Golub-Kahan bidiagonalization and revealing the noise level in data

Iveta Hnětynková *,**, Martin Plešinger *, Zdeněk Strakoš *,**

** Academy of Sciences of the Czech republic, Institute of Computer Science*

*** Charles University in Prague, Faculty of Mathematics and Physics*

Hamburg – September 2008

Outline

1. Problem formulation
2. Regularization by Golub-Kahan bidiagonalization
3. Revealing the noise level in data
4. Stopping criteria
5. Summary

1. Problem formulation

Consider an ill-posed linear system

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

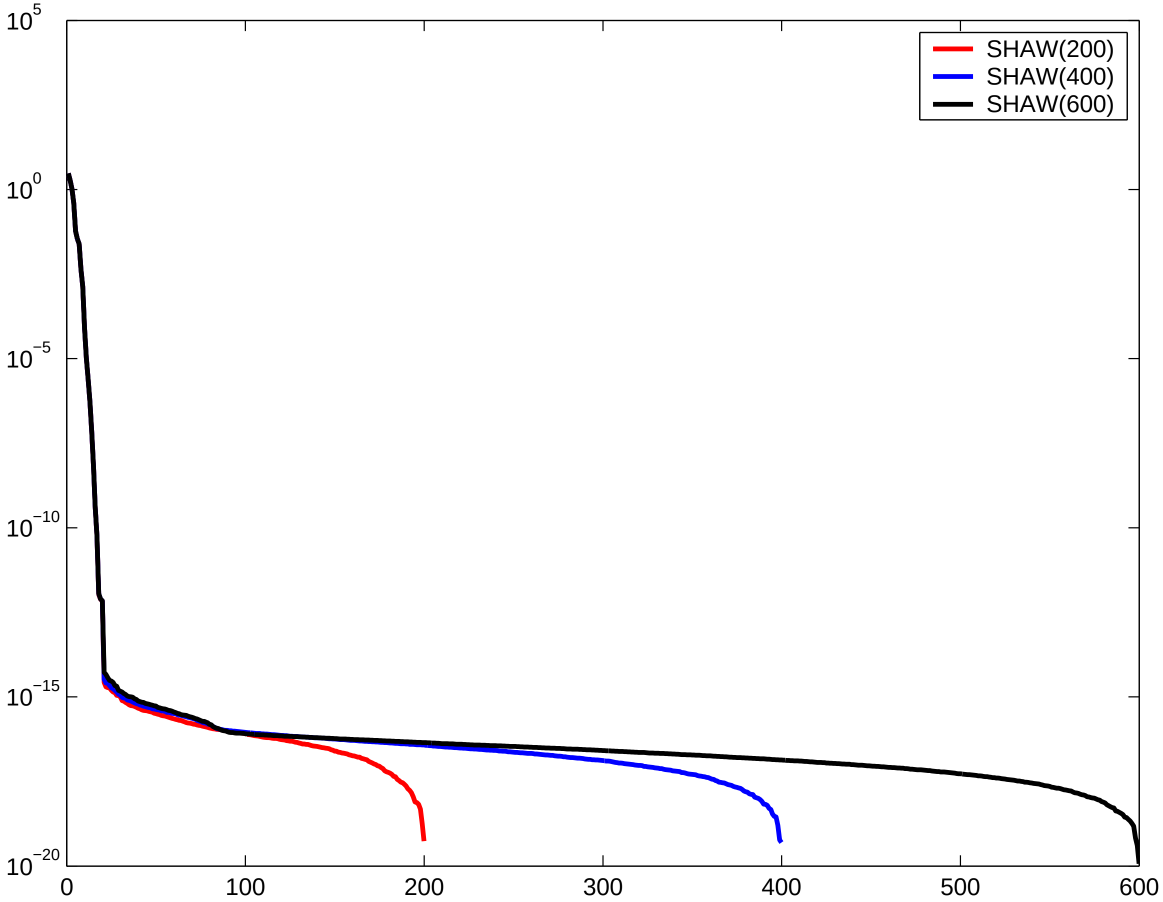
with a white noise contaminated right-hand side

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

Possible difficulties:

- the noise component b^{noise} is unknown;
- the rank of A is not well defined (singular values of A decay gradually to zero);
- the solution is sensitive on small perturbations in data.

Singular values of A decay gradually to zero:



Denote $l = \text{rank}(A)$. Consider the singular value decomposition

$$A = \tilde{U}\,\tilde{\Sigma}\,\tilde{V}^T = \sum_{j=1}^{l} \tilde{u}_j \tilde{\sigma}_j \tilde{v}_j^T\,,$$

$$\tilde{U} = [\tilde{u}_1, \ldots, \tilde{u}_l], \quad \tilde{V} = [\tilde{v}_1, \ldots, \tilde{v}_l], \quad \tilde{\Sigma} = \text{diag}(\tilde{\sigma}_1, \ldots, \tilde{\sigma}_l)\,.$$

The least squares method (LS) minimizes $\|b - Ax\|$ and

$$\begin{aligned} x^{LS} &= \sum_{j=1}^{l} \frac{\tilde{u}_j^T b}{\tilde{\sigma}_j}\,\tilde{v}_j \\ &= \sum_{j=1}^{l} \frac{\tilde{u}_j^T b^{exact}}{\tilde{\sigma}_j}\,\tilde{v}_j + \sum_{j=1}^{l} \frac{\tilde{u}_j^T b^{noise}}{\tilde{\sigma}_j}\,\tilde{v}_j\,. \end{aligned}$$

Thus components of the solution corresponding to small singular values may be dominated by errors in b, the solution is meaningless.

Regularization methods are used to suppress the effect of errors in the data and extract the essential information about the system, e.g.,

- truncated SVD, truncated total least squares, Tikhonov regularization, see [Hansen, O'Leary – 97], [Fierro, Golub, Hansen, O'Leary – 97], [Hansen - 98], [Golub, Hansen, O'Leary - 99], [Sima, Van Huffel, Golub - 04], [Kilmer, Hansen, Espanol - 06], . . .
- methods based on iterative Golub-Kahan bidiagonalization as LSQR, hybrid methods, see [Paige, Saunders – 82], [Bjorck – 96], [Hansen – 97], [Hanke – 01], [Kilmer, Hansen, Espanol - 06], . . .

2. Regularization by Golub-Kahan bidiagonalization

Consider Golub-Kahan bidiagonalization (GK) of A in the form

$$
\begin{aligned}
& w_0 = 0\,, \quad s_1 = b\,/\,\beta_1\,, \quad \text{where} \quad \beta_1 = \|b\|_2\,, \\
& \text{for} \quad j = 1, 2, 3, \ldots \\
& \quad \alpha_j\, w_j = A^T s_j - \beta_j\, w_{j-1}\,, \qquad \|w_j\| = 1\,, \\
& \quad \beta_{j+1}\, s_{j+1} = A\, w_j - \alpha_j\, s_j\,, \qquad \|s_{j+1}\| = 1\,, \\
& \text{end}\,.
\end{aligned}
$$

Denote $S_k = [s_1, \ldots, s_k]$, $W_k = [w_1, \ldots, w_k]$ resulting matrices with orthonormal columns and

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

Regularization methods based on GK compute the solution in two steps. First the problem is projected on the Krylov subspace using k steps of bidiagonalization (outer regularization), i.e.

$$A\,W_k = S_{k+1}\,L_{k+}\,.$$

Then an inner regularization is applied to the projected problem

$$A\,x \approx b \;\longrightarrow\; L_{k+}\,y \approx \beta_1\,e_1\,.$$

The bidiagonalization is stopped when the regularized solution of the projected problem matches selected stopping criteria, based on the L-curve, estimation of the distance between the exact and regularized solution, the discrepancy principle, cross validation methods etc.

3. Revealing the noise level in data

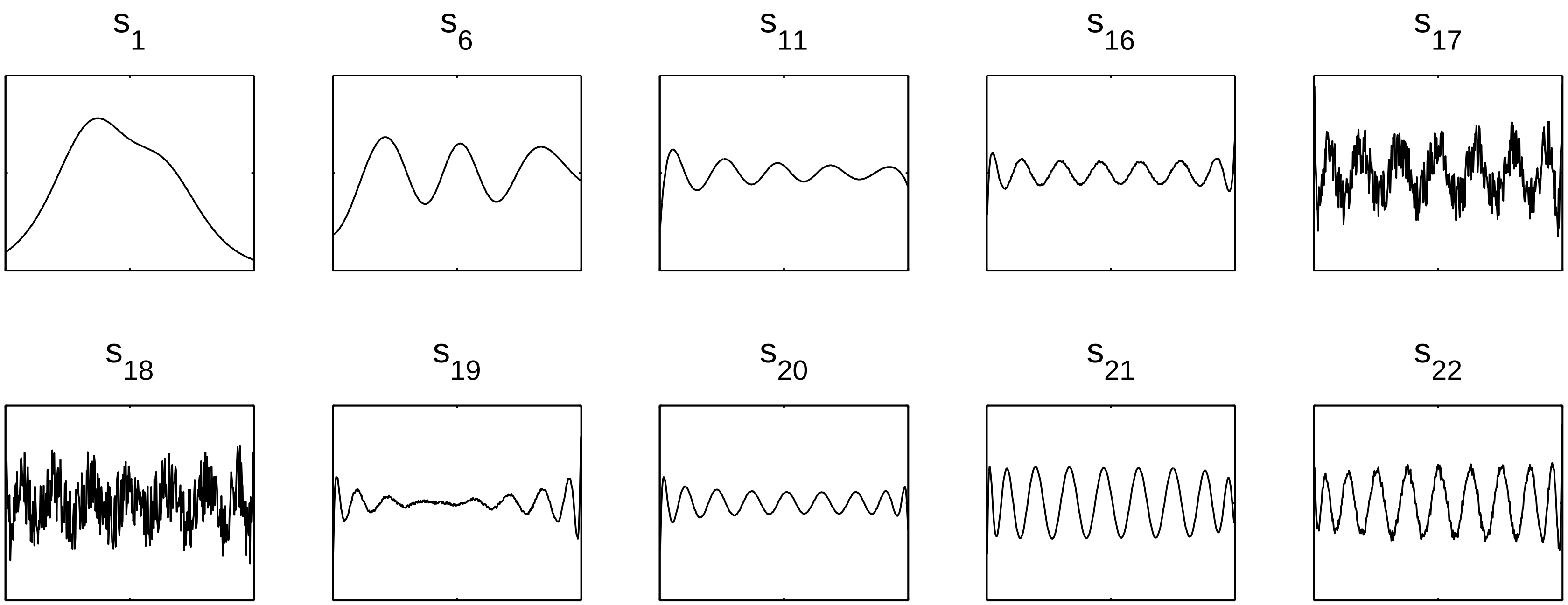
GK starts with the normalized noisy right-hand side $s_1 = b \,/\, \|b\|$, thus vectors s_j has to contain some information about the noise.

Can information about the noise level be obtained by analyzing the vectors s_j generated by GK?

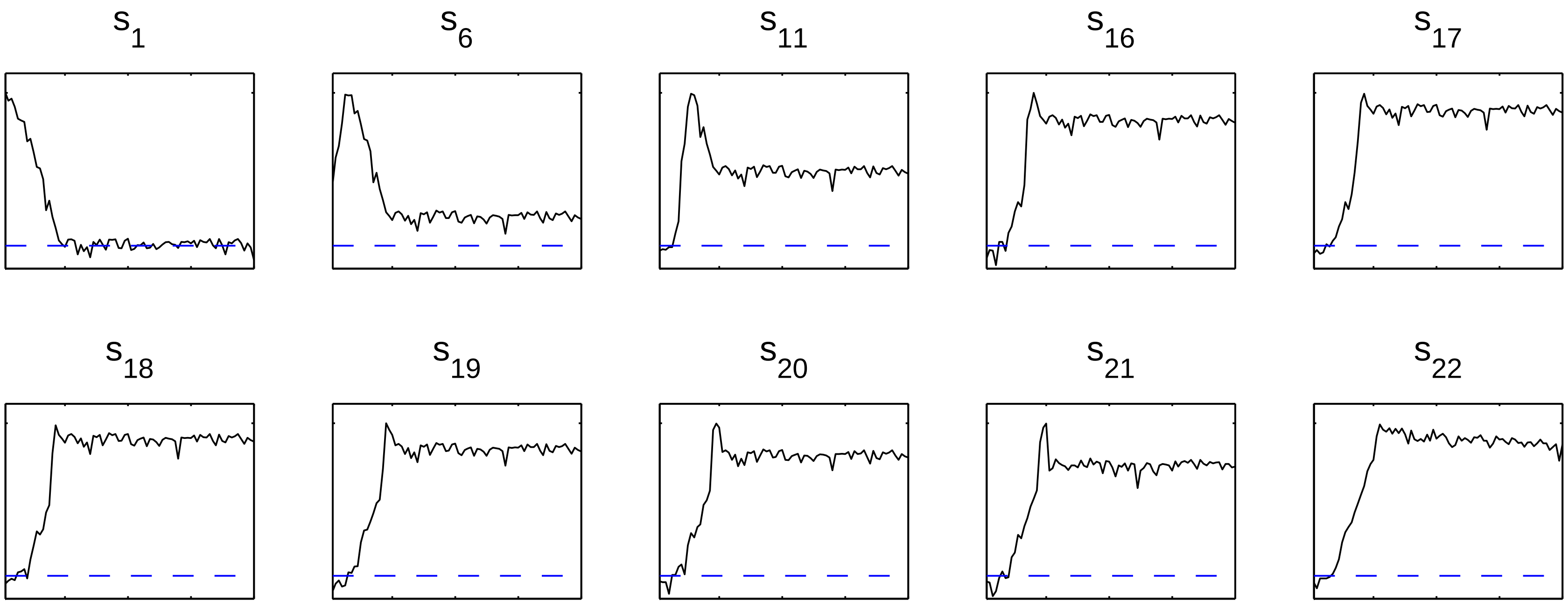
Example: Consider the problem SHAW(400) from [Hansen, RTools] with a noisy right-hand side (the noise was artificially added)

$$46.62 = \| \, b^{exact} \, \| \; \gg \; \| \, b^{noise} \, \| = 10^{-12} \, .$$

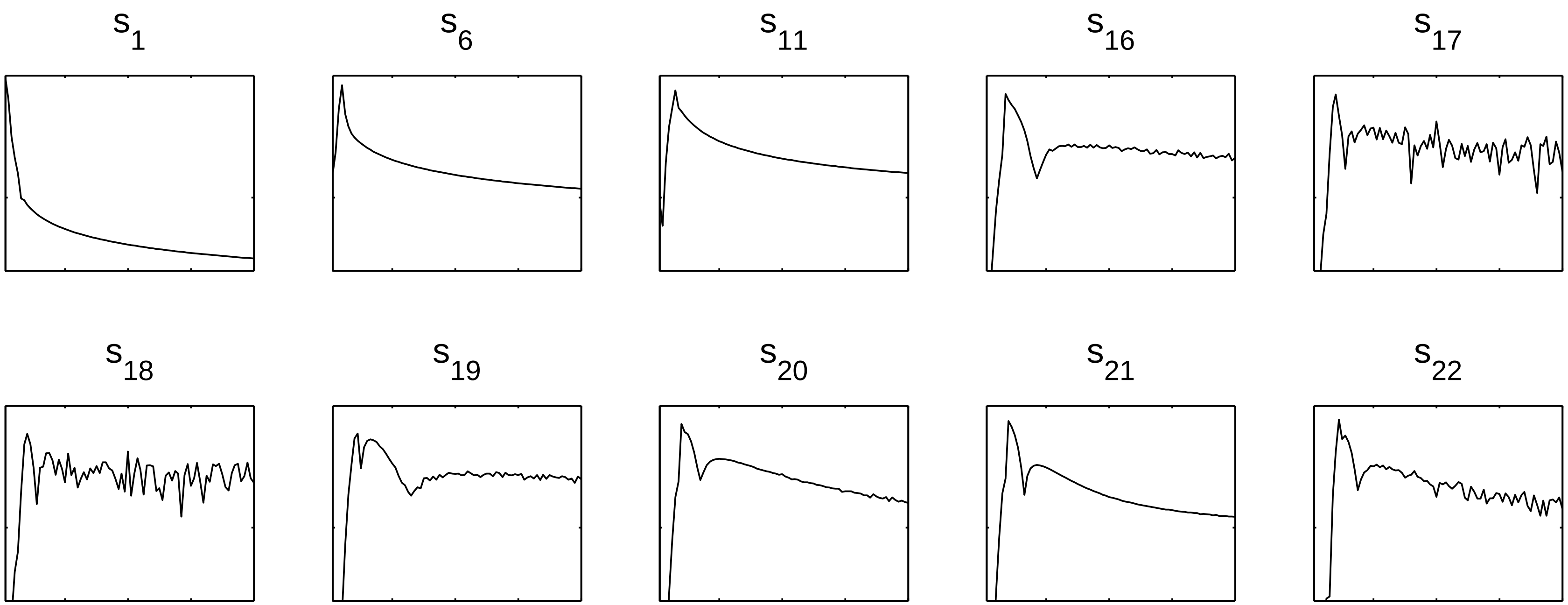
Components of several bidiagonalization vectors s_j:



Fourier coefficients of vectors s_j in the basis of the left singular vectors $\tilde{U}$ of A:



Fourier coefficients of vectors s_j computed by FFT algorithm:



Using the three-term recurrences,

$$\beta_2 s_2 = A w_1 - \alpha_1 s_1 = \frac{1}{\alpha_1} A A^T s_1 - \alpha_1 s_1,$$

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 the low frequency information is projected out.

In the example, vector s_{18} is fully dominated by noise –

the noise is revealed.

(In the 19th step, the noise is partially projected out because vectors s_j has to be mutually orthonormal.)

Now we get explicit information when the noise begins to cover useful information in the data. The solution of the original problem $A\,x \approx b$ computed through the bidiagonal problem

$$L_j\,y = \beta_1\,e_1$$

for $j > k^{noise} = 18$ can be significantly polluted by the noise.

Estimation of the noise level:

By some manipulations with the three-term recurrence for s_j,

$$\|b^{noise}\| \approx \beta_1 \prod_{j=2}^{k^{noise}} \frac{\beta_j}{\alpha_{j-1}}$$

where $\beta_1 = \|b\|$. Thus

$$\delta \equiv \frac{\beta_1}{\|b\|} \prod_{j=2}^{k^{noise}} \frac{\beta_j}{\alpha_{j-1}}$$

is an estimate of the relative noise level in the original data.

Examples: Noise levels in the data, estimate δ and iterations k^{noise}, for the problem SHAW(400)

level	2.71 e–14	2.71 e–10	2.71 e–6	2.71 e–4	2.71 e–2
δ	3.76 e–14	1.90 e–9	3.76 e–6	2.82 e–4	8.02 e–2
k^{noise}	18	16	11	8	4

and ILAPLACE(100,1).

level	4.68 e–13	4.68 e–10	4.68 e–7	4.68 e–2	4.68 e–1
δ	2.61 e–11	2.30 e–10	1.31 e–7	2.09 e–2	6.77 e–1
k^{noise}	23	21	17	7	3

4. Stopping criteria

There are two stages with different behavior.

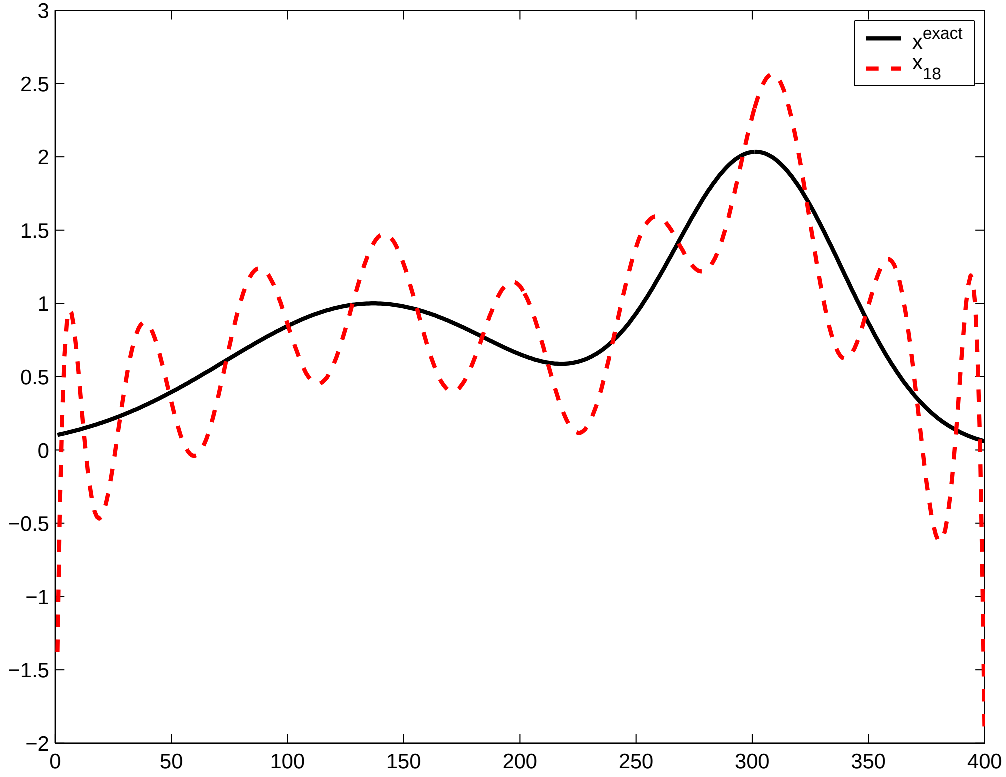
1. Stopping GK for $k \leq k^{noise}$:

   The bidiagonal problem $L_k\, y = \beta_1\, e_1$ can be solved directly, but some information is not absorbed in the problem yet and the result is unsatisfactory.

2. Stopping GK for $k > k^{noise}$:

   The bidiagonal problem inherits a part of the ill-posedness of the original problem, and therefore some form of inner regularization must be applied (idea of hybrid methods).

Exact solution and solution computed directly from the bidiagonal problem with $k = k_{noise}$, for SHAW(400) with $\| b^{noise} \| = 10^{-12}$:



Truncated SVD - regularized solution:

Consider the SVD of the bidiagonal matrix

$$L_k = U_{11} \Sigma_1 V_{11}^T$$

and denote

$$U_k \equiv S_k U_{11} , \quad V_k \equiv W_k V_{11} \quad \in \mathcal{R}^{n \times k} ,$$

then

$$U_k^T A V_k = \Sigma_1 .$$

Moreover we denote

$$U_k = [u_1, \ldots, u_k] , \quad V_k = [v_1, \ldots, v_k] , \quad \Sigma_1 = \text{diag}\,(\sigma_1, \ldots, \sigma_k) .$$

We define a k-r-TSVD solution of $Ax \approx b$ as

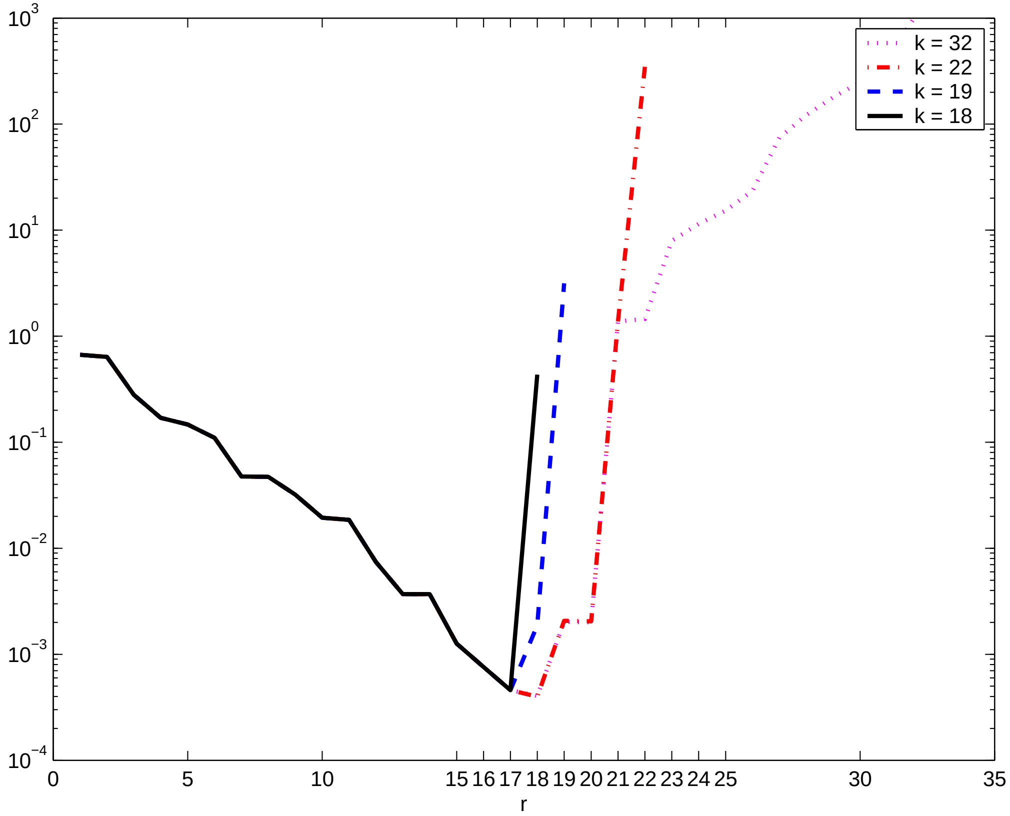
$$x_k^{\mathsf{TSVD},\,r} = \sum_{j=1}^{r} \frac{u_j^T\,b}{\sigma_j} v_j\,, \qquad r \leq k$$

and study the relative error

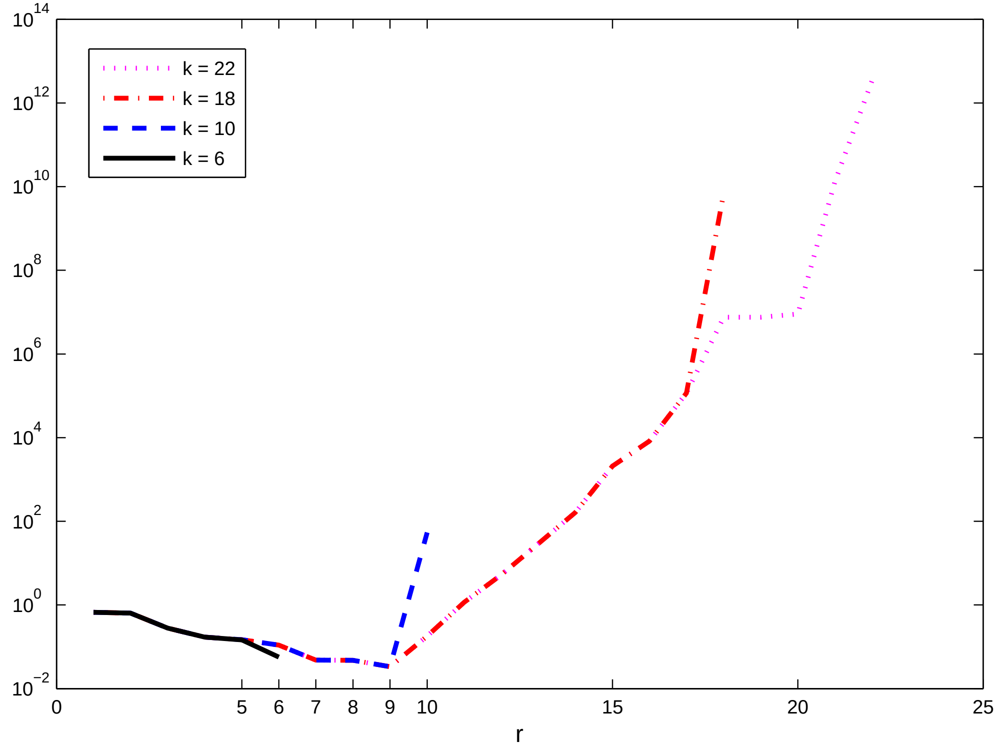
$$\frac{\|\, x^{exact} \;-\; x_k^{\mathsf{TSVD},\,r} \,\|_2}{\|\, x^{exact} \,\|_2}\,.$$

Remember k is the dimension of the bidiagonal problem, r is the truncation parameter in TSVD applied on the bidiagonal problem.

Relative error of k-r-TSVD solution, for SHAW(400) with $\| b^{noise} \| = 10^{-12}$ ($k^{noise} = 18$):



Relative error of k-r-TSVD solution, for SHAW(400) with $\| b^{noise} \| = 10^{-2}$ ($k^{noise} = 8$):



Choosing the TSVD truncation level r close to k^{noise} gives a good approximation of the solution.

Discrepancy principle:

Bidiagonalization is stopped for the smallest k where

$$\|b - Ax_k^{\mathsf{TSVD},r}\| \;=\; \alpha\;\|b^{\mathsf{noise}}\| \;\approx\; \alpha\;\delta\;\|b\|\,,$$

δ is the estimate of the noise level and α is a given real parameter.

5. Summary and future work

Information about the noise can be obtained directly from the bidiagonalization without solving the problem. It is possible to

- identify the iteration when the noise is significantly transfered to the projection,
- estimate cheaply the noise level in the original data.

Stopping criteria based on discrepancy principle can be used in subsequent iterations.

Thank you for your attention!

References

- Golub, Van Loan – An analysis of the total least squares problem, Numer. Anal., 1980.
- Hansen – Rank-Deficient and Discrete Ill-Posed Problems, SIAM Monographs Math. Modeling Comp., 1998.
- Hansen – Matlab package: REGULARIZATION TOOLS 3.2.
- Hansen, Kilmer, Kjeldsen – Exploiting residual information in the parameter choice for discrete ill-posed problems, BIT, 2006.
- H., Strakoš – Lanczos tridiagonalization and core problem, LAA, 2007.
- Paige, Strakoš – Core problem in linear algebraic systems, SIAM J. Matrix Anal. Appl., 2006.
- Sima – Regularization techniques in model fitting and parameter estimation, PhD. thesis, Faculty of Engineering, K. U. Leuven, 2006.