Ill–Posed Inverse Problems in Image Processing

Introduction, Structured matrices, Spectral filtering, Regularization, Noise revealing

I. Hnětynková[1], M. Plešinger[2], Z. Strakoš[3]

hnetynko@karlin.mff.cuni.cz, martin.plesinger@sam.math.ethz.ch, strakos@cs.cas.cz

[1,3]Faculty of Mathematics and Phycics, Charles University, Prague
[2]Seminar of Applied Mathematics, Dept. of Math., ETH Zürich
[1,2,3]Institute of Computer Science, Academy of Sciences of the Czech Republic

SNA '11, January 24—28

Recapitulation of Lecture I and II

Linear system

Consider an ill-posed (square nonsingular) problem

$$Ax = b, \quad b = b^{\mathrm{exact}} + b^{\mathrm{noise}}, \quad A \in \mathbb{R}^{N\times N}, \quad x,\, b \in \mathbb{R}^{N},$$

where

- A is a discretization of a smoothing operator,
- singular values of A decay,
- singular vectors of A represent increasing frequencies,
- b^{exact} is smooth and satisfies the discrete Picard condition,
- b^{noise} is unknown white noise,

$$\|b^{\mathrm{exact}}\| \gg \|b^{\mathrm{noise}}\|, \qquad \text{but} \qquad \|A^{-1}b^{\mathrm{exact}}\| \ll \|A^{-1}b^{\mathrm{noise}}\|.$$

We want to approximate

$$x^{\mathrm{exact}} = A^{-1}b^{\mathrm{exact}}.$$

Discrete Picard condition (DPC):

On average, the components $|(b^{\text{exact}}, u_j)|$ of the true right-hand side b^{exact} in the left singular subspaces of A decay faster than the singular values σ_j of A, $j = 1, \ldots, N$.

White noise:

The components $|(b^{\text{noise}}, u_j)|$, $j = 1, \ldots, N$ do not exhibit any trend.

Denote

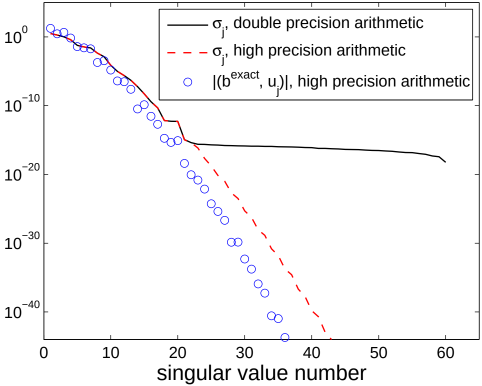
$$\delta^{\text{noise}} \equiv \frac{\| b^{\text{noise}} \|}{\| b^{\text{exact}} \|}$$

the **(usually unknown) noise level** in the data.

Recapitulation of Lecture I and II

Linear system

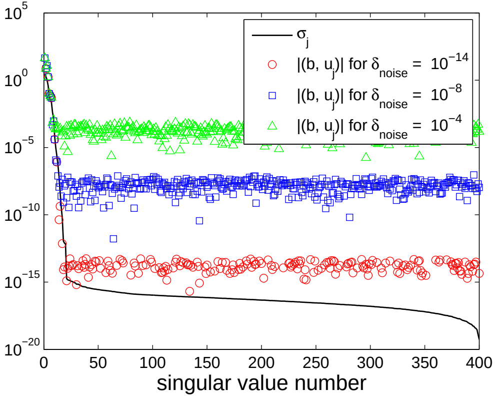
Singular values and DPC (SHAW(400)):



Recapitulation of Lecture I and II

Linear system

Violation of DPC for different noise levels (SHAW(400)):



Recapitulation of Lecture I and II

Naive solution

The components of the naive solution

$$x^{\mathrm{naive}} \equiv A^{-1}b = \underbrace{\sum_{j=1}^{k} \frac{u_j^T b^{\mathrm{exact}}}{\sigma_j} v_j}_{x^{\mathrm{exact}}} + \underbrace{\sum_{j=1}^{k} \frac{u_j^T b^{\mathrm{noise}}}{\sigma_j} v_j}_{\text{amplified noise}}$$

$$+ \underbrace{\sum_{j=k+1}^{N} \frac{u_j^T b^{\mathrm{exact}}}{\sigma_j} v_j}_{x^{\mathrm{exact}}} + \underbrace{\sum_{j=k+1}^{N} \frac{u_j^T b^{\mathrm{noise}}}{\sigma_j} v_j}_{\text{amplified noise}}$$

corresponding to small σ_j's are dominated by amplified noise.

Regularization is used to suppress the effect of errors and extract the essential information about the solution.

Recapitulation of Lecture I and II

Regularization methods

Direct regularization (TSVD, Tikhonov regularization): Suitable for solving small ill-posed problems.

Projection regularization: Suitable for solving large ill-posed problems. Regularization is often based on regularizing **Krylov subspace** iterations.

Hybrid methods: Here the **outer iterative regularization** is combined with an **inner direct regularization** of the projected small problem (i.e. of the reduced model).

The algorithm is stopped when the regularized solution of the **reduced model** matches some selected **stopping criteria** based, e.g., on the discrepancy principle, the generalized cross validation, the L-curve criterion, or the normalized cumulative periodograms.

Outline of the tutorial

- **Lecture I—Problem formulation:**

  Mathematical model of blurring, System of linear algebraic equations, Properties of the problem, Impact of noise.

- **Lecture II—Regularization:**

  Basic regularization techniques (TSVD, Tikhonov), Criteria for choosing regularization parameters, Iterative regularization, Hybrid methods.

- **Lecture III—Noise revealing:**

  Golub-Kahan iterative bidiagonalization and its properties, Propagation of noise, Determination of the noise level, Noise vector approximation, Open problems.

Outline of Lecture III

- **9. Golub-Kahan iterative bidiagonalization and its properties:**

  Basic algorithm, LSQR method.
- **10. Propagation of noise:**

  Spectral properties of bidiagonalization vectors, Noise amplification.
- **11. Determination of the noise level:**

  Motivation, Connection of GK with the Lanczos tridiagonalization, Approximation of the Riemann-Stieltjes distribution function, Estimate based on distribution functions, Identification of the noise revealing iteration.
- **12. Noise vector approximation:**

  Basic formula, Noise subtraction, Numerical illustration (SHAW and ELEPHANT image deblurring problem).
- **13. Open problems.**

9. Golub-Kahan iterative bidiagonalization and its properties

Golub-Kahan iterative bidiagonalization (GK) of A:

given $w_0 = 0$, $s_1 = b / \beta_1$, where $\beta_1 = \|b\|$, for $j = 1, 2, \ldots$

$$\begin{aligned} \alpha_j w_j &= A^T s_j - \beta_j w_{j-1}, \qquad \|w_j\| = 1, \\ \beta_{j+1} s_{j+1} &= A w_j - \alpha_j s_j, \qquad \|s_{j+1}\| = 1. \end{aligned}$$

Then $w_1, \ldots, w_k$ is an orthonormal basis of $\mathcal{K}_k(A^T A, A^T b)$, and $s_1, \ldots, s_k$ is an orthonormal basis of $\mathcal{K}_k(AA^T, b)$.

[Golub, Kahan: '65].

9. Golub-Kahan iterative bidiagonalization and its properties

Basic algorithm

Let $S_k = [s_1, \dots, s_k]$, $W_k = [w_1, \dots, w_k]$ be the associated matrices with orthonormal columns. Denote

$$L_k = \begin{bmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \ddots & \ddots & \\ & & \beta_k & \alpha_k \end{bmatrix}, \quad L_{k+} = \begin{bmatrix} L_k \\ e_k^T \beta_{k+1} \end{bmatrix}$$

the bidiagonal matrices containing the normalization coefficients.

Then GK can be written in the matrix form as

$$A^T S_k = W_k L_k^T,$$

$$A W_k = [S_k, s_{k+1}] L_{k+} = S_{k+1} L_{k+}.$$

9. Golub-Kahan iterative bidiagonalization and its properties

LSQR method

Regularization based on GK belong among popular approaches for solving **large ill-posed** problems. First the problem is **projected onto a Krylov subspace** using k steps of bidiagonalization (regularization by projection),

$$A\,x \;\approx\; b \;\longrightarrow\; S_{k+1}^T\, A\, W_k\, y = L_{k+}\, y \;\approx\; \beta_1\, e_1 = S_{k+1}^T\, b\,.$$

Then, e.g., the **LSQR method** minimizes the residual,

$$\min_{x \in x_0 + \mathcal{K}_k(A^T A, A^T b)} \|Ax - b\| = \min_{y \in \mathbb{R}^k} \|L_{k+} y - \beta_1 e_1\|\,,$$

i.e. the approximation has the form $x_k = W_k y_k$, where y_k is a least squares solution of the projected problem, [Paige, Saunders: '82].

9. Golub-Kahan iterative bidiagonalization and its properties

LSQR method

Choice of the Krylov subspace:

The vector b is dominated by low frequencies (data) and A^T has the smoothing property. Thus $A^T b$ and also

$$\mathcal{K}_k(A^T A, A^T b) = Span\{A^T b, (A^T A)A^T b, \ldots, (A^T A)^{k-1} A^T b\}$$

are dominated by **low frequencies**.

9. Golub-Kahan iterative bidiagonalization and its properties

LSQR method

Here k **is in fact the regularization parameter**:

- If k is too small, then the projected problem $L_{k+}\, y \approx \beta_1\, e_1$ does not contain enough information about the solution of the original system.
- If k is too large, then the projected problem is contaminated by noise.

Moreover, the projected problem may **inherit a part of the ill-posedness** of the original problem.

Therefore, in **hybrid methods**, some form of **inner regularization** (TSVD, Tikhonov regularization) is applied to the (small) projected problem. The method then, however, requires:

- stopping criteria for GK,
- parameter choice method for the inner regularization.

This usually requires solving the problem for **many values** of the regularization parameter and many iterations.

10. Propagation of noise

10. Propagation of noise

Spectral properties of bidiagonalization vectors

GK starts with the normalized **noisy right-hand side** $s_1 = b \,/\, \|b\|$. Consequently, vectors s_j contain information about the noise.

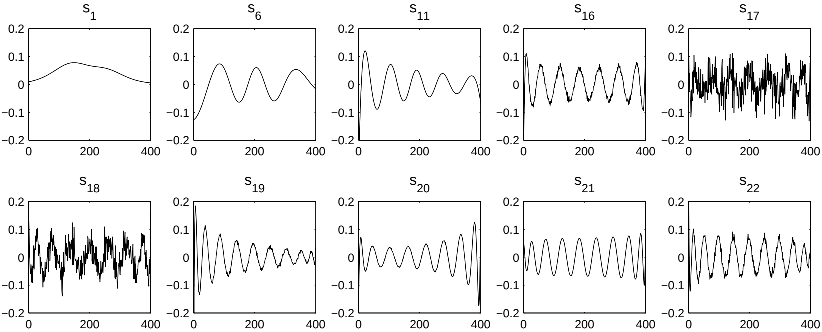
Consider the problem SHAW(400) from [Regularization Toolbox] with a noisy right-hand side (the noise was artificially added using the MatLab function `randn`). As an example we set

$$\delta^{\mathrm{noise}} \equiv \frac{\| b^{\mathrm{noise}} \|}{\| b^{\mathrm{exact}} \|} = 10^{-14} \,.$$

10. Propagation of noise

Spectral properties of bidiagonalization vectors

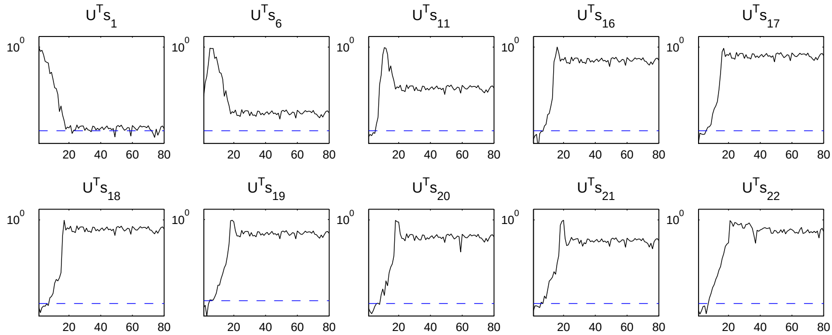
Components of several bidiagonalization vectors s_j computed via GK with double reorthogonalization:



10. Propagation of noise

Spectral properties of bidiagonalization vectors

The first 80 spectral coefficients of the vectors s_j in the basis of the left singular vectors u_j of A:



10. Propagation of noise

Spectral properties of bidiagonalization vectors

Using the three-term recurrences,

$$\beta_2 \alpha_1 s_2 = \alpha_1 (A w_1 - \alpha_1 s_1) = A A^T s_1 - \alpha_1^2 s_1 ,$$

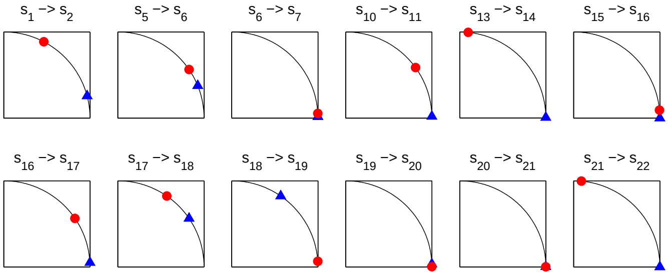
where AA^T has smoothing property. The vector s_2 is a linear combination of s_1 contaminated by the noise and $A A^T s_1$ which is smooth. Therefore the contamination of s_1 by the **high frequency part** of the noise is transferred to s_2, while a portion of the smooth part of s_1 is subtracted by orthogonalization of s_2 against s_1. **The relative level of the high frequency part of noise in s_2 must be higher than in s_1.**

In subsequent vectors $s_3, s_4, \ldots$ the relative level of the high frequency part of noise gradually increases, until the low frequency information is projected out.

10. Propagation of noise

Spectral properties of bidiagonalization vectors

Signal space – noise space diagrams:



s_k (triangle) and s_{k+1} (circle) in the signal space $\mathrm{span}\{u_1, \ldots, u_{k+1}\}$ (horizontal axis) and the noise space $\mathrm{span}\{u_{k+2}, \ldots, u_N\}$ (vertical axis).

Noise is amplified with the ratio $-\alpha_k/\beta_{k+1}$:

GK for the spectral components:

$$\begin{aligned} \alpha_1\,(V^T w_1) &= \Sigma\,(U^T s_1)\,, \\ \beta_2\,(U^T s_2) &= \Sigma\,(V^T w_1) - \alpha_1\,(U^T s_1)\,, \end{aligned}$$

and for $k = 2, 3, \ldots$

$$\begin{aligned} \alpha_k(V^T w_k) &= \Sigma\,(U^T s_k) - \beta_k(V^T w_{k-1})\,, \\ \beta_{k+1}(U^T s_{k+1}) &= \Sigma\,(V^T w_k) - \alpha_k(U^T s_k)\,. \end{aligned}$$

See [Hnětynková, Plešinger, Strakoš: '10] for a detailed derivation.

10. Propagation of noise

Noise amplification

Since dominance in $\Sigma(U^T s_k)$ and $(V^T w_{k-1})$ is shifted by one component, in $\alpha_k\,(V^T w_k) = \Sigma\,(U^T s_k) - \beta_k\,(V^T w_{k-1})\,,$ one can not expect a significant cancellation, and therefore

$$\alpha_k \approx \beta_k\,.$$

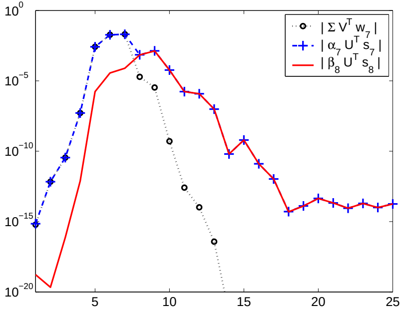
Whereas $\Sigma\,(V^T w_k)$ and $(U^T s_k)$ do exhibit dominance in the direction of the same components. If this dominance is strong enough, then the required orthogonality of s_{k+1} and s_k in $\beta_{k+1}\,(U^T s_{k+1}) = \Sigma\,(V^T w_k) - \alpha_k\,(U^T s_k)$ can not be achieved without a significant cancellation, and one can expect

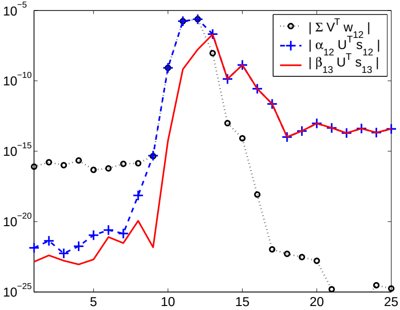
$$\beta_{k+1} \ll \alpha_k\,.$$

10. Propagation of noise

Noise amplification

Absolute values of the first 25 components of $\Sigma(V^T w_k)$, $\alpha_k(U^T s_k)$, and $\beta_{k+1}(U^T s_{k+1})$ for $k = 7$ (left) and for $k = 12$ (right), SHAW(400) with the noise level $\delta_{\text{noise}} = 10^{-14}$:





Summary:

- At the early steps of GK, the relative level of the high frequency part of noise in s_k gradually increases with k.
- At some point the low frequency information is projected out. Consequently, s_{k+1} is significantly smooter than s_k. Here the noise starts to **seriously affect** the projected problem.
- This point can be identified using **spectral analysis** of the vectors s_k (e.g. fft).

11. Determination of the noise level

11. Determination of the noise level

Motivation

If the noise level $\delta^{\text{noise}} = \| b^{\text{noise}} \| / \| b^{\text{exact}} \|$ in the data is known, many different approaches can be used for the stopping criterion in GK [Kilmer, O'Leary: '01], e.g., the discrepancy principle [Morozov: '66], [Morozov: '84], [Hansen: '98].

However, in most applications such apriory information is not available.

Can this information be obtained directly from GK?

11. Determination of the noise level

Connection of GK with the Lanczos tridiagonalization

GK is closely related to the **Lanczos tridiagonalization** [Lanczos: '50] of the symmetric matrix $A\,A^T$ with the starting vector $s_1 \;=\; b\,/\,\beta_1$,

$$A\,A^T\,S_k \;=\; S_k\,T_k \;+\; \alpha_k\,\beta_{k+1}\,s_{k+1}\,e_k^T\,,$$

where

$$T_k \;=\; L_k\,L_k^T \;=\; \begin{bmatrix} \alpha_1^2 & \alpha_1\,\beta_1 & & \\ \alpha_1\,\beta_1 & \alpha_2^2 + \beta_2^2 & \ddots & \\ & \ddots & \ddots & \alpha_{k-1}\,\beta_k \\ & & \alpha_{k-1}\,\beta_k & \alpha_k^2 + \beta_k^2 \end{bmatrix}.$$

11. Determination of the noise level

Connection of GK with the Lanczos tridiagonalization

Consequently, the matrix L_k from GK represents a **Cholesky factor** of the symmetric tridiagonal matrix T_k from the Lanczos tridiagonalization of $A\,A^T$ with the starting vector $s_1 = b\,/\,\beta_1$, see [Hnětynková, Strakoš: '07] and the references given there.

11. Determination of the noise level

Approximation of the Riemann-Stieltjes distribution function

Consider the non-decreasing piecewise constant **Riemann-Stieltjes distribution function** $\omega(\lambda)$ with the N points of increase (nodes) associated with the given (SPD) matrix $B \in \mathbb{R}^{N \times N}$, and the normalized initial vector s.

For simplicity, let **eigenvalues** $\lambda_1 < \lambda_2 < \cdots < \lambda_N$ of B be distinct. Then

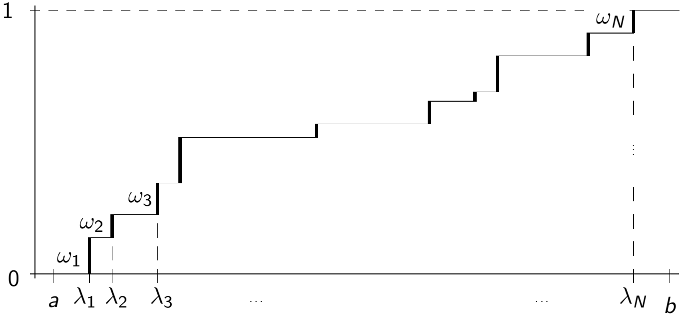
$$\omega(\lambda) = \begin{cases} 0 & \lambda < \lambda_1 \,, \\ \sum_{j=1}^{i} \omega_j & \lambda_i \leq \lambda < \lambda_{i+1} \,, \\ \sum_{j=1}^{N} \omega_j = 1 & \lambda_N \leq \lambda \,, \end{cases}$$

where the **weight** $\omega_j = |(s, v_j)|^2$ is the squared component of s in the direction of the jth invariant subspace of B.

11. Determination of the noise level

Approximation of the Riemann-Stieltjes distribution function

An example of a distribution function $\omega(\lambda)$:



11. Determination of the noise level

Approximation of the Riemann-Stieltjes distribution function

The **Lanczos tridiagonalization** of B with the starting vector s generates at each step k a non-decreasing piecewise constant distribution function $\omega^{(k)}$, with the nodes being the (distinct) eigenvalues $\eta_j^{(k)}$ of the Lanczos matrix T_k and the weights $\omega_j^{(k)}$ being the squared first entries of the corresponding normalized eigenvectors, [Hestenes, Stiefel: '52].

The distribution functions $\omega^{(k)}(\lambda)$, $k = 1, 2, \ldots$ represent **Gauss-Christoffel quadrature approximations** of the distribution function $\omega(\lambda)$, [Hestenes, Stiefel: '52], [Fischer: '96], [Meurant, Strakoš: '06].

11. Determination of the noise level

Approximation of the Riemann-Stieltjes distribution function

The Riemann-Stieltjes integral of a function $f(\lambda)$ defined on a closed interval $< a, b >$, where $a \leq \lambda_1, \lambda_N \leq b$,

$$\int_a^b f(\lambda)\, d\omega(\lambda) \equiv \sum_{j=1}^{N} \omega_j\, f(\lambda_j)\,,$$

is in step k of the Lanczos tridiagonalization approximated by the k-th Gauss-Christoffel quadrature rule

$$\sum_{j=1}^{k} \omega_j^{(k)}\, f(\eta_j^{(k)})\,.$$

In our case, $B = A\,A^T$, $s = s_1 = b/\beta_1$ and $T_k = L_k\, L_k^T$, where L_k is the bidiagonal matrix from the GK bidiagonalization of A.

Consider the SVD

$$L_k \;=\; P_k\, \Theta_k\, Q_k{}^T,$$

$P_k = [p_1^{(k)}, \ldots, p_k^{(k)}]\,,\;\; Q_k = [q_1^{(k)}, \ldots, q_k^{(k)}]\,,$
$\Theta_k = \mathrm{diag}\,(\theta_1^{(k)}, \ldots, \theta_n^{(k)})\,,$
with the singular values ordered in the **increasing** order,

$$0 \;<\; \theta_1^{(k)} < \;\ldots\; < \theta_k^{(k)}\,.$$

Then $T_k = L_k \, L_k^T = P_k \, \Theta_k^2 \, P_k^T$ is the spectral decomposition of T_k ,

$(\theta_\ell^{(k)})^2$ are its **eigenvalues** (the Ritz values of AA^T) and

$p_\ell^{(k)}$ its **eigenvectors** (which determine the Ritz vectors of AA^T), $\ell = 1, \ldots, k$.

Consequently, the GK bidiagonalization generates at each step k the distribution function

$$\omega^{(k)}(\lambda) \quad \textbf{with nodes} \quad (\theta_\ell^{(k)})^2 \quad \textbf{and weights} \quad \omega_\ell^{(k)} = |(p_\ell^{(k)}, e_1)|^2$$

that approximates the distribution function

$$\omega(\lambda) \quad \textbf{with nodes} \quad \sigma_j^2 \quad \textbf{and weights} \quad \omega_j = |(b/\beta_1, u_j)|^2 \, ,$$

where σ_j^2, u_j are the eigenpairs of $A\,A^T$, for $j = N, \ldots, 1$, [Hestenes, Stiefel: '52], [Fischer: '96], [Meurant, Strakoš: '06].

Note that unlike the Ritz values $(\theta_\ell^{(k)})^2$, the squared singular values σ_j^2 are enumerated in *descending* order.

11. Determination of the noise level

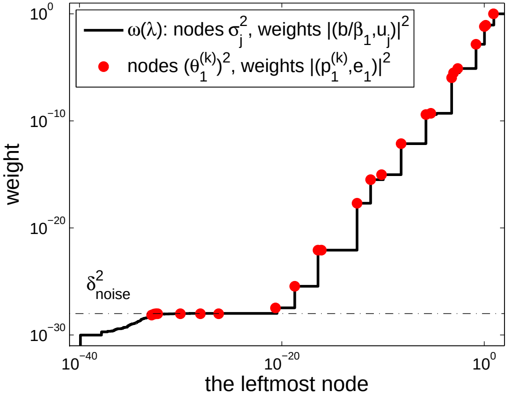
Approximation of the Riemann-Stieltjes distribution function

MatLab demo for the discrete **ill-posed problem** SHAW(400) ...

11. Determination of the noise level

Approximation of the Riemann-Stieltjes distribution function

The smallest node and weight in approximation of $\omega(\lambda)$ for the discrete **ill-posed problem** SHAW(400):



11. Determination of the noise level

Estimate based on distribution functions

The distribution function $\omega(\lambda)$:

The large nodes $\sigma_1^2, \sigma_2^2, \ldots$ of $\omega(\lambda)$ are well-separated (relatively to the small ones) and their weights on average decrease faster than σ_1^2, σ_2^2 due to the DPC. Therefore the **large nodes** essentially **control the behavior of the early stages of the Lanczos process.**

11. Determination of the noise level

Estimate based on distribution functions

Depending on the noise level, the weights corresponding to **smaller nodes** are completely dominated by noise, i.e., there exists an index J_{noise} such that

$$|(b/\beta_1, u_j)|^2 \approx |(b^{\mathrm{noise}}/\beta_1, u_j)|^2\,, \qquad \text{for } j \geq J_{\mathrm{noise}}\,.$$

The **weight of the set of the associated nodes** is given by

$$\delta^2 \equiv \sum_{j=J_{\mathrm{noise}}}^{N} |(b^{\mathrm{noise}}/\beta_1, u_j)|^2 \approx 1/\beta_1^2 \sum_{j=1}^{N} |(b^{\mathrm{noise}}, u_j)|^2 = \delta_{\mathrm{noise}}^2\,.$$

11. Determination of the noise level

Estimate based on distribution functions

The distribution functions $\omega^{(k)}(\lambda)$:

At **any** iteration step, the weight of $\omega^{(k)}(\lambda)$ corresponding to the **smallest node** $(\theta_1^{(k)})^2$ must be larger than the sum of weights of all σ_j^2 smaller than this $(\theta_1^{(k)})^2$, see [Fischer, Freund: '94] (this result goes back to Chebychev).

As k increases, some $(\theta_1^{(k)})^2$ eventually approaches (or becomes smaller than) the node $\sigma_{J_{\mathrm{noise}}}^2$, and its weight becomes

$$|(p_1^{(k)}, e_1)|^2 \approx \delta^2 \approx \delta_{\mathrm{noise}}^2 .$$

Summarizing:

The weight $|(p_1^{(k)}, e_1)|^2$ corresponding to the smallest Ritz value $(\theta_1^{(k)})^2$ of $A\,A^T$ is strictly decreasing. At some iteration step it sharply **starts to (almost) stagnate close to the squared noise level** $\delta^2_{\mathrm{noise}}$, see [Hnětynková, Plešinger, Strakoš: '10].

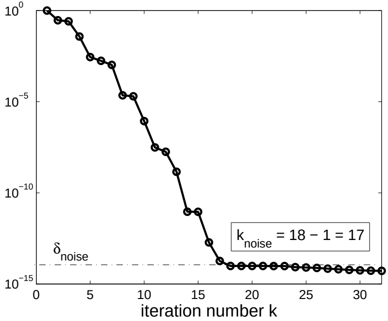
The **last iteration before** this happens is called the **noise revealing iteration** k_{noise}.

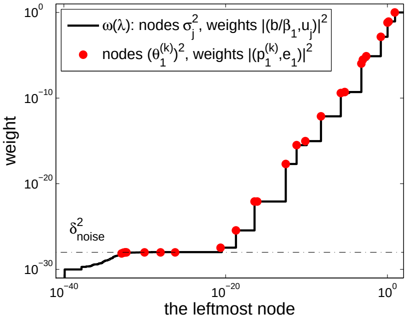
Note that computation of $|(p_1^{(k)}, e_1)|^2$ can be realized without forming the SVD of L_k using the shift-invert strategy.

11. Determination of the noise level

Estimate based on distribution functions

Square roots of the weights $|(p_1^{(k)}, e_1)|^2$, $k = 1, 2, \ldots$ (left), and the smallest node and weight in approximation of $\omega(\lambda)$ (right), SHAW(400) with the noise level $\delta_{\mathrm{noise}} = 10^{-14}$:

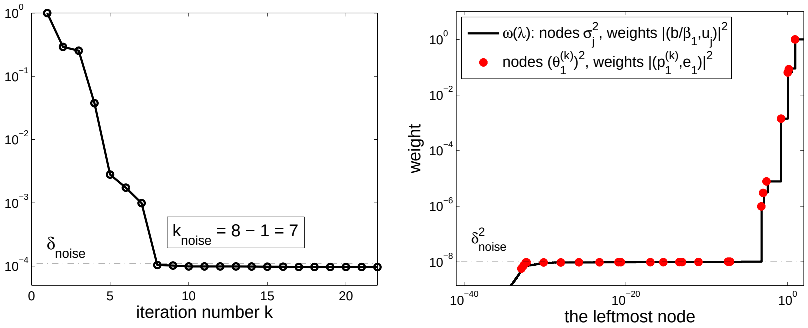




11. Determination of the noise level

Estimate based on distribution functions

Square roots of the weights $|(p_1^{(k)}, e_1)|^2$, $k = 1, 2, \ldots$ (left), and the smallest node and weight in approximation of $\omega(\lambda)$ (right), SHAW(400) with the noise level $\delta_{\mathrm{noise}} = 10^{-4}$:



11. Determination of the noise level

Identification of the noise revealing iteration

In order to estimate δ_{noise}, the iteration k_{noise} must be identified. This can be done by an **automated procedure** that does not rely on human interaction.

For example, in our experiments k_{noise} was determined as the first iteration for which

$$\frac{|(p_1^{(k+1)}, e_1)|}{|(p_1^{(k+1+step)}, e_1)|} < \left(\frac{|(p_1^{(k)}, e_1)|}{|(p_1^{(k+1)}, e_1)|}\right)^{\zeta},$$

where ζ was set to 0.5 and *step* was set to 3.

Noise level δ_{noise} in the data, iteration k_{noise}, and the estimated noise level $|(p_1^{(k_{\text{noise}}+1)}, e_1)|$, for two problems from [Regularization Toolbox]. The estimates represent average values computed using 1000 randomly chosen vectors b^{noise}:

SHAW(400)				
δ_{noise}	1×10^{-14}	1×10^{-6}	1×10^{-4}	1×10^{-2}
k_{noise}	16	9	7	4
estimate	1.80×10^{-14}	1.31×10^{-6}	1.01×10^{-4}	1.03×10^{-2}
ILAPLACE(100,1)				
δ_{noise}	1×10^{-13}	1×10^{-7}	1×10^{-2}	1×10^{-1}
k_{noise}	22	15.30	6.02	2
estimate	9.12×10^{-14}	1.34×10^{-7}	1.02×10^{-2}	1.11×10^{-1}

12. Noise vector approximation

12. Noise vector approximation

Basic formula

In the noise revealing iteration

$$\delta_{\text{noise}} \approx |(p_1^{(k_{\text{noise}+1})}, e_1)|,$$

and the bidiagonalization vector $s_{k_{\text{noise}}}$ **is fully dominated by the high frequency noise**. Thus

$$b^{\text{noise}} \approx \|b^{\text{noise}}\| \; s_{k_{\text{noise}}} \approx \beta_1 \, |(p_1^{(k_{\text{noise}}+1)}, e_1)| \, s_{k_{\text{noise}}} \, ,$$

represents an approximation of the unknown noise.

We can **subtract the reconstructed noise from the noisy observation vector** b. Hopefully, the noise level in the corrected system will be **lower** than in the original one.

What happens if we **repeat** this process several times?

12. Noise vector approximation

Noise subtraction

Algorithm: Given A, b; $b^{(0)} := b$;

for $j = 1, \ldots, t$

- GK bidiagonalization of A with the starting vector $b^{(j-1)}$;
- identification of the noise revealing iteration k_{noise};
- $\delta^{(j-1)} := |(p_1^{(k_{\text{noise}})}, e_1)|$;
- $b^{\text{noise},(j-1)} := \beta_1 \, \delta^{(j-1)} \, s_{k_{\text{noise}}}$; `// noise approximation`
- $b^{(j)} := b^{(j-1)} - b^{\text{noise},(j-1)}$; `// correction`

end;

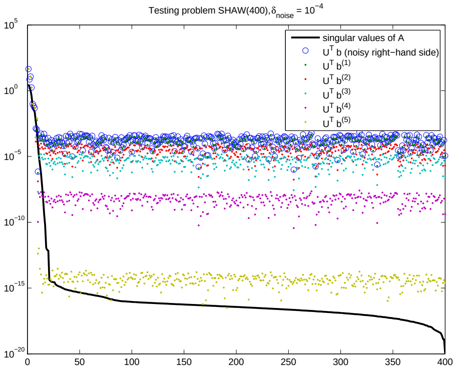
The **accumulated noise approximation** is

$$\hat{b}^{\text{noise}} \equiv \sum_{j=0}^{t-1} b^{\text{noise},(j)} \,.$$

12. Noise vector approximation

Numerical illustration - SHAW problem

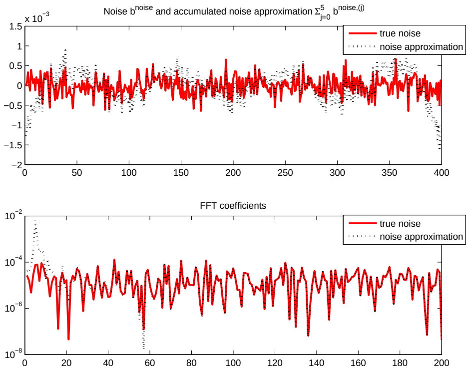
Singular values of A, and spectral coeffs. of the original and corrected observation vector $b^{(j)}$, $j = 1, \ldots, 5$, SHAW(400) with the noise level $\delta_{\text{noise}} = 10^{-4}$ ($k_{\text{noise}} = 10$ is **fixed**):



12. Noise vector approximation

Numerical illustration - SHAW problem

Individual components (top) and Fourier coeffs. (bottom) of $\hat{b}^{\mathrm{noise}}$, SHAW(400) with the noise level $\delta_{\mathrm{noise}} = 10^{-4}$:



12. Noise vector approximation

Numerical illustration - ELEPHANT image deblurring problem

Elephant image deblurring problem: image size 324×470 pixels, problem dimension $N = 152280$, the exact solution (left) and the noisy right-hand side (right), $\delta_{\mathrm{noise}} = 3 \times 10^{-3}$:

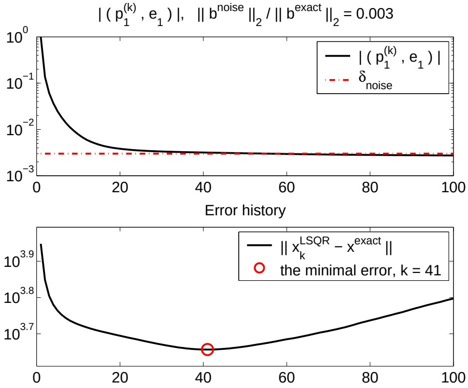
x^{exact}

$b^{exact} + b^{noise}$

12. Noise vector approximation

Numerical illustration - ELEPHANT image deblurring problem

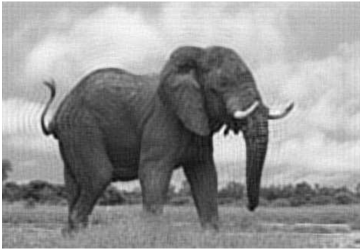
Square roots of the weights $|(p_1^{(k)}, e_1)|^2$, $k = 1, 2, \ldots$ (top) and error history of LSQR solutions (bottom):

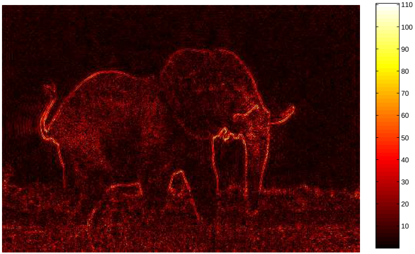


12. Noise vector approximation

Numerical illustration - ELEPHANT image deblurring problem

The best LSQR reconstruction (left), x_{41}^{LSQR}, and the corresponding componentwise error (right). GK without any reorthogonalization:

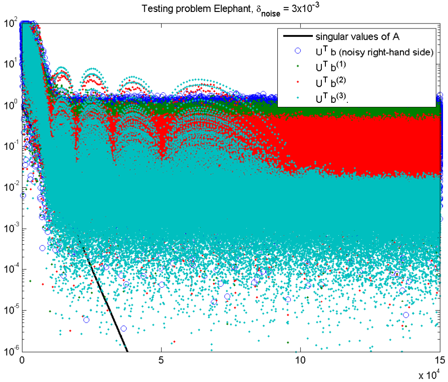




12. Noise vector approximation

Numerical illustration - ELEPHANT image deblurring problem

Singular values of A, and spectral coeffs. of the original and corrected observation vector $b^{(j)}$, $j = 1, \ldots, 3$, Elephant image deblurring problem with $\delta_{\mathrm{noise}} = 3 \times 10^{-3}$ (k_{noise} corresponds to the best LSQR approximation of x):



13. Open problems

Message:

Using GK, information about the noise can be obtained in a straightforward and cheap way.

Open problems:

- Large scale problems (determining k_{noise});
- Behavior in finite precision arithmetic
  (GK without reorthogonalization);
- Regularization;
- Denoising;
- Colored noise.

The full version of our presentations will be available at

http://www.cs.cas.cz/krylov/

Thank you for your kind attention!