mu)^2 + (\hat{\eta}'_\mu)^2)^{2/3}},$$

where

$$\eta_\mu = \|x_\mu\|^2, \qquad \hat{\eta}_\mu = \log(\eta_\mu),$$

$$\rho_\mu = \|b - Ax_\mu\|^2, \qquad \hat{\rho}_\mu = \log(\rho_\mu).$$

It can be shown that

$$\kappa_\mu = 2\frac{\rho_\mu \eta_\mu}{\eta'_\mu} \frac{\mu \eta'_\mu \rho_\mu + \rho_\mu \eta_\mu + \mu^2 \eta_\mu \eta'_\mu}{(\mu^2 \eta_\mu^2 + \rho_\mu^2)^{3/2}}.$$

Since

$$\eta'_\mu = -2b^T A (A^T A + \mu I)^{-3} A^T b,$$

upper and lower bounds for η'_μ can be computed by Gauss and Gauss-Radau quadrature rules also.

Thus, upper and lower bounds for the curvature of the L-curve can be obtained inexpensively from the partial Lanczos bidiagonalization.

We keep improving the bounds by increasing ℓ until it is easy to locate the maximizer.

Example. Consider the test problem shaw from the REGULARIZATION TOOLS package. This problem is an integral equation of the first kind

$$\int_\alpha^\beta K(s,t)\,x(t)\,dt = b(s), \qquad \alpha \le s \le \beta$$

with

$$\begin{aligned} K(s,t) &:= (\cos(s)+\cos(t))^2 \left(\frac{\sin(u(s,t))}{u(s,t)}\right)^2, \\ u(s,t) &:= \pi(\sin(s)+\sin(t)), \\ x(t) &:= 2\exp\left(-6(t-0.8)^2\right) + \exp\left(-2(t+0.5)^2\right) \end{aligned}$$

and $\alpha := -\pi/2$, $\beta := \pi/2$. The right-hand side function b is determined by the kernel K and the solution x.

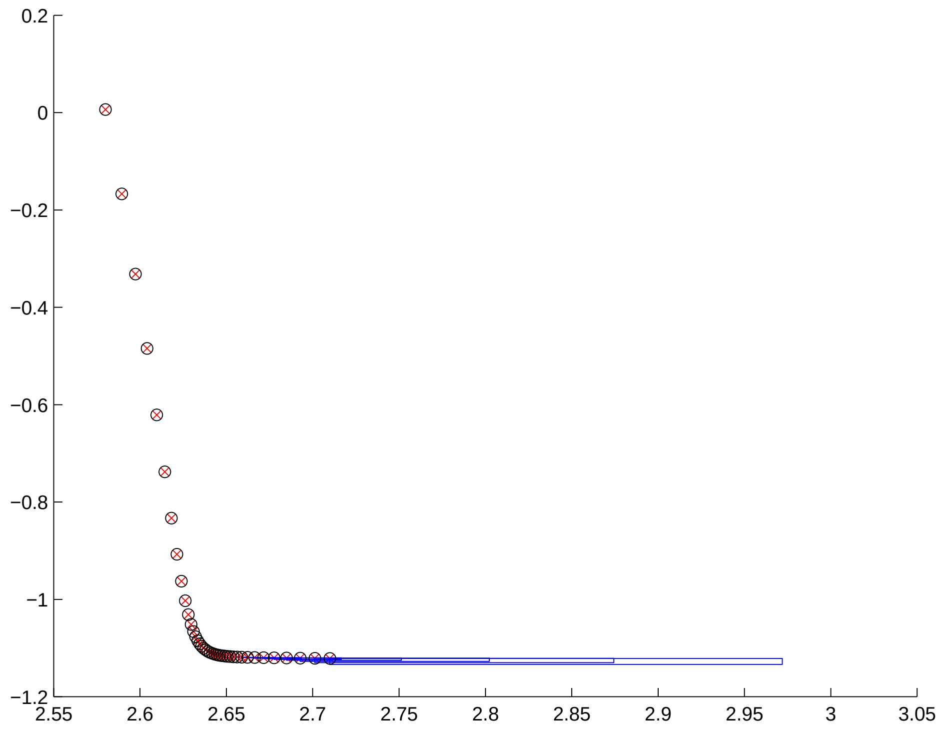
The matrix $A \in \mathrm{R}^{200\times200}$ of the linear system

$$Ax = b$$

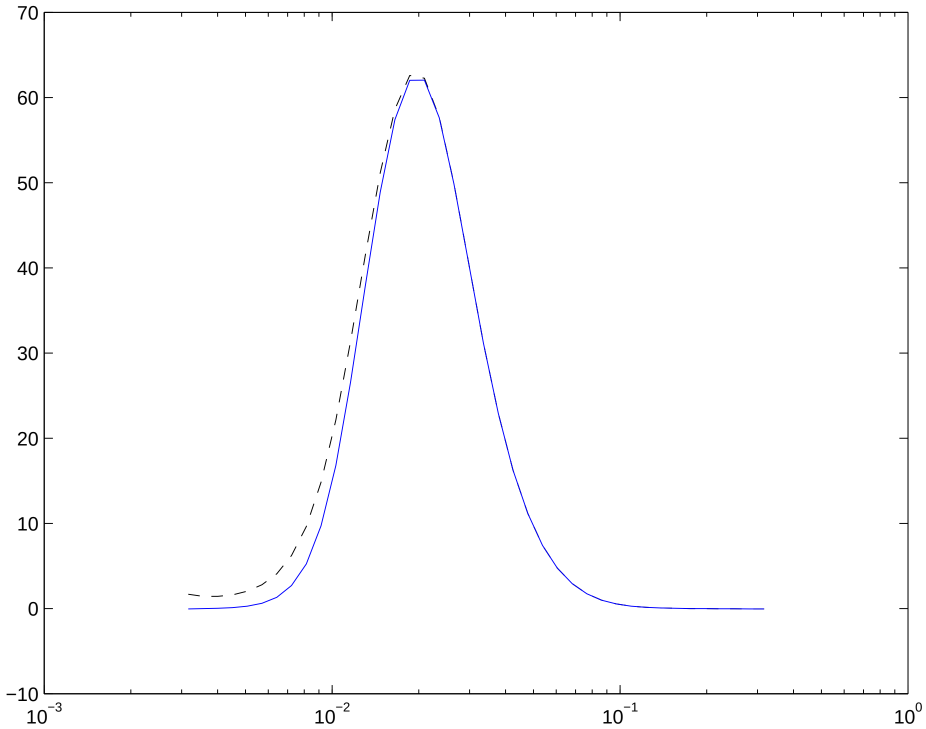
is obtained by discretizing the integral equation by a Nyström method using the composite midpoint rule with 200 equidistant nodes.

The right-hand side vector $b \in \mathrm{R}^{200}$ is contaminated by 1% noise.

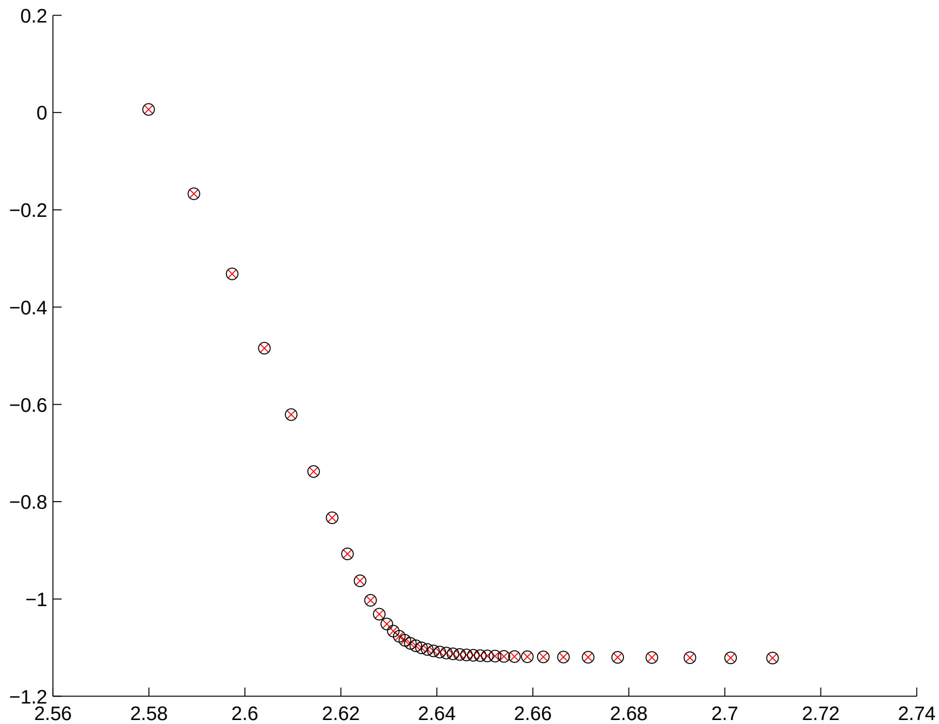
The matrix A is symmetric, indefinite and numerically singular.



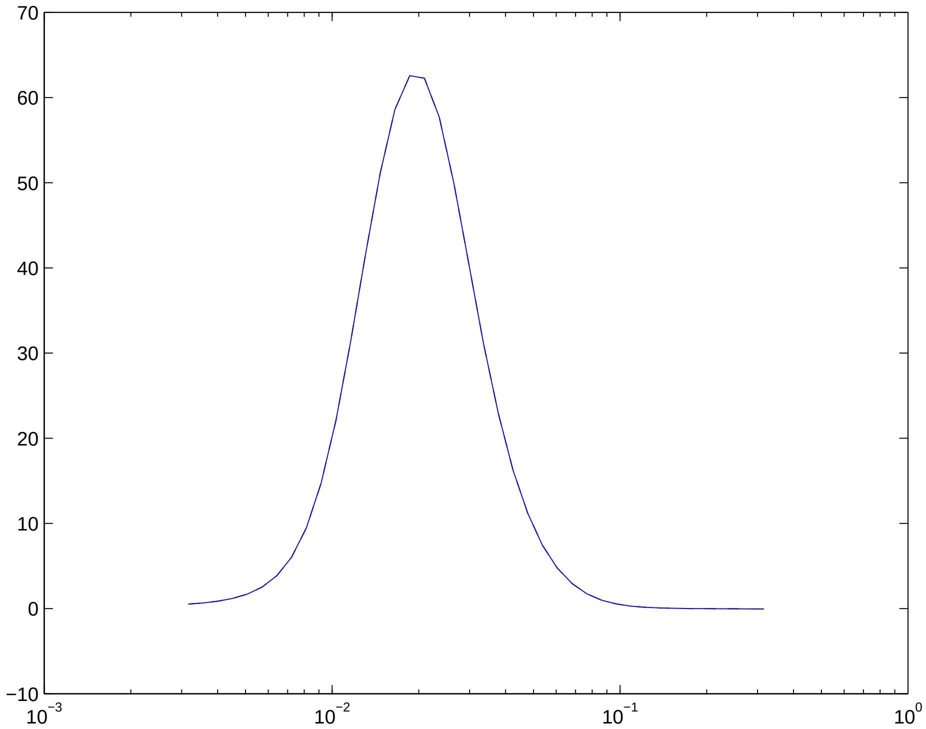
L-ribbon: $\ell = 8$ Lanczos bidiagonalization steps



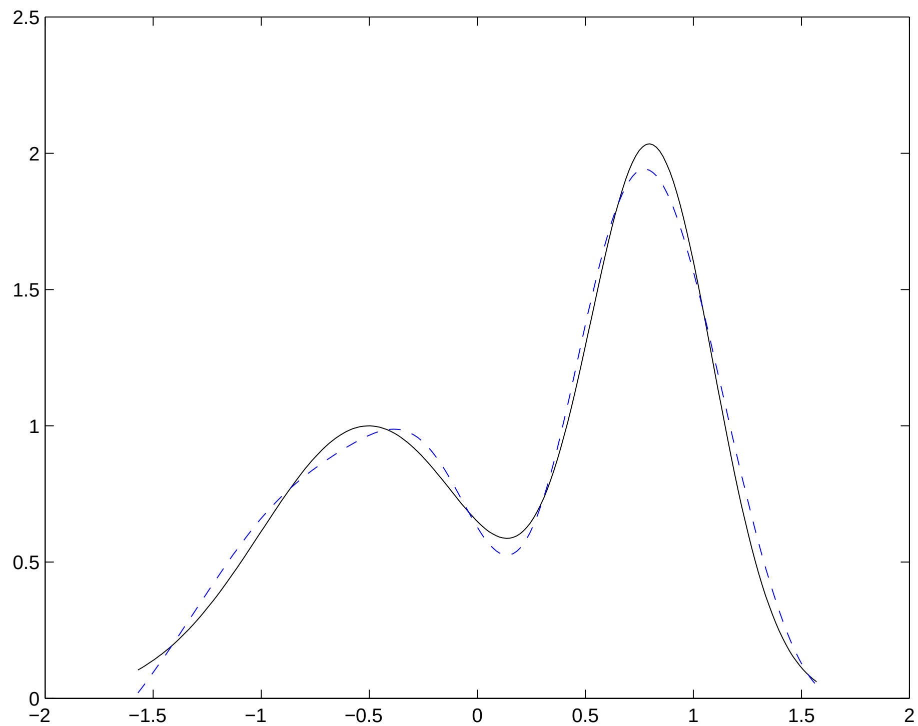
Curvature-ribbon: $\ell = 8$ Lanczos bidiagonalization steps



L-ribbon: $\ell = 9$ Lanczos bidiagonalization steps



Curvature-ribbon: $\ell = 9$ Lanczos bidiagonalization steps



Computed solution (dashed curve) with $\ell = 9$ Lanczos bidiagonalization steps and exact solution (solid curve)

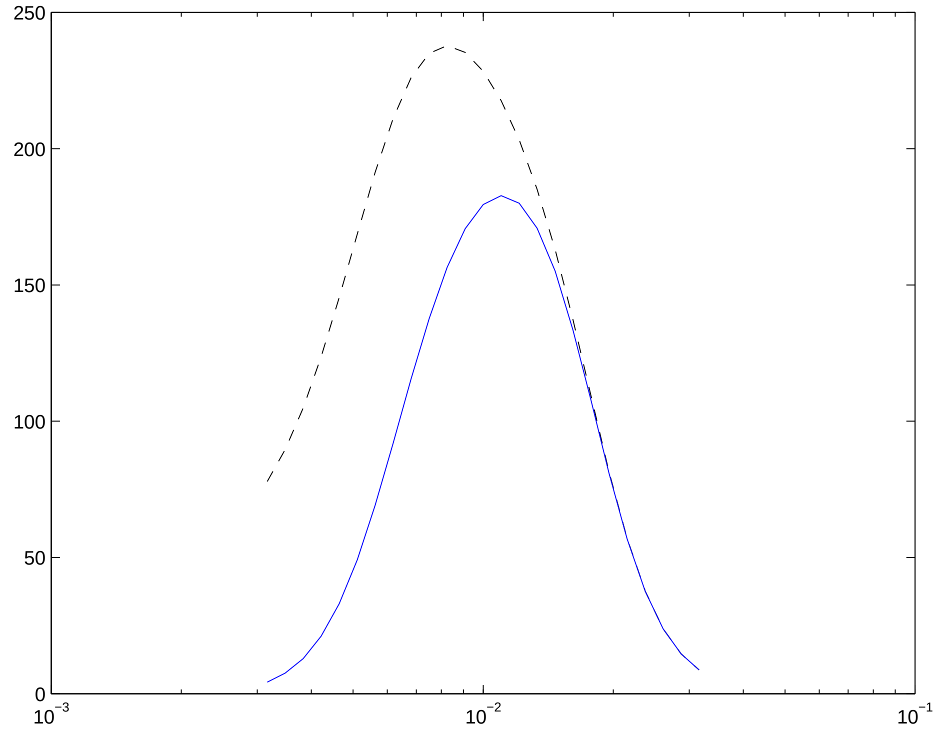
Example. The first kind integral equation models geomagnetic prospecting with

$$K(s,t) := \frac{d}{(d^2 + (s-t)^2)^{3/2}}, \qquad d := \frac{1}{4},$$

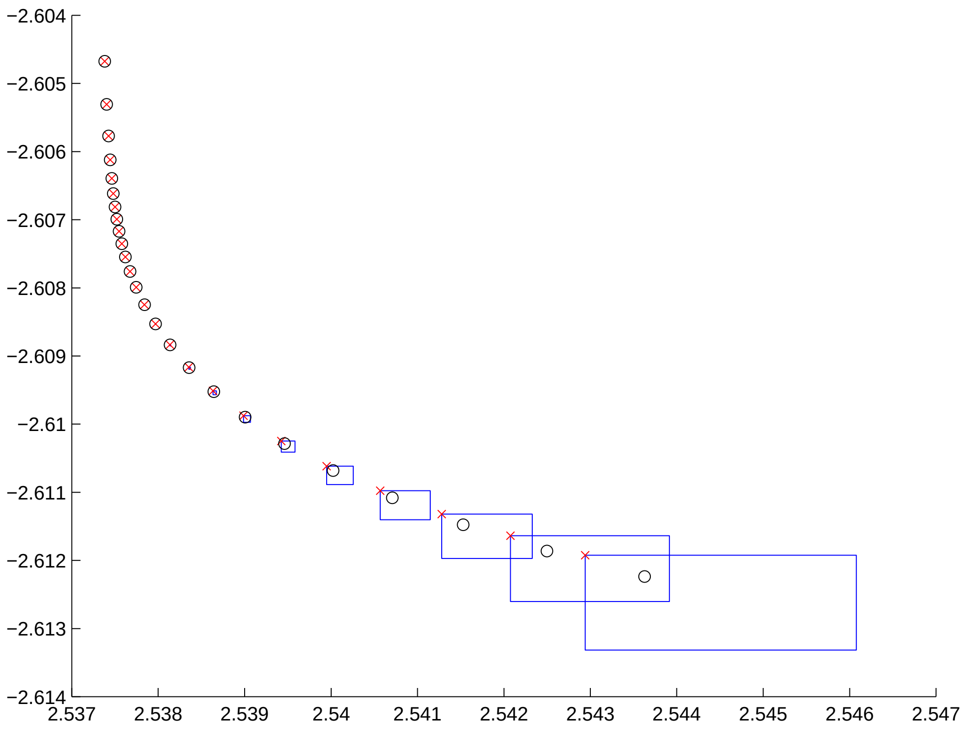
$$x(t) := \sin(\pi t) + \frac{1}{2}\sin(2\pi t), \tag{1}$$

and $\alpha := 0$, $\beta := 1$. The right-hand side function b models the vertical component of the magnetic field from a source distribution x of magnetic dipoles at depth d.

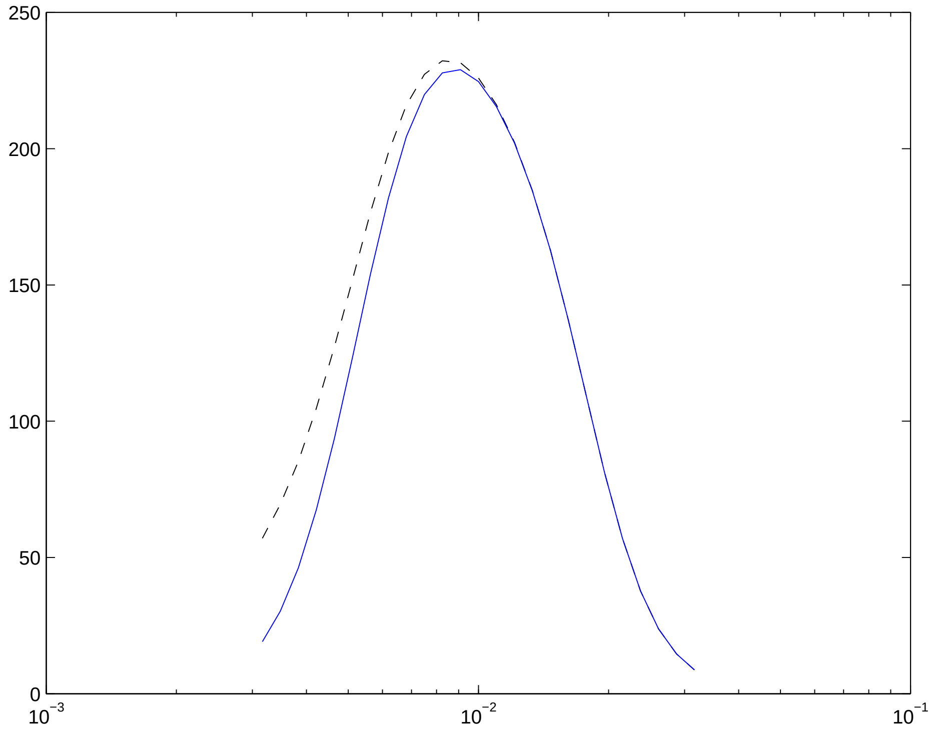
Discretization by a simple Nyström method gives a symmetric, numerically singular, matrix $A \in \mathrm{R}^{256\times 256}$.



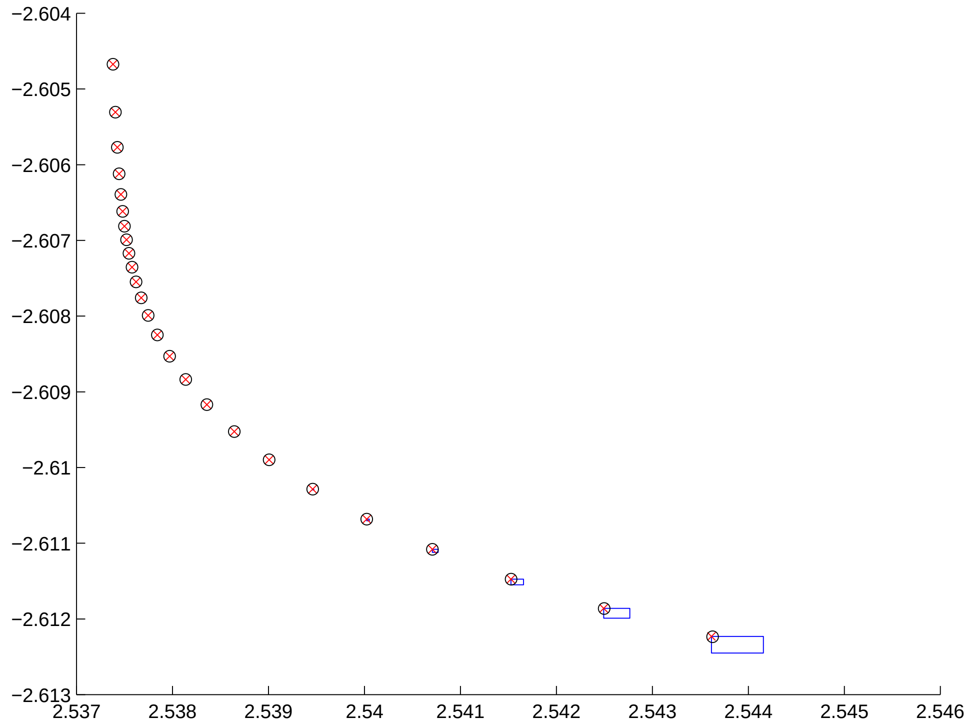
Curvature ribbon: $\ell = 12$ Lanczos bidiagonalization steps



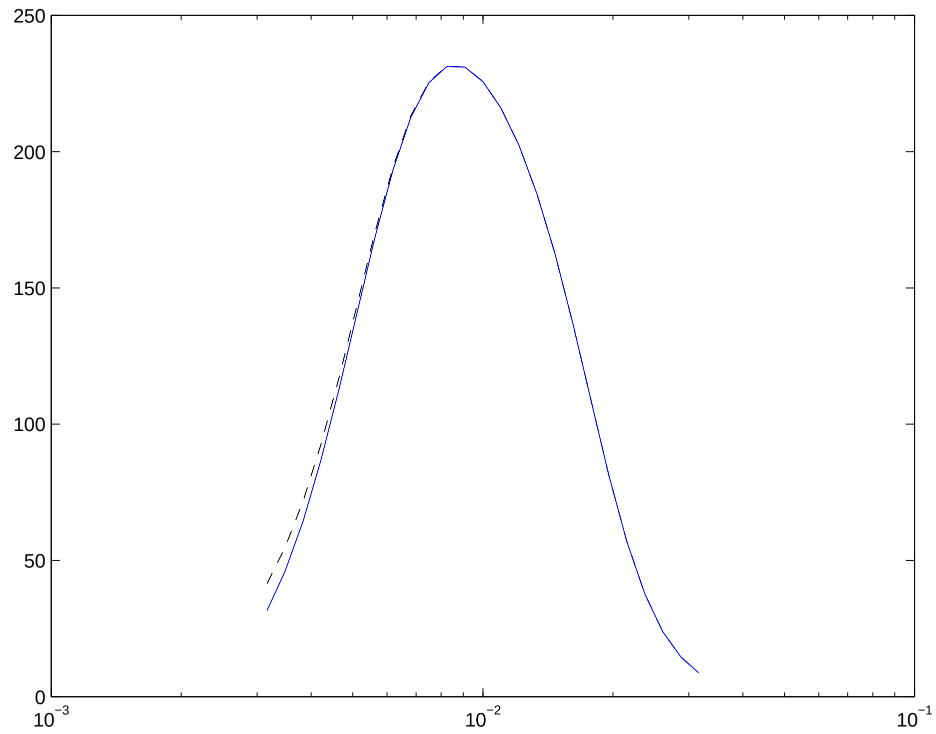
L-ribbon: $\ell = 13$ Lanczos bidiagonalization steps



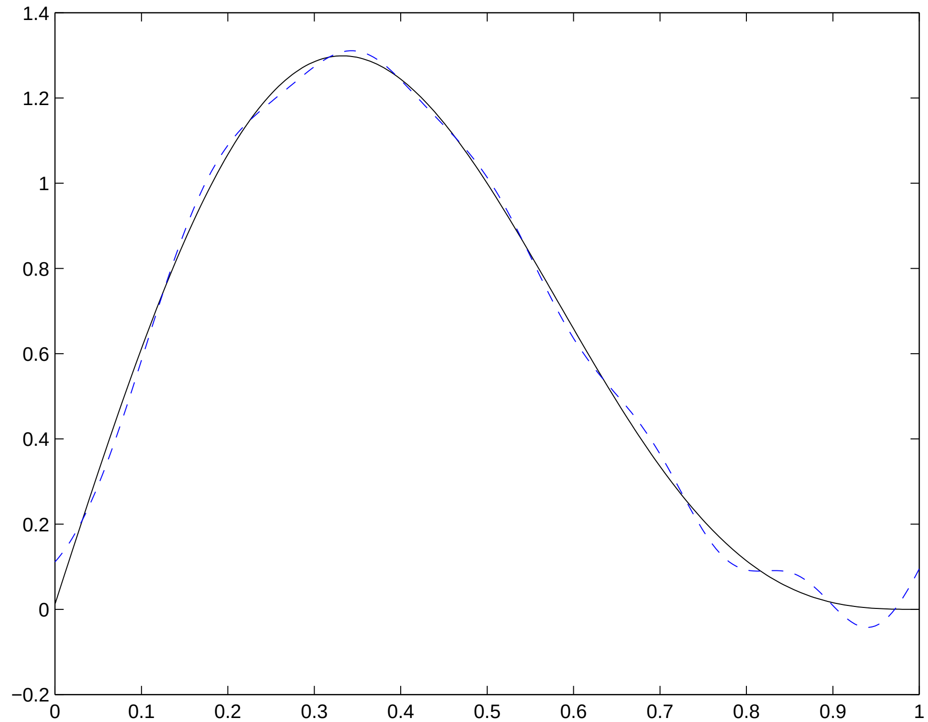
Curvature-ribbon: $\ell = 13$ Lanczos bidiagonalization steps



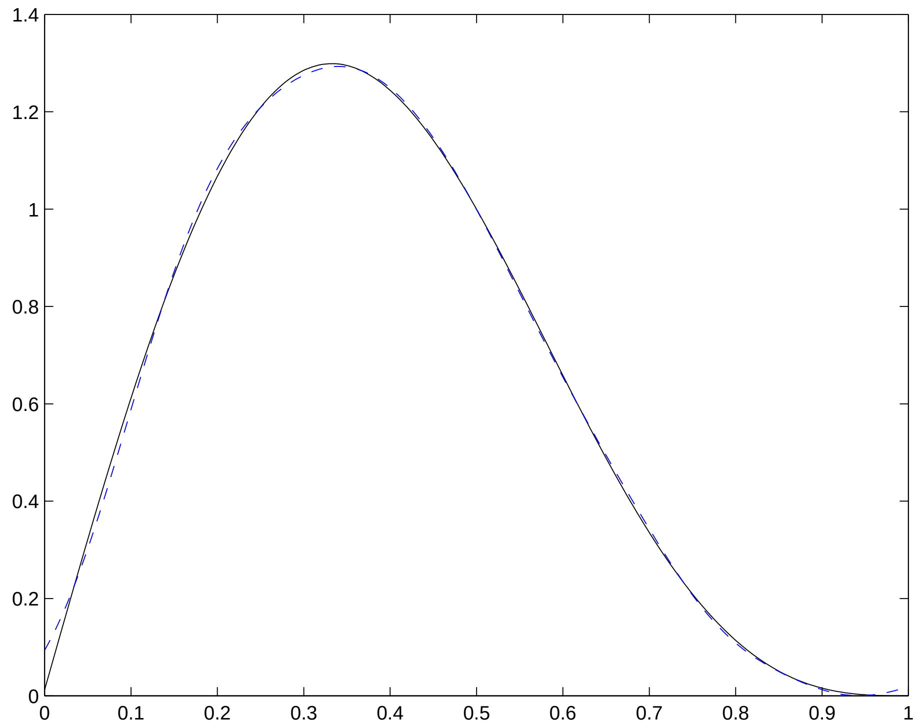
L-ribbon: $\ell = 14$ Lanczos bidiagonalization steps



Curvature-ribbon: $\ell = 14$ Lanczos bidiagonalization steps



Computed solution for $\mu = 9 \cdot 10^{-3}$ determined from curvature-ribbon (blue dashed curve) and exact solution to the noise-free continuous problem (black solid curve). $\ell = 14$ Lanczos bidiagonalization steps.



Computed solution for $\mu = 3 \cdot 10^{-3}$ and exact solution to the noise-free continuous problem (black solid curve). $\ell = 14$ Lanczos bidiagonalization steps.

Example: Computerized tomography - an underdetermined problem

256×256 pixels, 256 X-ray emittors, 256 X-ray detectors, 90 equidistant angles in $[0, \pi)$. Gives linear system

$$Ax = b, \qquad A \in \mathrm{R}^{23040 \times 65536}, \quad b \in \mathrm{R}^{23040},$$

Instead of solving $(A^T A + \mu I)x = A^T b$, we solve

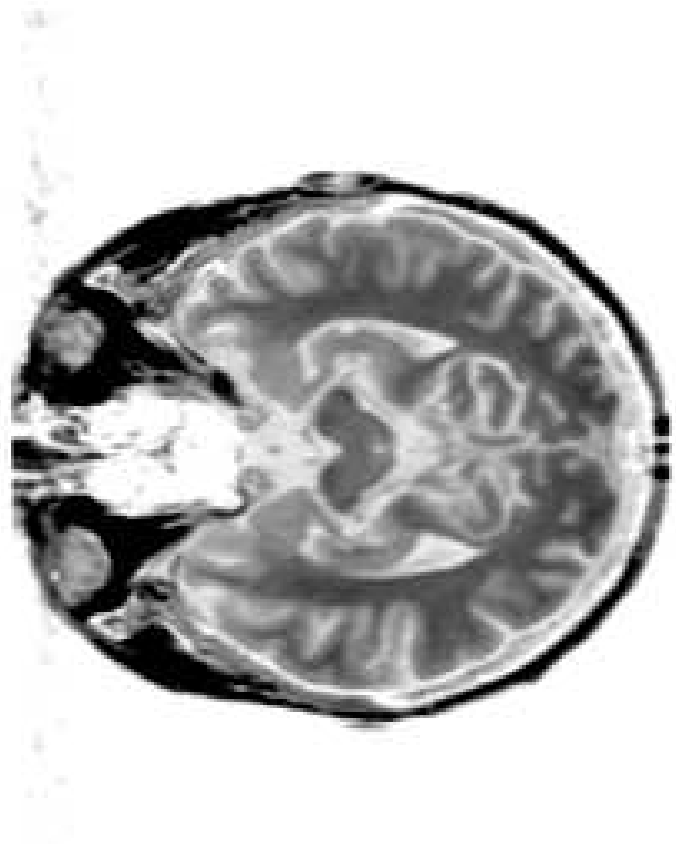
$$(AA^T + \mu I)y = b$$

and then form $x = A^T y$. We only need to store basis for y.

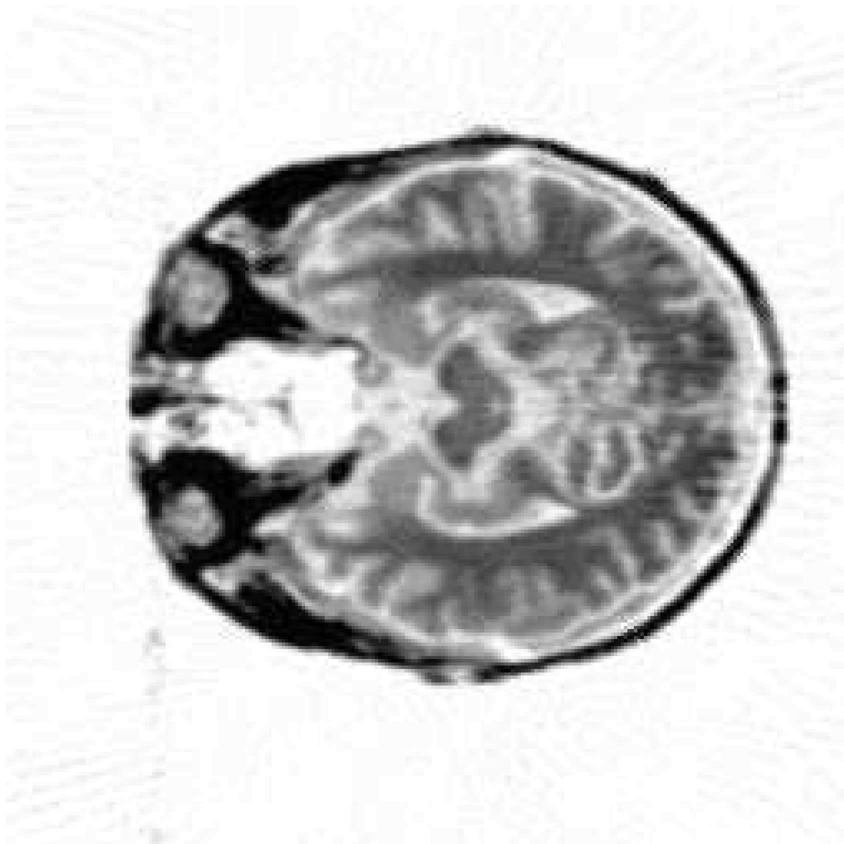
Note

- L-ribbon contains the L-curve, but
- Galerkin solution can correspond to point outside L-ribbon.

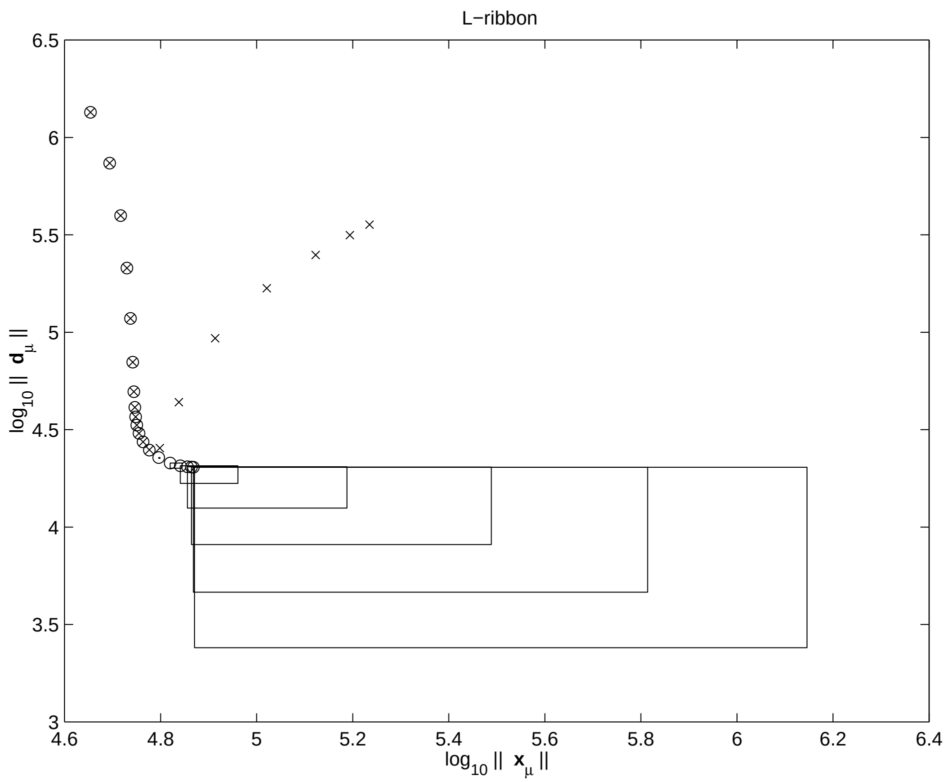
Error-free image 256×256 pixels (assumed to be unknown)



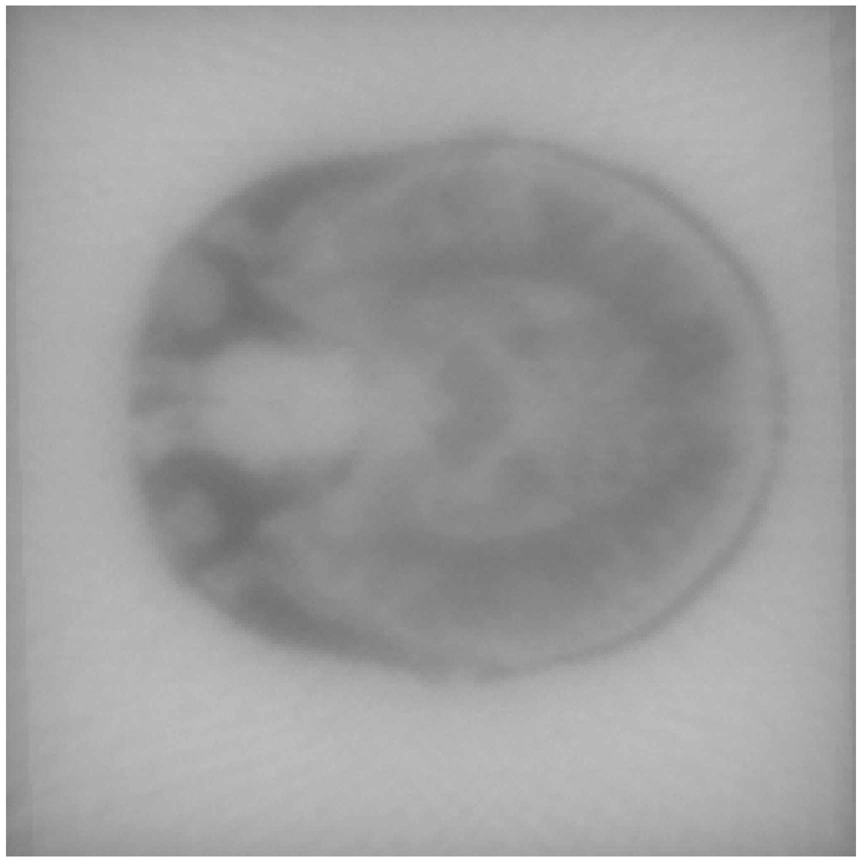
Reconstructed image from data with 1% noise, $\mu = 0.9283$, 200 Lanczos bidiagonalization steps.



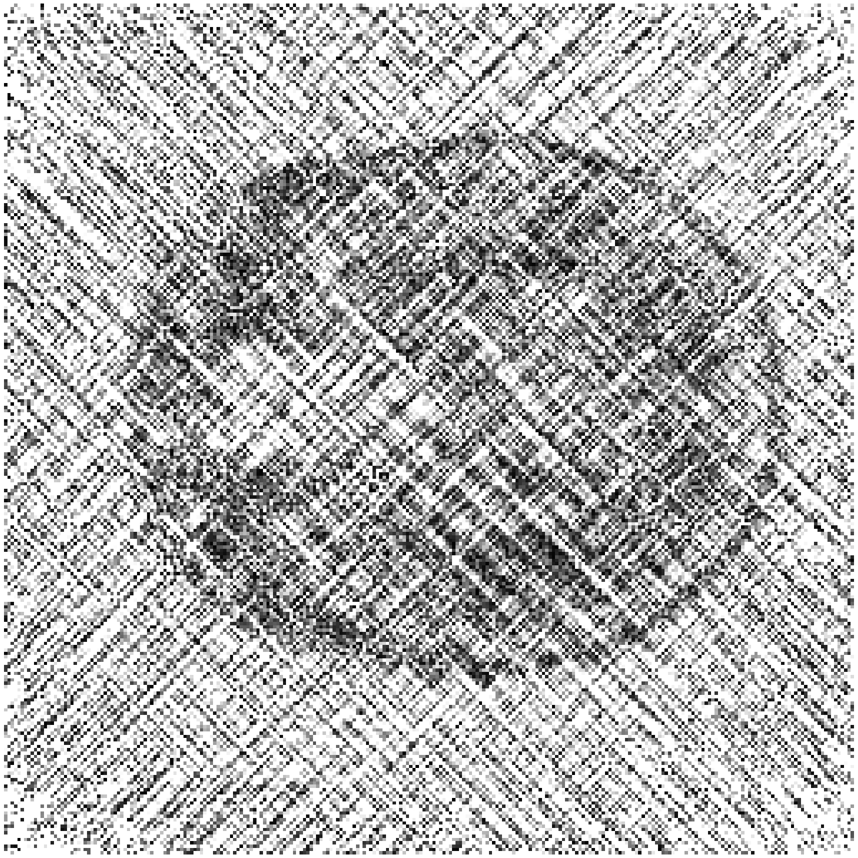
L-ribbon, × computed Galerkin solutions



Reconstructed image, μ too large.



Reconstructed image, μ too small.



Vielen Dank!