Golub-Kahan iterative bidiagonalization and determining the noise level in the data

Iveta Hnětynková, Marie Michenková, Martin Plešinger,
Zdeněk Strakoš

Charles University in Prague

Technical University Liberec

Academy of Sciences of the Czech republic, Prague

TU Liberec & Preciosa, a.s., September 2013

Six tons large real world ill-posed problem:

Solving large scale discrete ill-posed problems is frequently based upon **orthogonal projections-based model reduction** using Krylov subspaces, see, e.g., hybrid methods. This can be viewed as

approximation of a Riemann-Stieltjes distribution function
via matching moments.

Outline

1. **Problem formulation**
2. Propagation of noise in the Golub-Kahan iterative bidiagonalization
3. Numerical illustration

The underlying problem is a linear algebraic system

$$Ax \approx b$$

which can arise, e.g., in discretization of a Fredholm integral equation of the 1st kind

$$b(\mathbf{s})^{\text{exact}} = \int K(\mathbf{s},\mathbf{t})\, x(\mathbf{t})\, d\mathbf{t} \equiv \mathcal{A}\, x(\mathbf{t}).$$

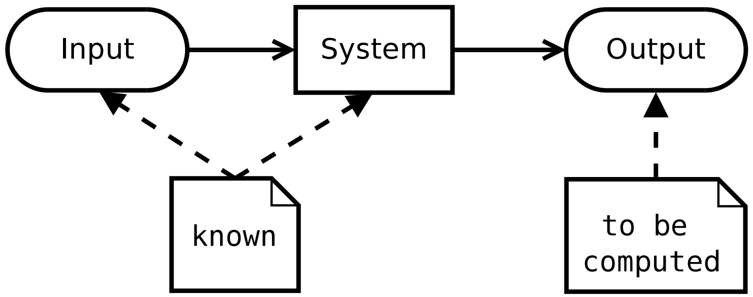
The right-hand side b is contaminated by **noise**

$$b \,=\, b^{\text{exact}} + b^{\text{noise}}, \qquad \delta_{\text{noise}} \,\equiv\, \frac{\|b^{\text{noise}}\|}{\|b^{\text{exact}}\|} \,\ll\, 1\,.$$

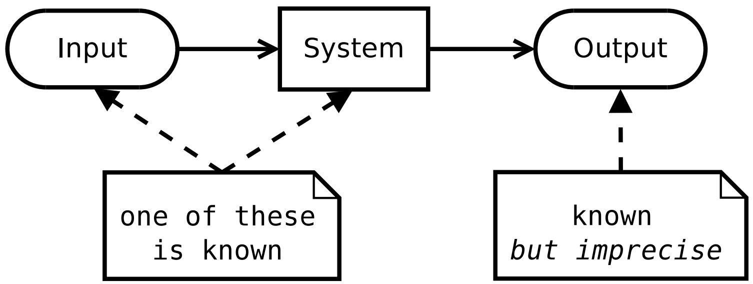
The goal is to approximate

$$x^{\text{exact}} \,\equiv\, A^{\dagger}\, b^{\text{exact}}.$$

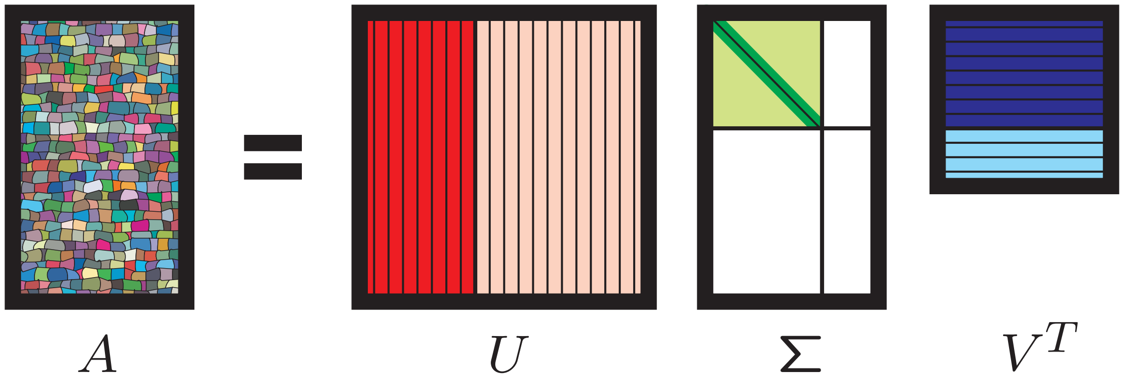
Forward Problem



Inverse Problem



Singular value decomposition in discrete ill-posed problems



$$
\begin{array}{ccccc}
 & A & & A^* & \\
v_1 & \xrightarrow{\sigma_1} & u_1 & \xrightarrow{\sigma_1} & v_1\,, \\
v_2 & \xrightarrow{\sigma_2} & u_2 & \xrightarrow{\sigma_2} & v_2\,, \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
v_r & \xrightarrow{\sigma_r} & u_r & \xrightarrow{\sigma_r} & v_r\,,
\end{array}
$$

$$
\left.\begin{array}{c} v_{r+1} \\ \vdots \\ v_m \end{array}\right\} \longrightarrow \approx 0\,, \qquad
\left.\begin{array}{c} u_{r+1} \\ \vdots \\ u_n \end{array}\right\} \longrightarrow \approx 0\,.
$$

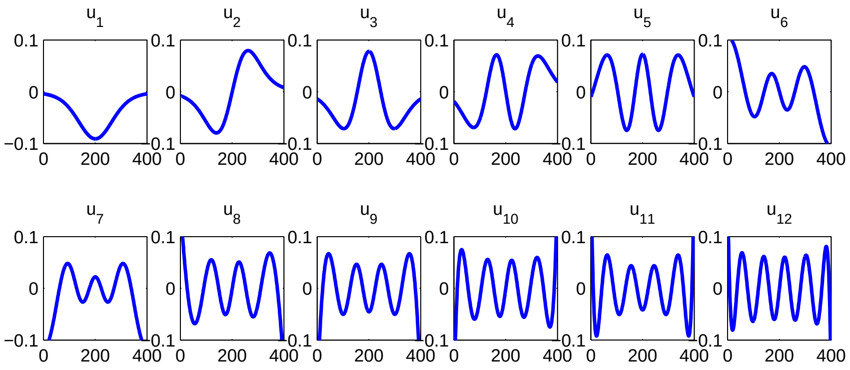
Properties (assumptions):

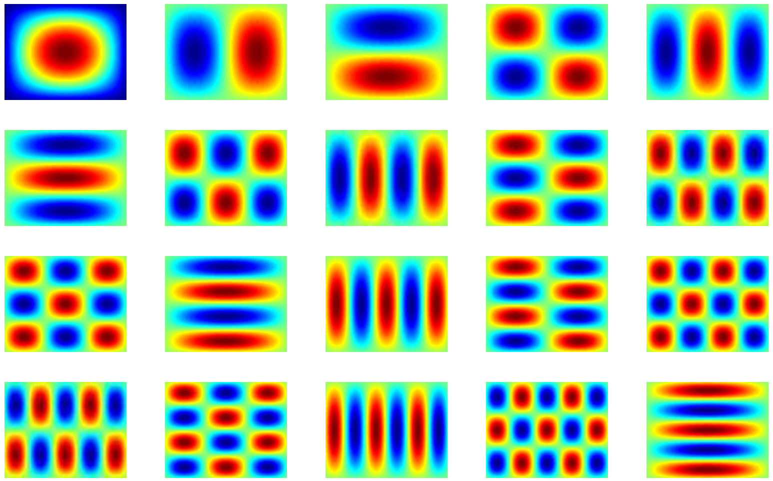
- matrices A, A^T, AA^T have a smoothing property;
- left singular vectors u_j of A represent increasing frequencies as j increases;
- the system $A\,x^{\text{exact}} = b^{\text{exact}}$ satisfies the discrete Picard condition.

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, $j = 1, \dots, n$.

Left singular vectors of A represent a basis with increasing frequencies; reshaped right singular vectors of A (singular images) for the Gaussian blur





Using the SVD the solution of $Ax = b$ can be written as

$$x = A^{-1}b = V\Sigma^{-1}U^Tb = \sum_{j=1}^{N} \frac{u_j^T b}{\sigma_j} v_j.$$

Recall that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_N$ and the exact components and the noise components behave differently,

$$x = \sum_{j=1}^{N} \frac{u_j^T b^{\text{exact}}}{\sigma_j} v_j + \sum_{j=1}^{N} \frac{u_j^T b^{\text{noise}}}{\sigma_j} v_j .$$

Outline

1. Problem formulation
2. **Propagation of noise in the Golub-Kahan bidiagonalization**
3. Numerical illustration

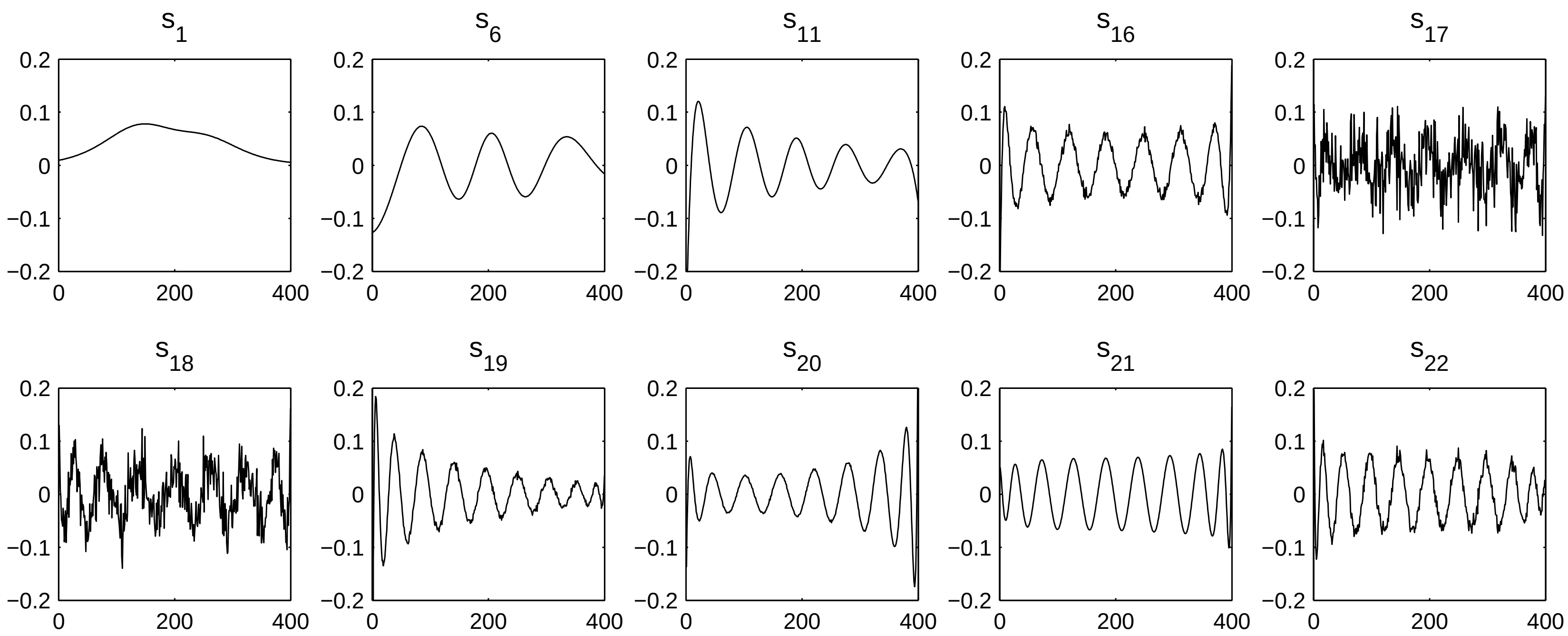
Krylov subspace methods are projection methods. Golub-Kahan iterative bidiagonalization (GK) of A:

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

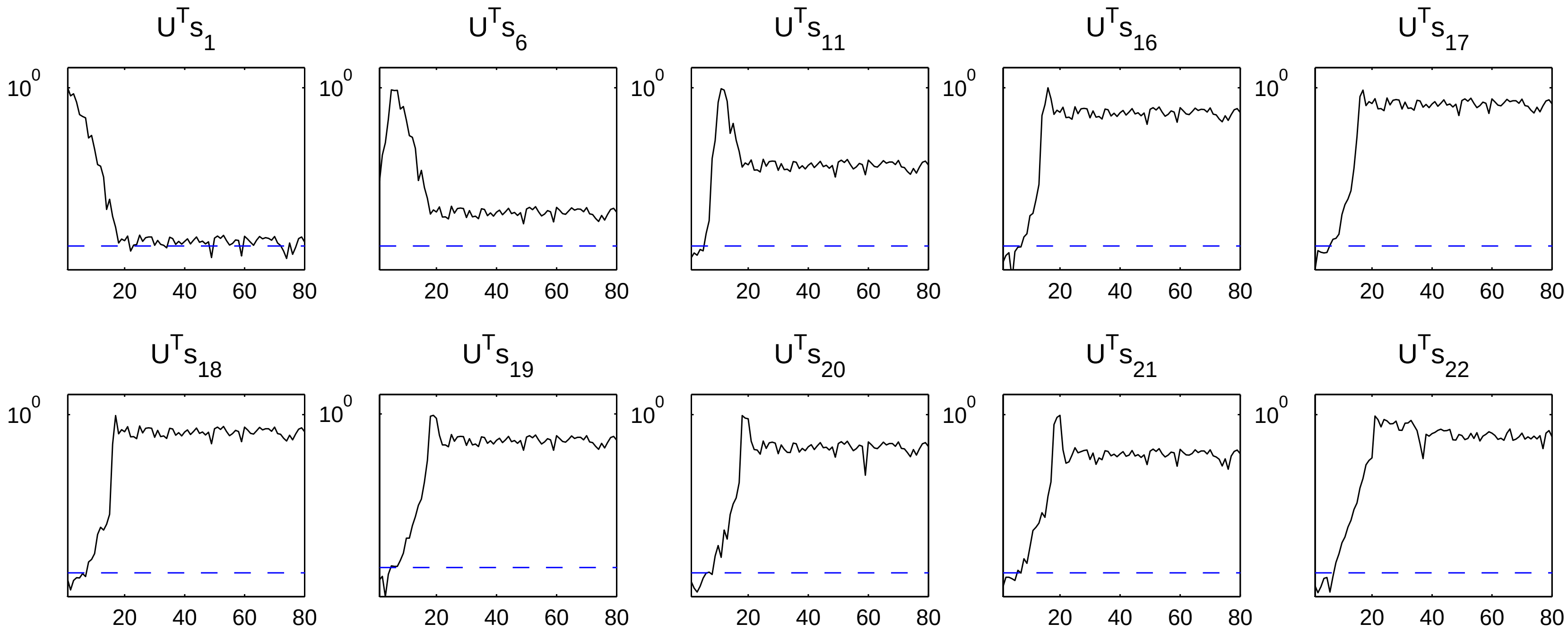
$$\begin{aligned} \alpha_j \, w_j &= A^T s_j - \beta_j \, w_{j-1} , & \|w_j\| = 1 , \\ \beta_{j+1} \, s_{j+1} &= A \, w_j - \alpha_j \, s_j , & \|s_{j+1}\| = 1 . \end{aligned}$$

$S_k = [s_1, \ldots, s_k]$, $W_k = [w_1, \ldots, w_k]$, $S_k^T A W_k \equiv L_k$, where L_k is lower bidiagonal, S_k and W_k have orthonormal columns. 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 estimate the noise level?

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:



Noise is amplified with the ratio α_k/β_{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 $k = 2, 3, \ldots$

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

Since the 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 the 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

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

Noise level estimation :

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

Spectral properties of

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

determine an approximation of the Riemann-Stieltjes distribution function related to the original mapping $\mathcal{A}$.

Consider the SVD of the bidiagonal matrix

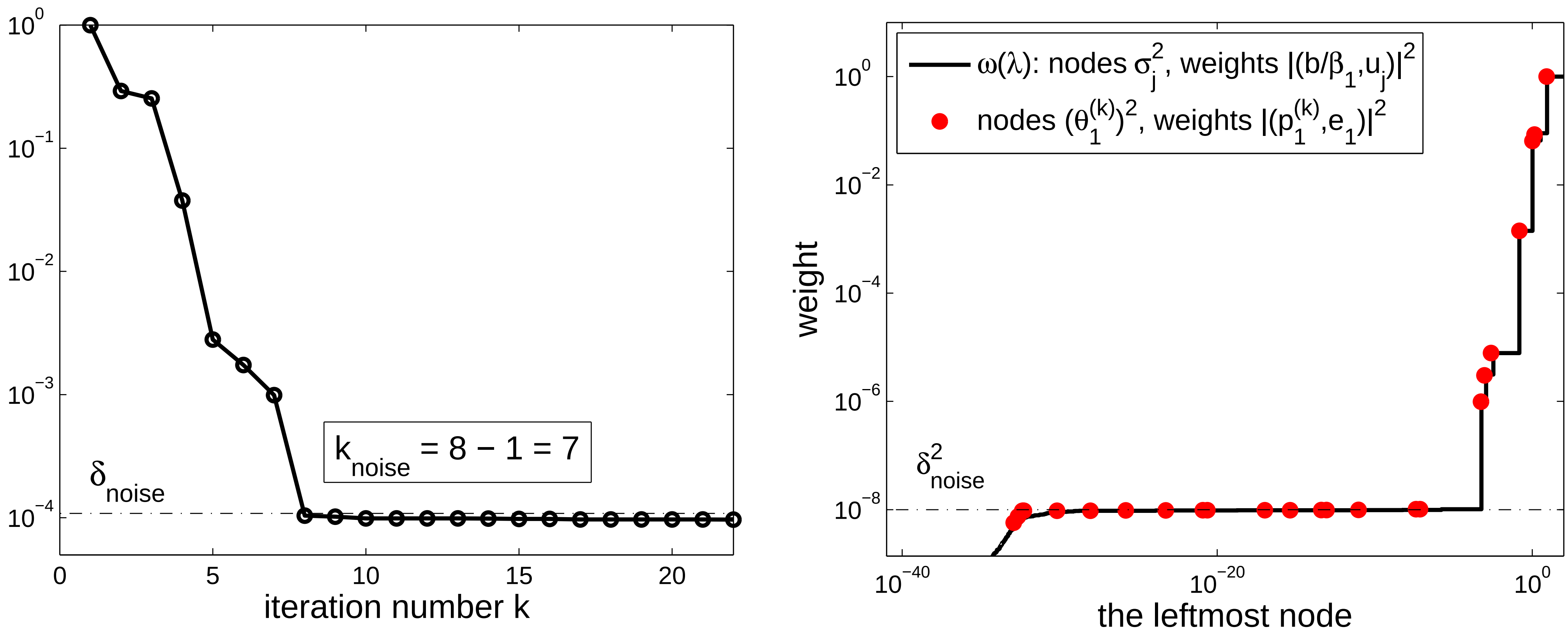
$$L_k = P_k \, \Theta_k \, Q_k{}^T,$$

$$P_k = [p_1^{(k)}, \ldots, p_k^{(k)}] \,, \; Q_k = [q_1^{(k)}, \ldots, q_k^{(k)}] \,, \; \Theta_k = \mathrm{diag}\,(\theta_1^{(k)}, \ldots, \theta_n^{(k)}) \,,$$

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

The weight $|(p_1^{(k)}, e_1)|^2$ of the approximate distribution function corresponding to the smallest $(\theta_1^{(k)})^2$ is strictly decreasing. At the so called noise revealing iteration, it sharply starts to (almost) stagnate on the level close to the squared noise level $\delta_{\mathrm{noise}}^2$.

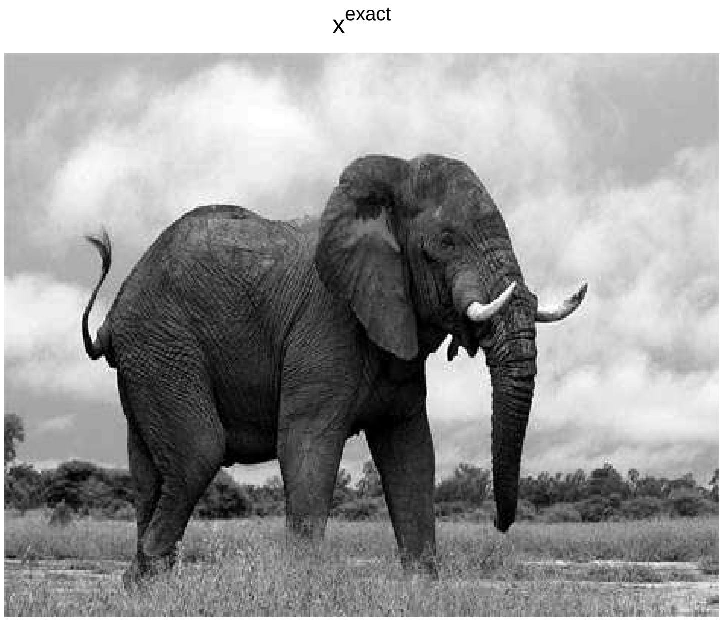
Square roots of the weights (left), approximation of $\omega(\lambda)$ (right):



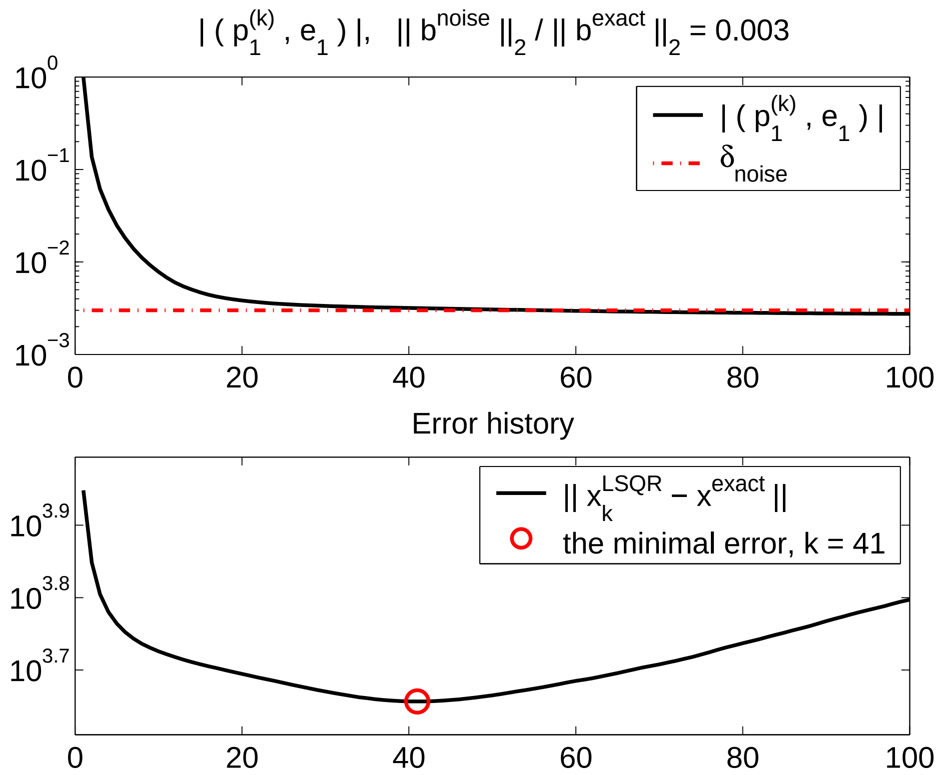
Outline

1. Problem formulation
2. Propagation of noise in the Golub-Kahan bidiagonalization
3. **Numerical illustration**

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

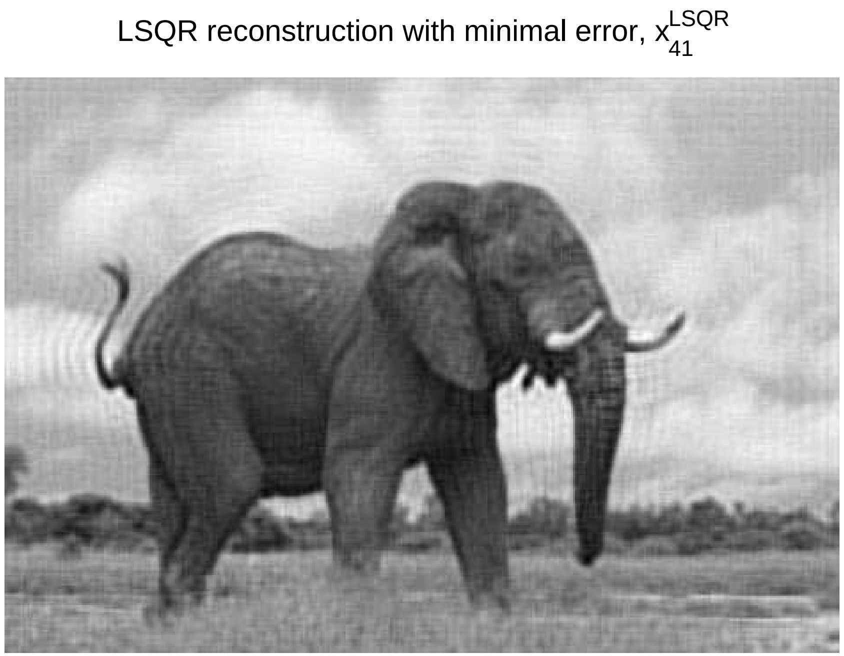


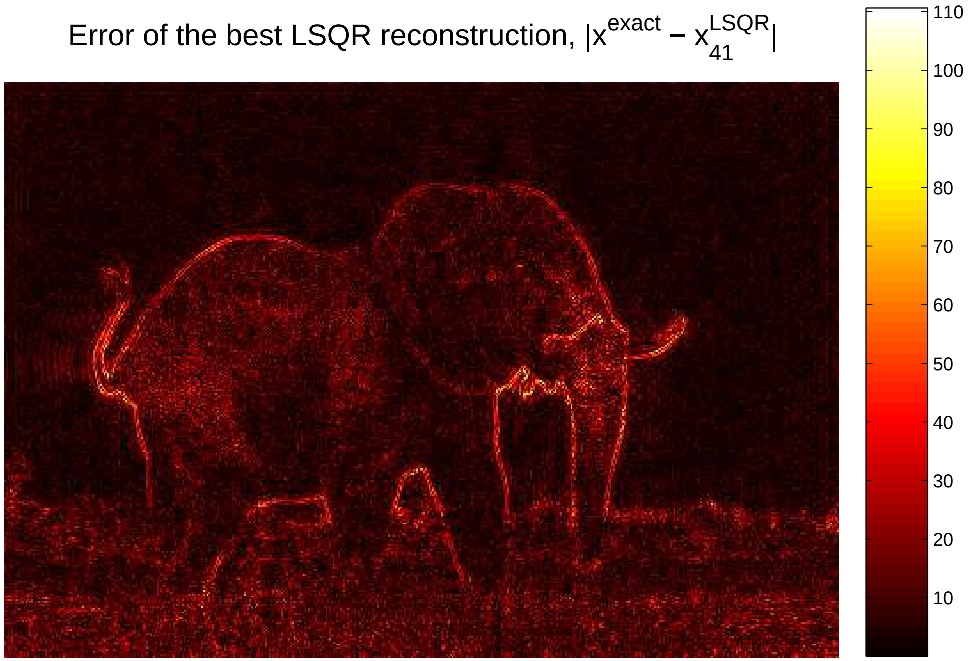
Square roots of the weights $|(p_1^{(k)}, e_1)|^2$, $k = 1, 2, \ldots$ (top) and error history of LSQR solutions (bottom):



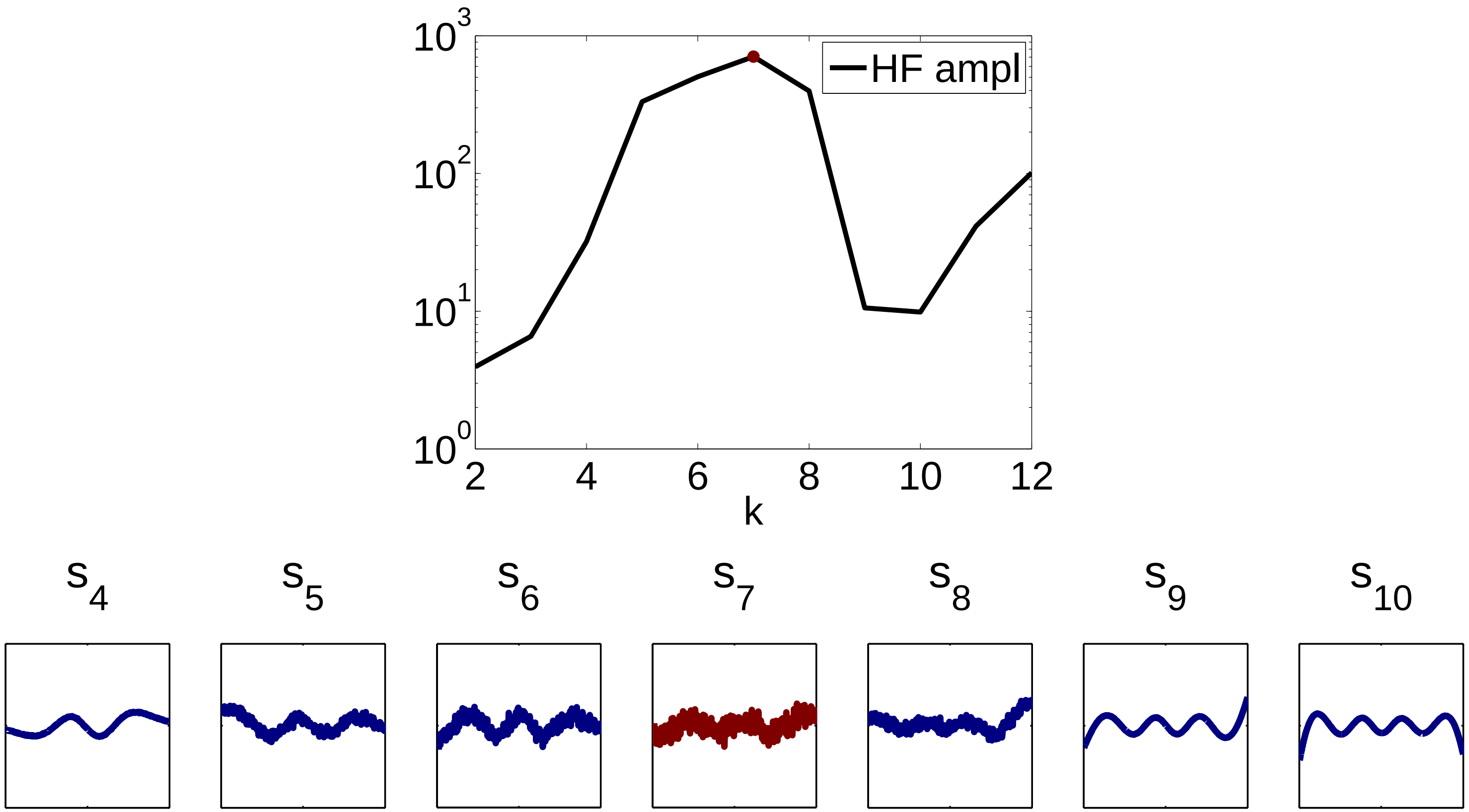
The best LSQR reconstruction (left), x_{41}^{LSQR},
and the corresponding componentwise error (right).

GK without any reorthogonalization!

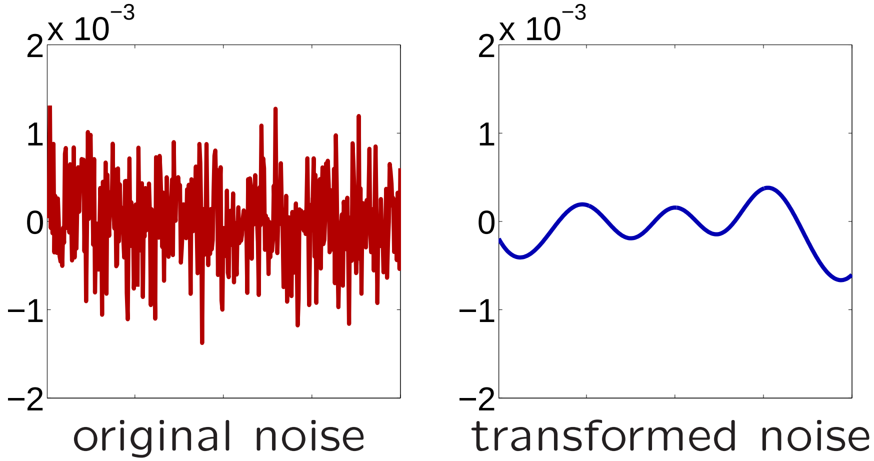




Denoising, problem SHAW(400), maximal amplification factor



Denoising?



Subtraction of the approximate noise from the data leads to the remaining much smoother "transformed noise"

Main message :

Whenever you see a blurred elephant which is a bit too noisy,
the best thing is to apply, as quickly as possible,
the GK iterative bidiagonalization.

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.
- Meurant, Strakoš: *The Lanczos and CG algorithms in finite precision arithmetic*, Acta Numerica, 2006.
- Hnětynková, Strakoš: *Lanczos tridiag. and core problem*, LAA, 2007.
- Hnětynková, Plešinger, Strakoš: *The regularizing effect of the Golub-Kahan iterative bidiagonalization and revealing the noise level*, BIT, 2009.
- Michenková: *Regularization techniques based on the least squares method*, diploma thesis, MFF UK, 2013.
- Liesen, Strakoš: *Krylov Subspace Methods, Principles and Analysis.* Oxford University Press, 2013.
- ...

Thank you for your kind attention!