Revealing the noise in the data via the Golub-Kahan iterative bidiagonalization

Iveta Hnětynková, Martin Plešinger, Zdeněk Strakoš

Charles University, Prague

Technical University, Liberec

Academy of Sciences of the Czech republic, Prague

with thanks to P. C. Hansen, M. Kilmer and many others

Prague, July 2010

Outline

1. **Problem formulation**
2. Golub-Kahan iterative bidiagonalization, and approximation of the Riemann-Stieltjes distribution function
3. Propagation of the noise in the Golub-Kahan bidiagonalization
4. Determination of the noise level
5. Noise reconstruction and subtraction – a numerical experiment
6. Summary and future work

Consider an ill-posed square nonsingular linear algebraic system

$$A\,x \approx b, \qquad A \in \mathbb{R}^{n\times n}, \qquad b \in \mathbb{R}^{n},$$

with the right-hand side corrupted by a white noise

$$b = b^{\text{exact}} + b^{\text{noise}} \neq 0 \in \mathbb{R}^{n}, \quad \|\, b^{\text{exact}} \,\| \gg \|\, b^{\text{noise}} \,\|,$$

and the goal to approximate $x^{\text{exact}} \equiv A^{-1}\, b^{\text{exact}}$.

The noise level $\delta_{\text{noise}} \equiv \dfrac{\|b^{\text{noise}}\|}{\|b^{\text{exact}}\|} \ll 1\,.$

Properties (assumptions):

- matrices A, A^T, AA^T have a smoothing property;
- left singular vectors u_j of A represent increasing frequencies as j increases;
- the system $A\,x = b^{\text{exact}}$ satisfies the discrete Picard condition.

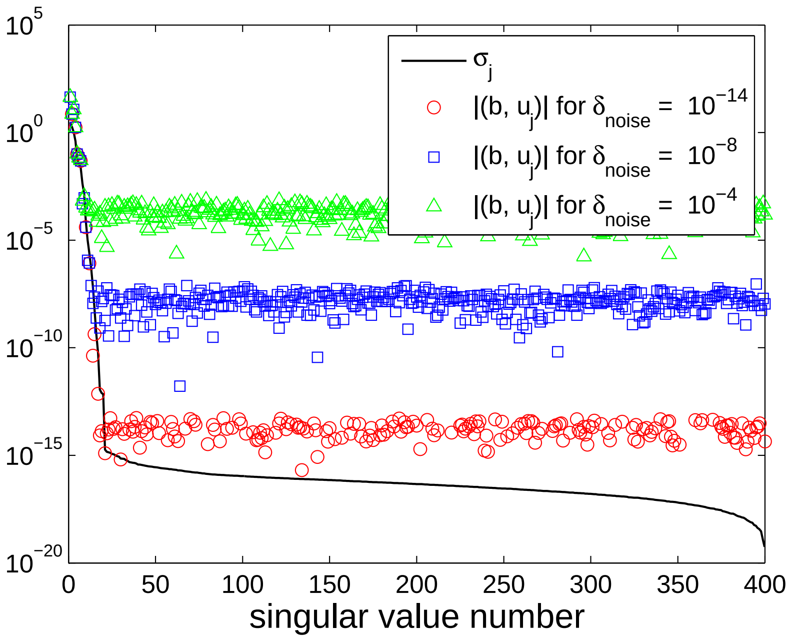
Discrete Picard condition (DPC):

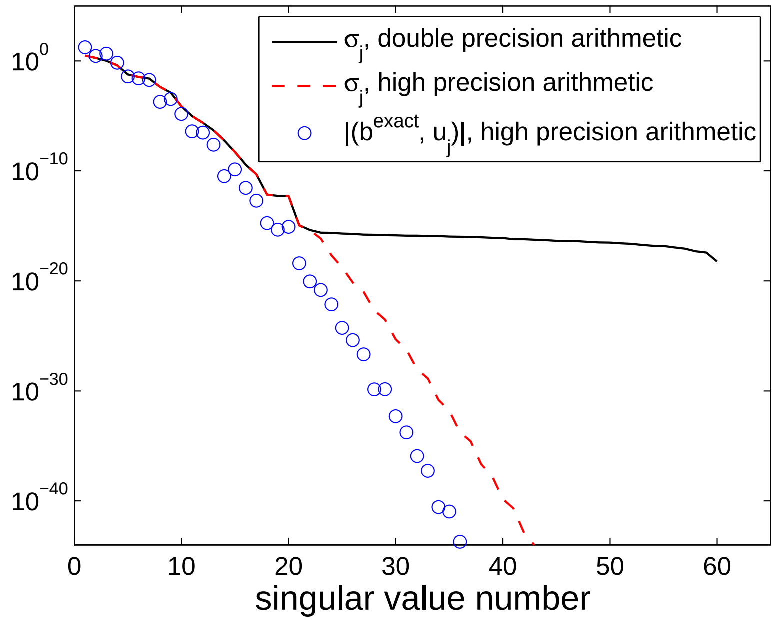
On average, the components $|(b^{\text{exact}}, u_j)|$ of the true right-hand side b^{exact} in the left singular subspaces of A decay faster than the singular values σ_j of A, $j = 1, \ldots, n$.

White noise:

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

Problem Shaw: Noise level, Singular values, and DPC: [Hansen – Regularization Tools]





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

In hybrid methods, see [O'Leary, Simmons – 81], [Hansen – 98], or [Fiero, Golub Hansen, O'Leary – 97], [Kilmer, O'Leary – 01], [Kilmer, Hansen, Español – 06], [O'Leary, Simmons – 81], the outer bidiagonalization is combined with an inner regularization – e.g., truncated SVD (TSVD), or Tikhonov regularization – of the projected small problem (i.e. of the reduced model).

The bidiagonalization is stopped when the regularized solution of the reduced model matches some selected stopping criteria.

Stopping criteria based on information from residual vectors:

A vector $\widehat{x}$ is a good approximation to $x^{\mathrm{exact}} = A^{-1}\, b^{\mathrm{exact}}$ if the corresponding residual approximates the (white) noise in the data

$$\widehat{r} \equiv b - A\,\widehat{x} \approx b^{\mathrm{noise}}\,.$$

Behavior of $\widehat{r}$ can be expressed in the frequency domain using

- discrete Fourier transform, see [Rust – 98], [Rust – 00], [Rust, O'Leary – 08], or
- discrete cosine transform, see [Hansen, Kilmer, Kjeldsen – 06],

and then analyzed using statistical tools – cumulative periodograms.

This talk:

Describe, how the noise propagates in the Golub-Kahan iterative bidiagonalization.

Under the given (quite natural) assumptions, the Golub-Kahan iterative bidiagonalization reveals the unknown noise level δ_{noise} .

A similar approach can possibly be used for approximating the noise vector b^{noise} .

Outline

1. Problem formulation
2. **Golub-Kahan iterative bidiagonalization, and approximation of the Riemann-Stieltjes distribution function**
3. Propagation of the noise in the Golub-Kahan bidiagonalization
4. Determination of the noise level
5. Noise reconstruction and subtraction – a numerical experiment
6. Summary and future work

Golub-Kahan iterative bidiagonalization (GK) of A:

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

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

Let $S_{k+1} = [s_1, \ldots, s_{k+1}]$, $W_k = [w_1, \ldots, w_k]$ 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

$$A^T S_k = W_k L_k^T,$$
$$A W_k = \left[S_k, s_{k+1} \right] L_{k+} = S_{k+1} L_{k+}.$$

GK is closely related to the **Lanczos tridiagonalization** of the symmetric matrix $A\,A^T$ with the starting vector $s_1 = b\,/\,\beta_1$,

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

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

i.e. the matrix L_k from GK represents a Cholesky factor of the symmetric tridiagonal matrix T_k from the Lanczos process.

Approximation of the distribution function:

The Lanczos tridiagonalization of the given (SPD) matrix $B \in \mathbb{R}^{n \times n}$ generates at each step k a non-decreasing piecewise constant distribution function $\omega^{(k)}$, with the nodes being the (distinct) eigenvalues 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 (i.e. minimal partial realization) approximations of the distribution function $\omega(\lambda)$, [Hestenes, Stiefel – 52], [Fischer – 96], [Meurant, Strakoš – 06].

Consider the SVD

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

$P_k = [p_1^{(k)}, \ldots, p_k^{(k)}]$, $Q_k = [q_1^{(k)}, \ldots, q_k^{(k)}]$, $\Theta_k = \text{diag}\,(\theta_1^{(k)}, \ldots, \theta_n^{(k)})$,
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),
$\ell = 1, \ldots, k$.

Summarizing:

The GK bidiagonalization generates at each step k the distribution function

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

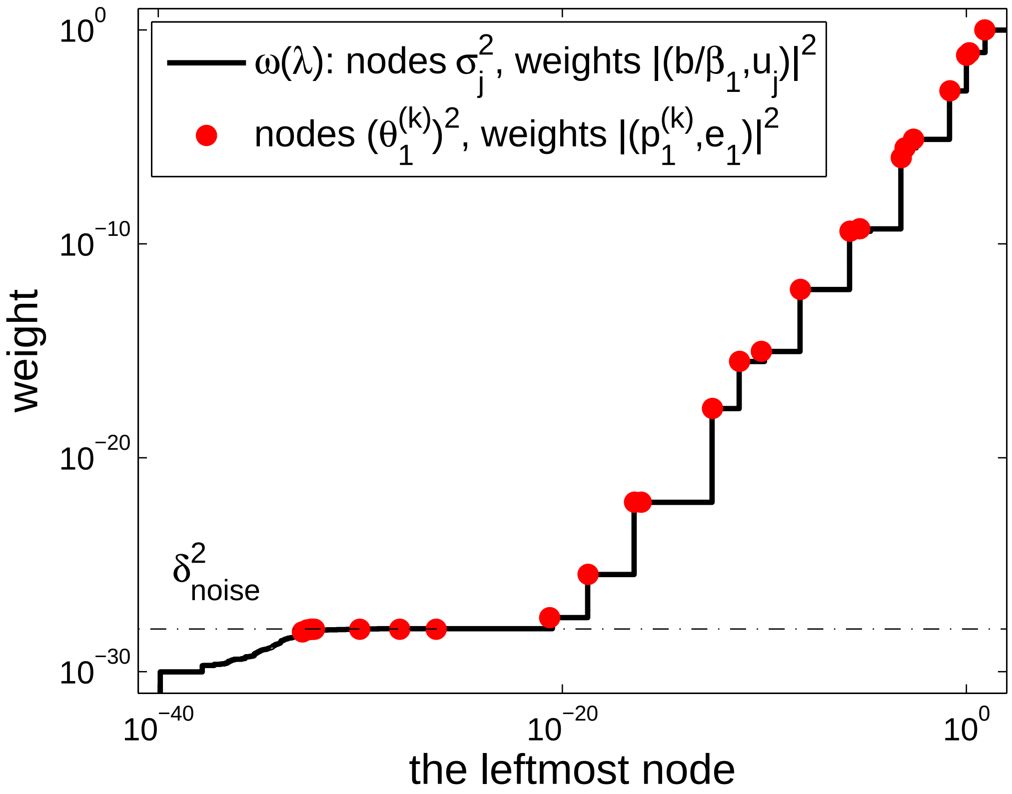
that approximates the distribution function

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

where σ_j^2, u_j are the eigenpairs of $A\,A^T$, for $j = n, \ldots, 1\,.$

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

Discrete ill-posed problem,
the smallest node and weight in approximation of $\omega(\lambda)$:



Outline

1. Problem formulation
2. Golub-Kahan iterative bidiagonalization, and approximation of the Riemann-Stieltjes distribution function
3. **Propagation of the noise in the Golub-Kahan bidiagonalization**
4. Determination of the noise level
5. Noise reconstruction and subtraction – a numerical experiment
6. Summary and future work

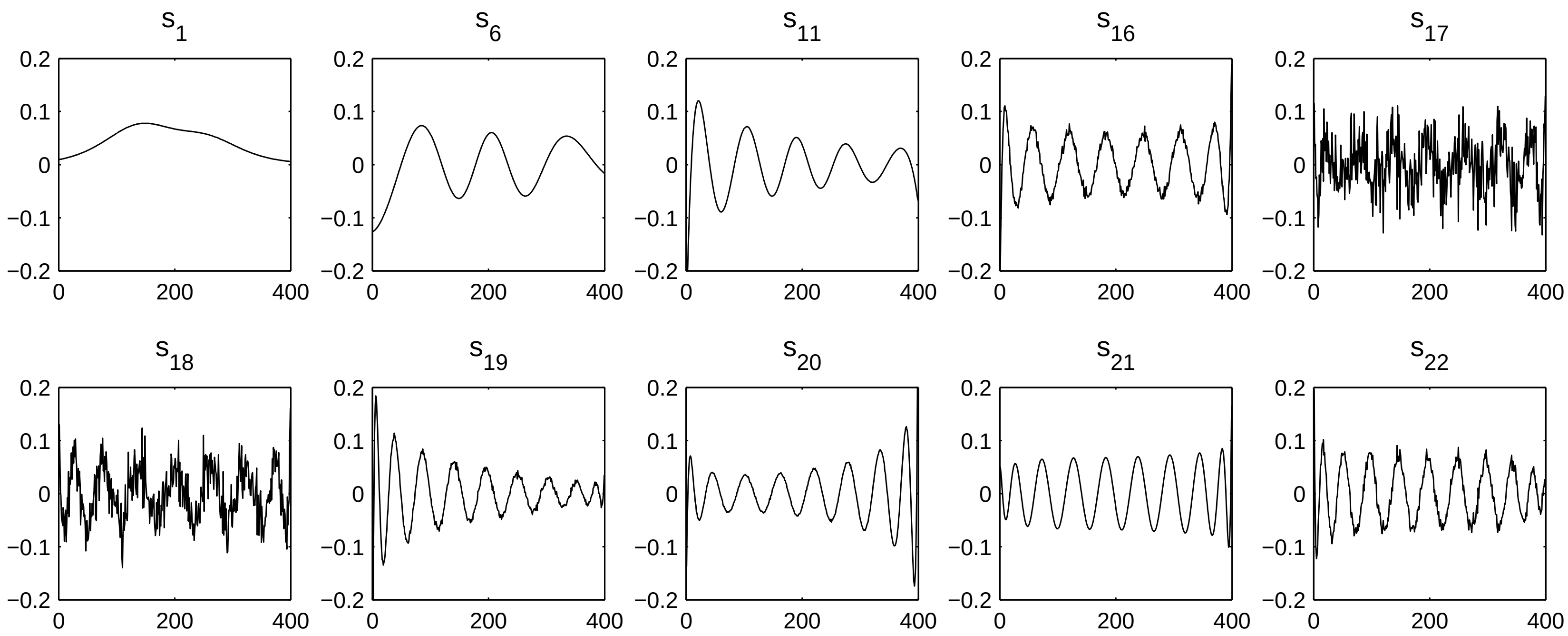
GK starts with the normalized noisy right-hand side $s_1 = b / \|b\|$. Consequently, vectors s_j contain information about the noise.

Can this information be used to determine the level of the noise in the observation vector b?

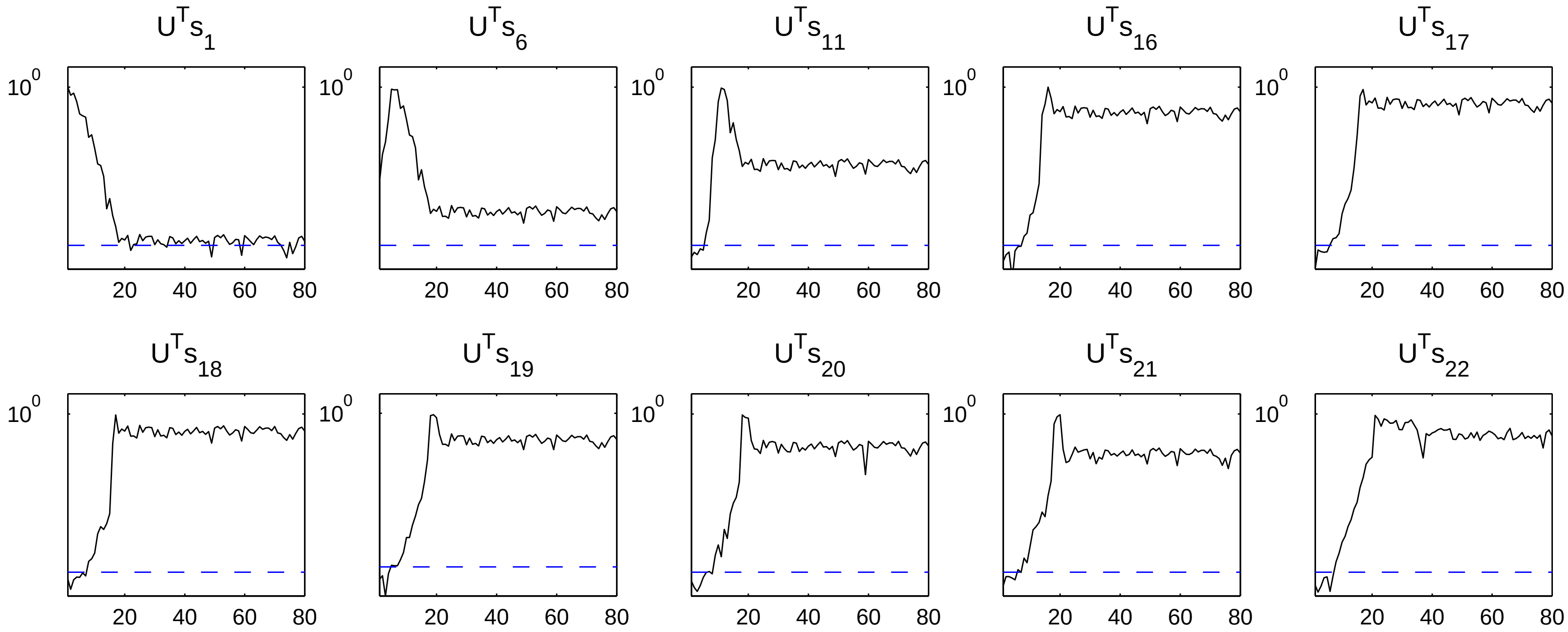
Consider the problem Shaw from [Hansen – Regularization Tools] (computed via **[A,b_exact,x] = shaw(400)**) with a noisy right-hand side (the noise was artificially added using the MATLAB function **randn**). As an example we set

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

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



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



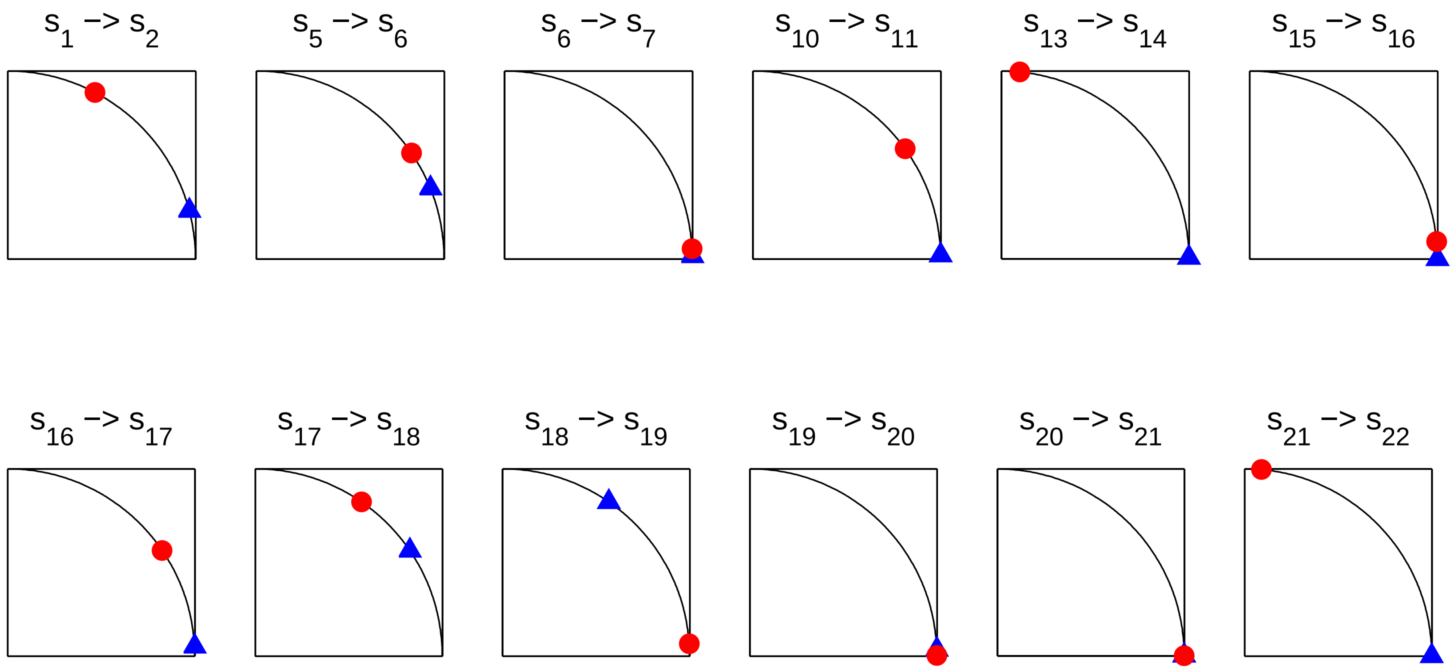
Using the three-term recurrences,

$$\beta_2 \alpha_1 s_2 = \alpha_1 (Aw_1 - \alpha_1 s_1) = AA^T s_1 - \alpha_1^2 s_1,$$

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 substracted 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 at some point the low frequency information is projected out (iteration 18).

Signal space – noise space diagrams:



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

The noise is amplified with the ratio α_k/β_{k+1} that on average (rapidly) grows with k, see [H., Plešinger, Strakoš - 09].

Using the SVD of $A = U\Sigma V^T$, GK gives for the spectral components

$$\begin{aligned} \alpha_1 \,(V^T w_1) &= \Sigma \,(U^T s_1)\,, \\ \beta_2 \,(U^T s_2) &= \Sigma \,(V^T w_1) - \alpha_1 \,(U^T s_1) \\ &= (1/\alpha_1 \,\Sigma^2 - \alpha_1)\,(U^T s_1)\,, \end{aligned}$$

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

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

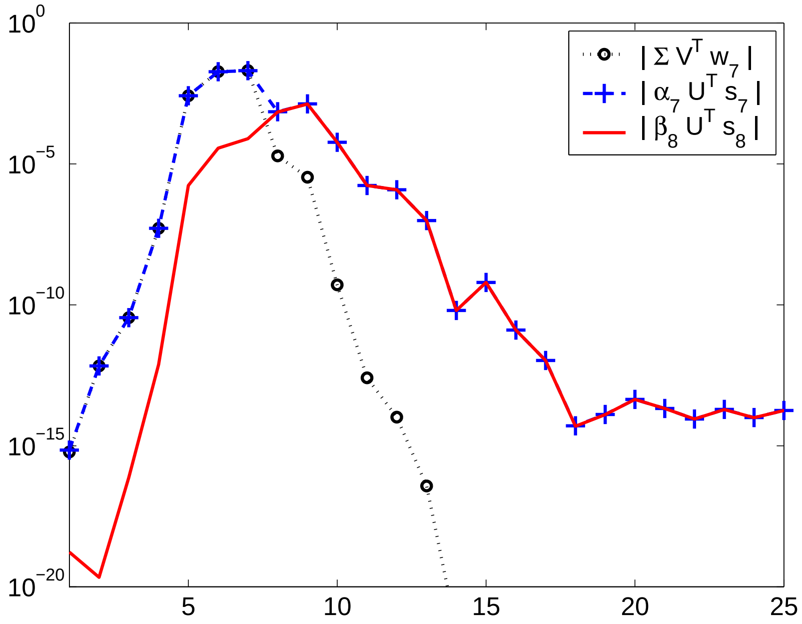
Since dominance in $\Sigma(U^T s_k)$ and $(V^T w_{k-1})$ is shifted by one component, in $\alpha_k\,(V^T w_k) = \Sigma\,(U^T s_k) - \beta_k\,(V^T w_{k-1})$, one can not expect a significant cancelation, and therefore

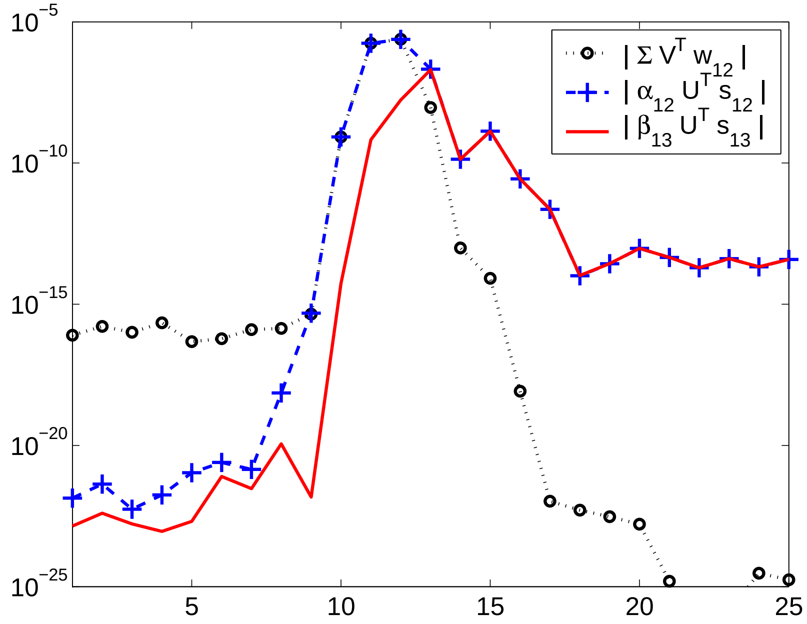
$$\alpha_k \approx \beta_k\,.$$

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 cancelation, and one can expect

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

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 $k = 7$, $\beta_8/\alpha_7 = 0.0524$ (left) and for $k = 12$, $\beta_{13}/\alpha_{12} = 0.677$ (right), Shaw with the noise level $\delta_{\text{noise}} = 10^{-14}$:





Outline

1. Problem formulation
2. Golub-Kahan iterative bidiagonalization, and approximation of the Riemann-Stieltjes distribution function
3. Propagation of the noise in the Golub-Kahan bidiagonalization
4. **Determination of the noise level**
5. Noise reconstruction and subtraction – a numerical experiment
6. Summary and future work

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, see (DPC). Therefore the large nodes essentially control the behavior of the early stages of the Lanczos process.

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

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

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

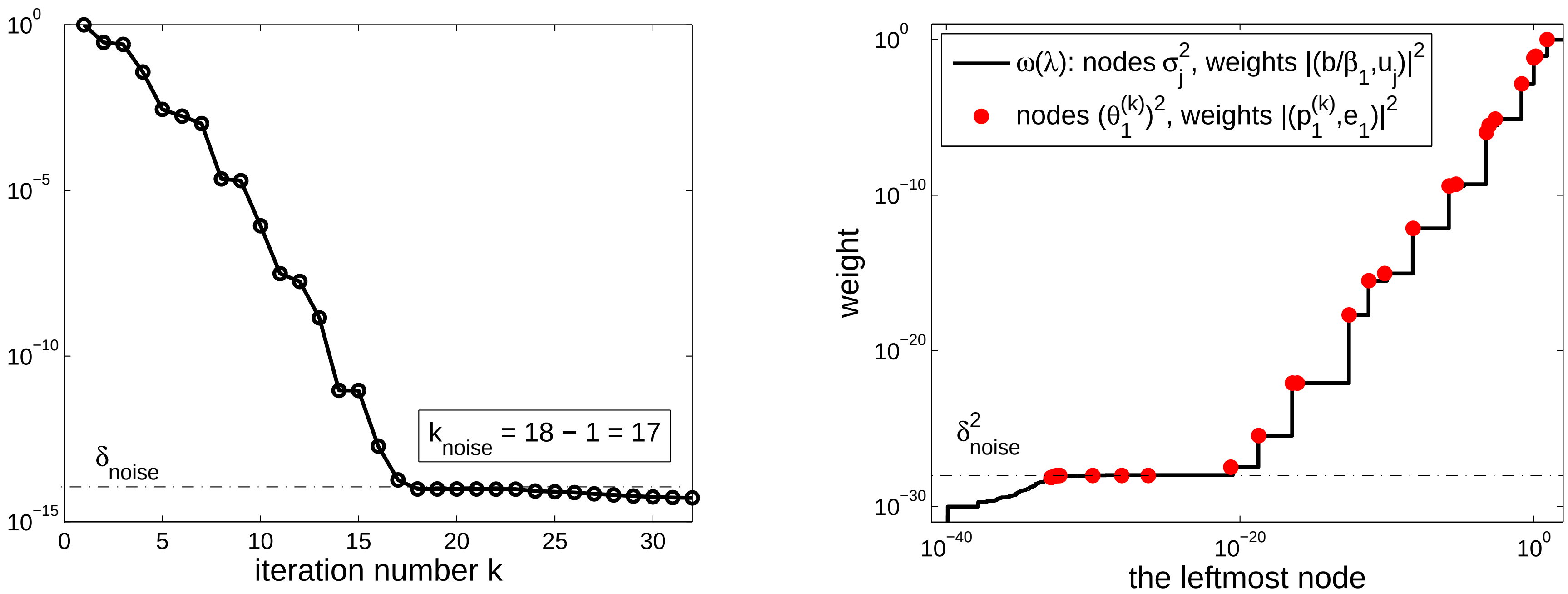
$$\delta^2 \equiv \sum_{j=J_{\text{noise}}}^{n} |(b^{\text{noise}}/\beta_1, u_j)|^2 \approx \delta_{\text{noise}}^2.$$

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 [Karlin, Shapley – 53], [Fischer, Freund – 94]. As k increases, some $(\theta_1^{(k)})^2$ eventually approaches (or becomes smaller than) the node $\sigma_{J_{\text{noise}}}^2$, and its weight becomes

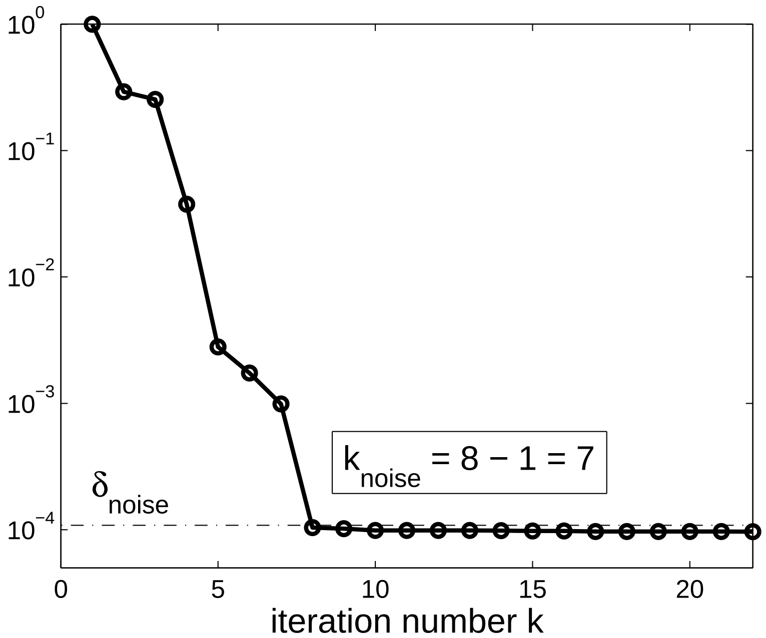
$$|(p_1^{(k)}, e_1)|^2 \approx \delta^2 \approx \delta_{\text{noise}}^2 .$$

The weight $|(p_1^{(k)}, e_1)|^2$ corresponding to the smallest Ritz value $(\theta_1^{(k)})^2$ is strictly decreasing. At some iteration step it sharply starts to (almost) stagnate close to the squared noise level δ_{noise}^2 .

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 with the noise level $\delta_{\text{noise}} = 10^{-14}$:



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 with the noise level $\delta_{\text{noise}} = 10^{-4}$:



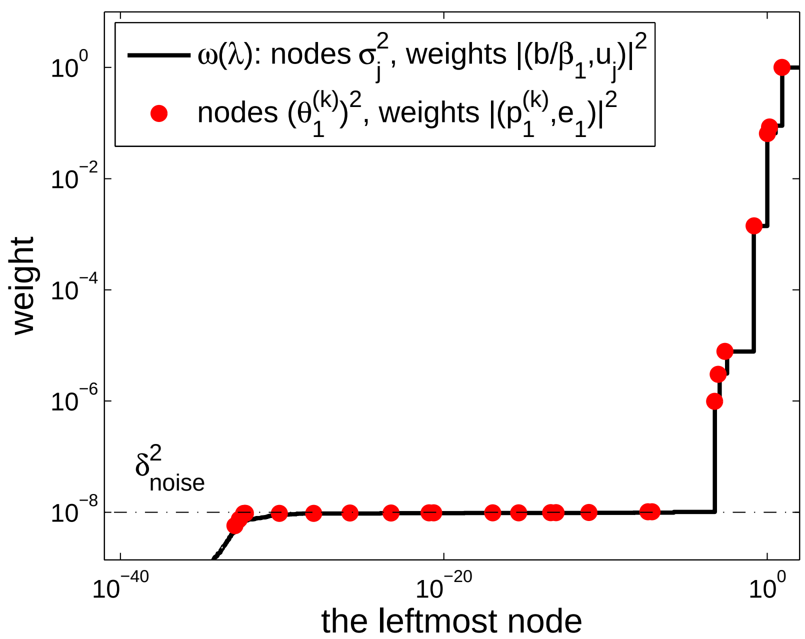
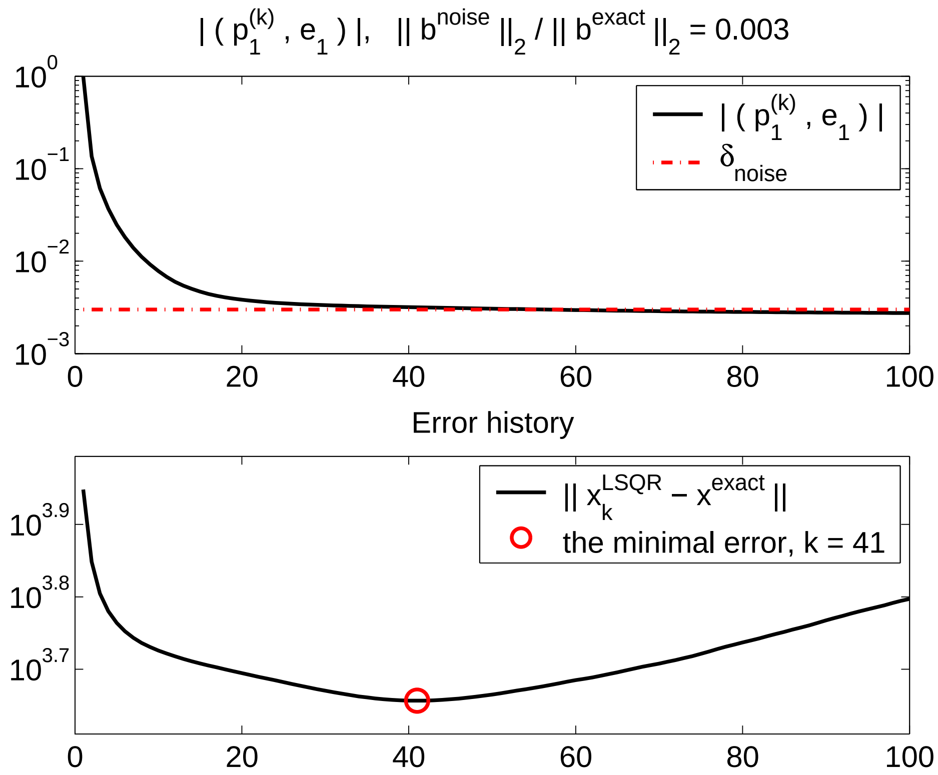


Image deblurring problem: image size 324×470 pixels, problem dimension $n = 152280$, the exact solution (left) and the noisy right-hand side (right), $\delta_{\text{noise}} = 3 \times 10^{-3}$:

x^{exact}

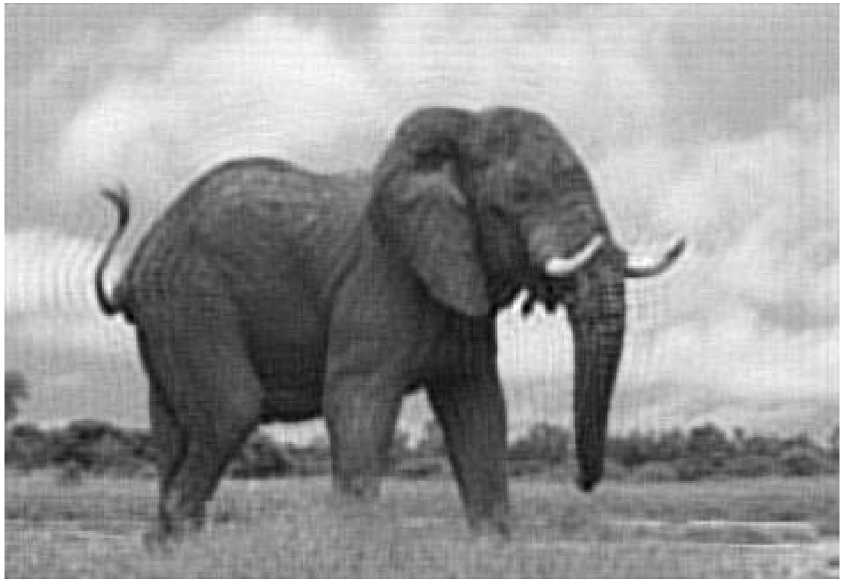
$b^{exact} + b^{noise}$

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

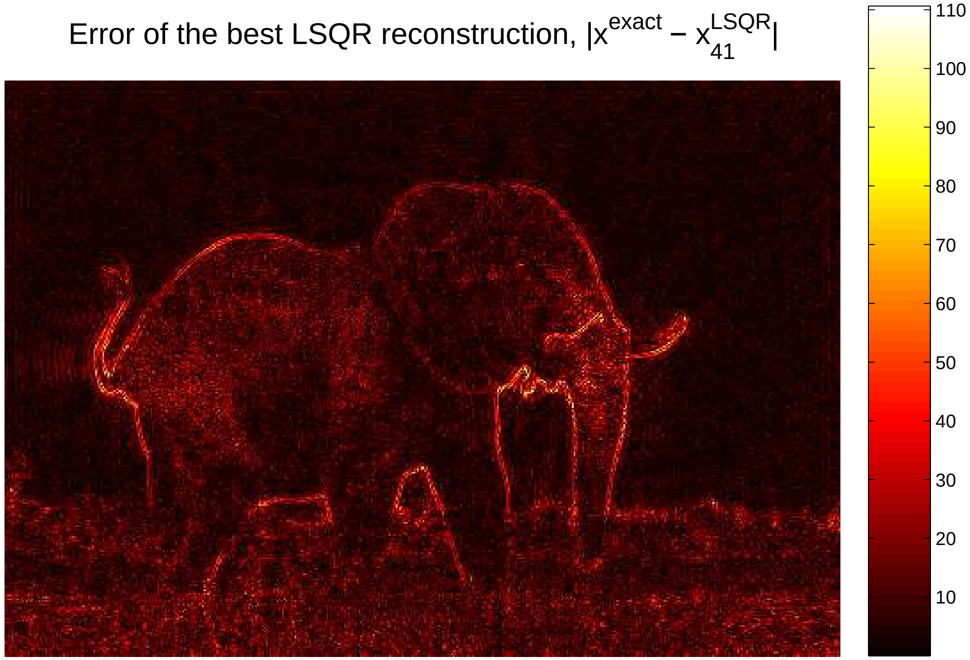


The best LSQR reconstruction (left), x_{41}^{LSQR},
and the corresponding componentwise error (right).
GK without any reorthogonalization!

LSQR reconstruction with minimal error, x_{41}^{LSQR}



Error of the best LSQR reconstruction, $|x^{\text{exact}} - x_{41}^{\text{LSQR}}|$



Outline

1. Problem formulation
2. Golub-Kahan iterative bidiagonalization, and the approximation of the Riemann-Stieltjes distribution function
3. Propagation of the noise in the Golub-Kahan bidiagonalization
4. Determination of the noise level
5. **Noise reconstruction and subtraction – a numerical experiment**
6. Summary and future work

Noise reconstruction:

Let k_{noise} be the noise revealing iteration, then

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

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

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

represents an approximation of the unknown noise.

Subtracting the reconstructed noise from the noisy observation vector?

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

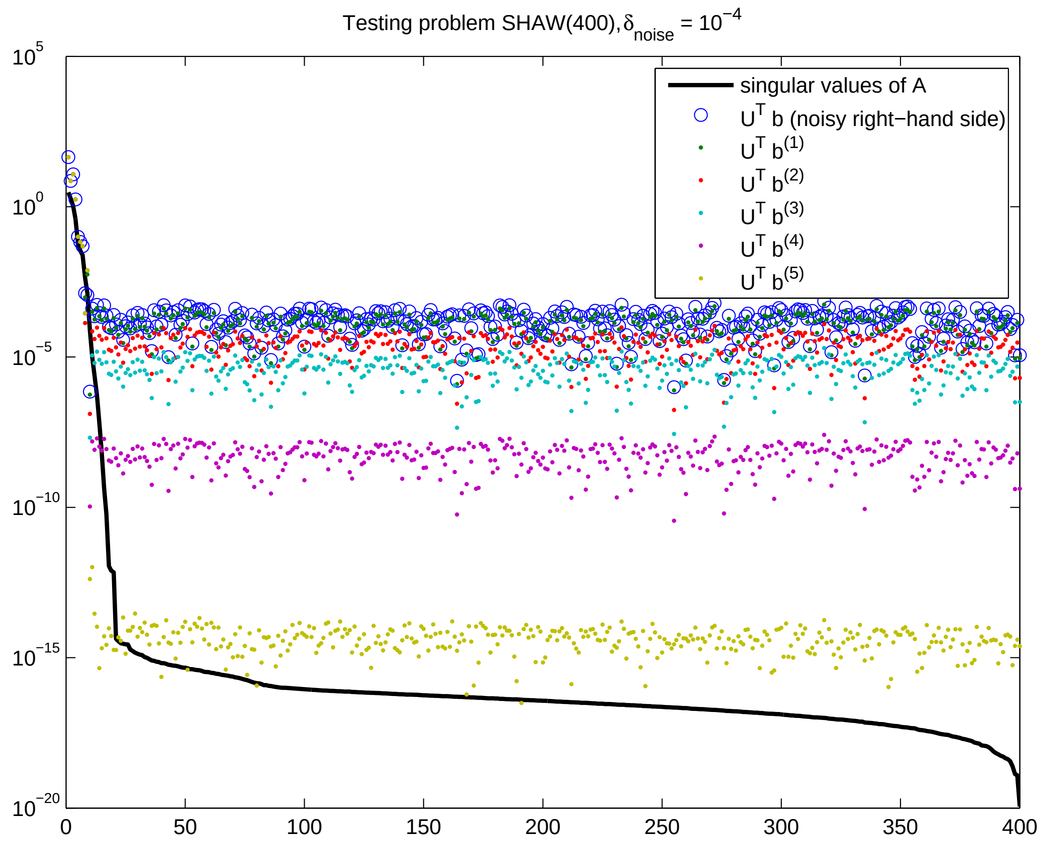
for $j = 1, \ldots, t$
- GK bidiagonalization of A with the starting vector $b^{(j-1)}$;
- identification of the noise revealing iteration k_{noise};
- $\delta^{(j-1)} := |(p_1^{(k_{\text{noise}})}, e_1)|$;
- $b^{\text{noise},(j-1)} := \beta_1 \, \delta^{(j-1)} \, s_{k_{\text{noise}}}$; `// noise approximation`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

end;

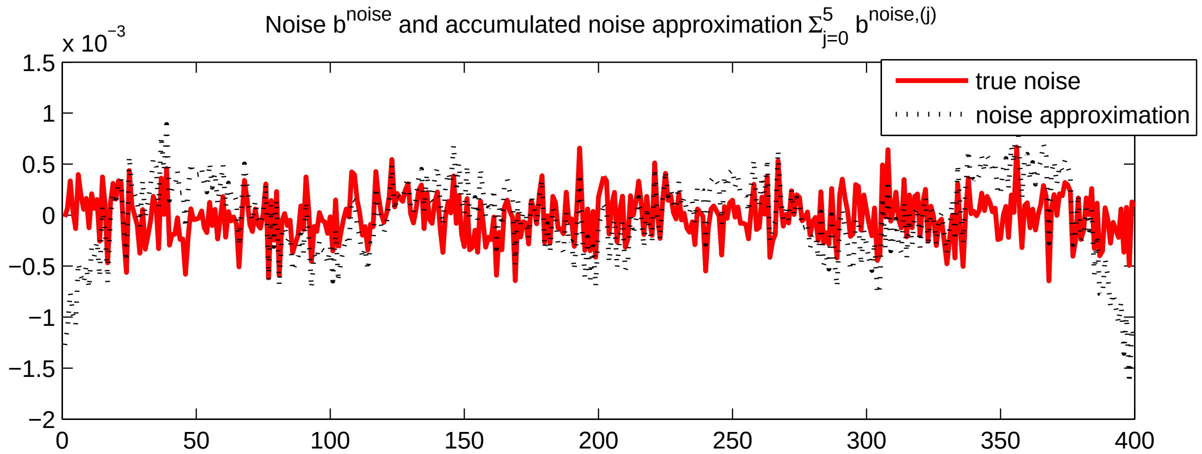
The **accumulated noise approximation** is

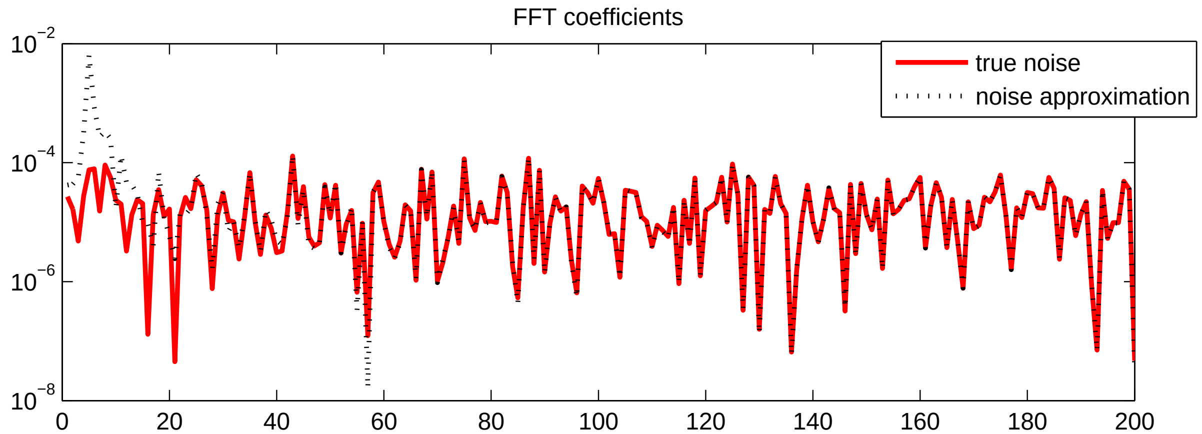
$$\hat{b}^{\text{noise}} \equiv \sum_{j=0}^{t-1} b^{\text{noise},(j)} \, .$$

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

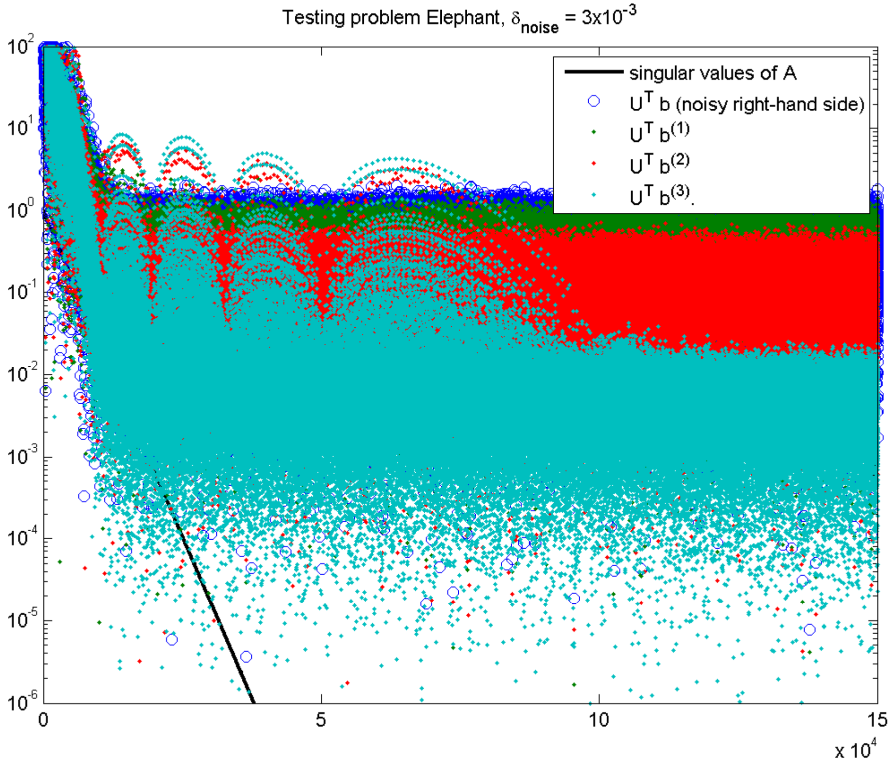


Individual components (top) and Fourier coeffs. (bottom) of $\widehat{b}^{\text{noise}}$, Shaw with the noise level $\delta_{\text{noise}} = 10^{-4}$:





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 $\delta_{\text{noise}} = 3 \times 10^{-3}$:



(k_{noise} corresponds to the best LSQR approximation of x)

Outline

1. Problem formulation
2. Golub-Kahan iterative bidiagonalization, and the approximation of the Riemann-Stieltjes distribution function
3. Propagation of the noise in the Golub-Kahan bidiagonalization
4. Determination of the noise level
5. Noise reconstruction and subtraction –a numerical experiment
6. **Summary and future work**

Message:

Using GK, information about the noise can be obtained in a straight-forward way.

Future work:

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

References

- Golub, Kahan: *Calculating the singular values and pseudoinverse of a matrix*, SIAM J. B2, 1965.
- Hansen: *Rank-deficient and discrete ill-posed problems*, SIAM Monographs Math. Modeling Comp., 1998.
- Hansen, Kilmer, Kjeldsen: *Exploiting residual information in the parameter choice for discrete ill-posed problems*, BIT, 2006.
- Hnětynková, Strakoš: *Lanczos tridiag. and core problem*, LAA, 2007.
- Meurant, Strakoš: *The Lanczos and CG algorithms in finite precision arithmetic*, Acta Numerica, 2006.
- Paige, Strakoš: *Core problem in linear algebraic systems*, SIMAX, 2006.
- Rust: *Truncating the SVD for ill-posed problems*, Technical Report, 1998.
- Rust, O'Leary: *Residual periodograms for choosing regularization parameters for ill-posed problems*, Inverse Problems, 2008.
- Hnětynková, Plešinger, Strakoš: *The regularizing effect of the Golub-Kahan iterative bidiagonalization and revealing the noise level*, BIT, 2009.
- ...

Thank you for your kind attention!