Golub-Kahan bidiagonalization and stopping criteria in solving ill-posed problems

Iveta Hnětynková, Martin Plešinger, Zdeněk Strakoš

with thanks to: Per Christian Hansen

ALA 2006 – Düsseldorf

Outline

1. Problem formulation
2. How to identify the noise
3. Direct truncation
4. TSVD-regularized solution
5. Summary

Problem formulation

Consider an ill-posed linear system

$$Ax = b, \quad A \in \mathcal{R}^{n \times n}, \quad b = b_{\text{exact}} + b_{\text{noise}} \neq 0 \in \mathcal{R}^{n},$$

A nonsingular,

$$\| b_{\text{exact}} \|_2 \gg \| b_{\text{noise}} \|_2 ,$$

- singular values of A decay gradually to zero,
- the noise component b_{noise} is unknown.

We wish to approximate

$$x_{\text{exact}} = A^{-1} \, b_{\text{exact}} .$$

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

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

stopped as soon as $\alpha_j = 0$ or $\beta_{j+1} = 0$ or $j = n$.

The GK algorithm generates matrices $S_j = [s_1, \dots, s_j]$, $W_j = [w_1, \dots, w_j]$ with orthonormal columns, and lower bidiagonal matrices

$$L_j = \begin{bmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \ddots & \ddots & \\ & & \beta_j & \alpha_j \end{bmatrix}, \quad L_{j+} = \begin{bmatrix} L_j \\ e_j^T \beta_{j+1} \end{bmatrix},$$

such that

$$L_j = S_j^T A W_j, \quad L_{j+} = S_{j+1}^T A W_j.$$

Core problem:

From the core problem formulation [Paige, Strakoš, 2006] it follows that there exists a fundamental decomposition

$$P^T \, [\, b \,|\, A \, Q \,] \; = \; \left[\begin{array}{c||c|c} b_1 & A_{11} & \\ \hline & & A_{22} \end{array} \right] , \quad P^T = P^{-1} , \quad Q^T = Q^{-1} ,$$

yielding for A nonsingular a compatible subproblem

$$A_{11} \, x_1 \; = \; b_1 \, ,$$

of dimension $k \leq n$, which contains all necessary and sufficient information to solve the original problem

$$x \; = \; Q \left[\begin{array}{c} x_1 \\ 0 \end{array} \right] .$$

In exact arithmetic, A nonsingular

GK stops with $\beta_{k+1} = 0$ (or $j = n$). Then $L_k \equiv A_{11}$ is a bidiagonal matrix and the matrices S_k, W_k represents first k columns of the matrices P, Q, respectively; i. e.

$$L_k\, y = \beta_1\, e_1$$

is the core problem.

In floating-point arithmetic

we have to stop GK using some stopping criteria. We can view the stopping as a perturbation of the bidiagonal matrix

$$\beta_{k+1} \longrightarrow \tilde{\beta}_{k+1} = 0\,,$$

which yields the modified matrix $\tilde{L}_k$.

How to identify the noise

Consider the singular value decomposition (SVD)

$$A = \widehat{U}\,\widehat{\Sigma}\,\widehat{V}^T,$$

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

We analyze four sets of vectors of the same length n:

- s_j, w_j : the left and right vectors from the Golub-Kahan bidiagonalization.
- $\widehat{u}_j$, $\widehat{v}_j$: the left and right singular vectors of A.

- s_j, w_j : contain the noise, especially vectors s_j, because $s_1 = b/\|b\|_2$. Vectors w_j are smoothed by the actions of the operator A^T.

- $\widehat{u}_j$, $\widehat{v}_j$: has no information about the noise, which is contained in the right-hand-side vector b. SVD of the matrix A can be expressed from the SVDs of $A_{11} \equiv L_k$ and A_{22}, the singular vectors of L_k have no information about the noise.

Our idea is: **an information about the level of noise has to be revealed from the vectors s_j, w_j generated by the Golub-Kahan bidiagonalization**.

An example:

Consider problem SHAW(400) from [Hansen – RTools] with noisy right hand side,

$$46.6225 = \| b_{\text{exact}} \|_2 \gg \| b_{\text{noise}} \|_2 = 10^{-12} .$$

The noise was artificially added.

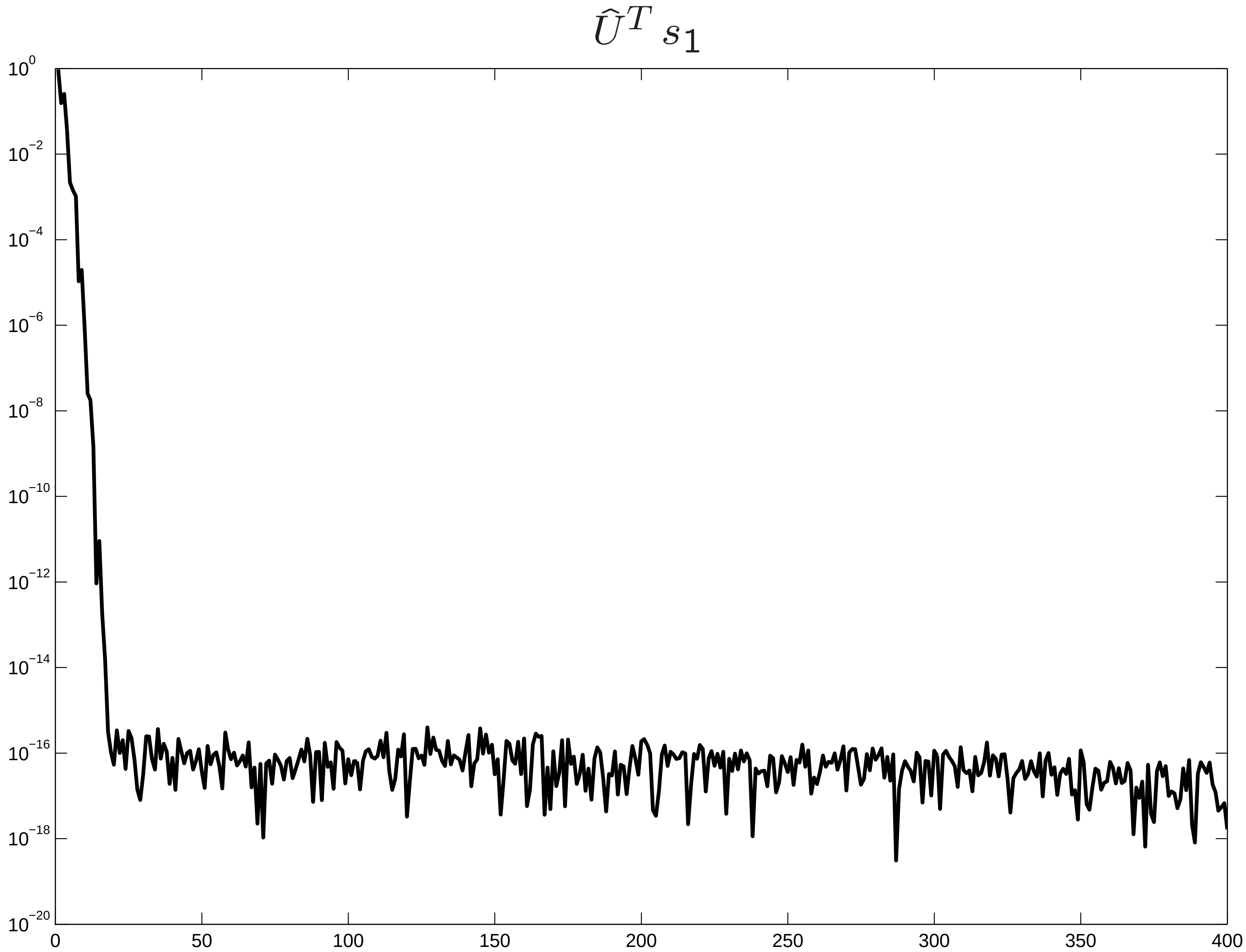
(Similar situation arises if the noise is not added, due to the rounding errors.)

We look at the noise-contaminated vectors s_j in the noise-free basis $[\widehat{u}_1, \ldots, \widehat{u}_n]$, i. e. we study vectors

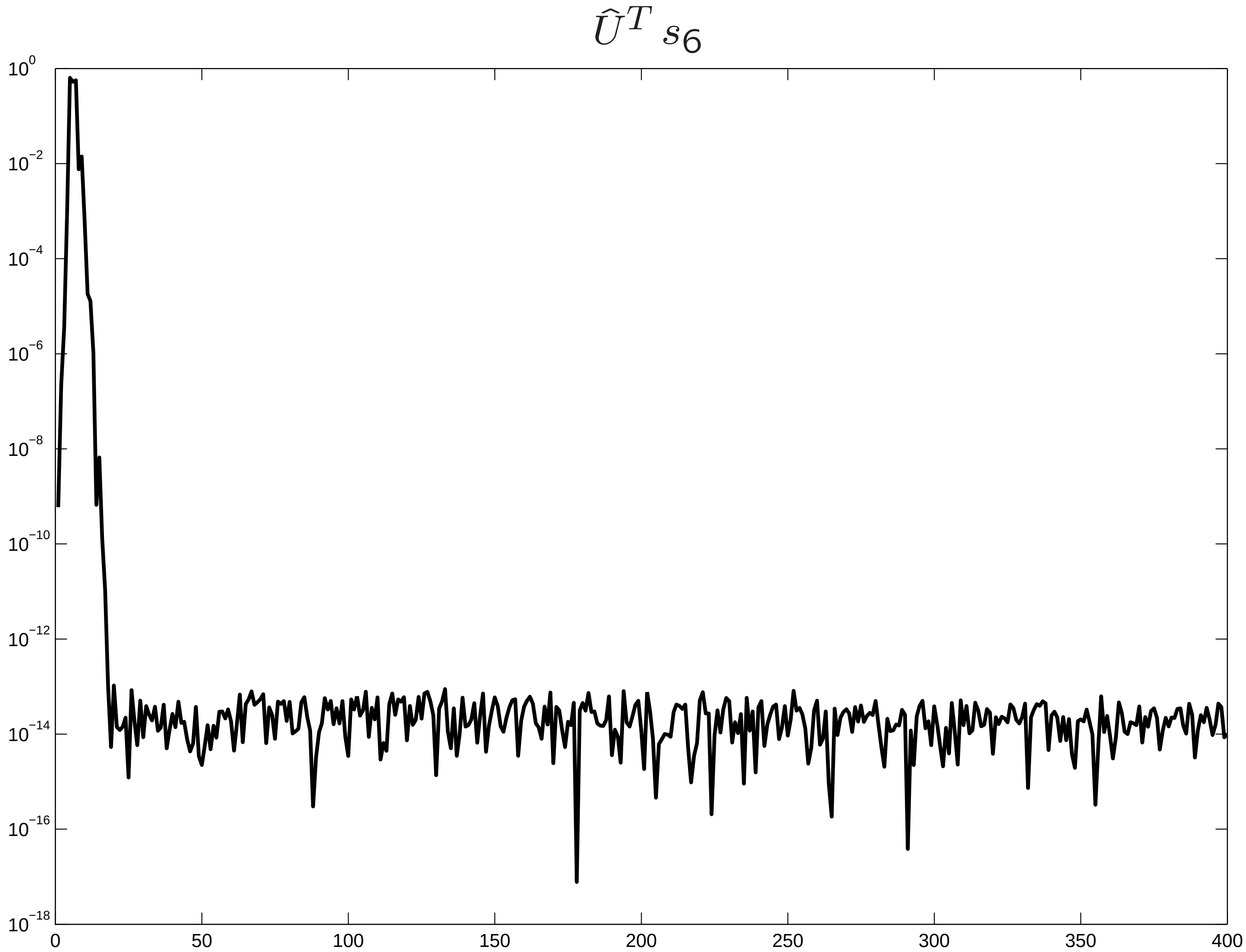
$$(\widehat{U}^T s_j), \quad j = 1, 2, \ldots .$$

The vector s_1 dominated by low frequencies, will have dominant projection in the direction of left singular vector $\widehat{u}_1$ and possibly several next vectors. Analogously with s_2, s_3, ...

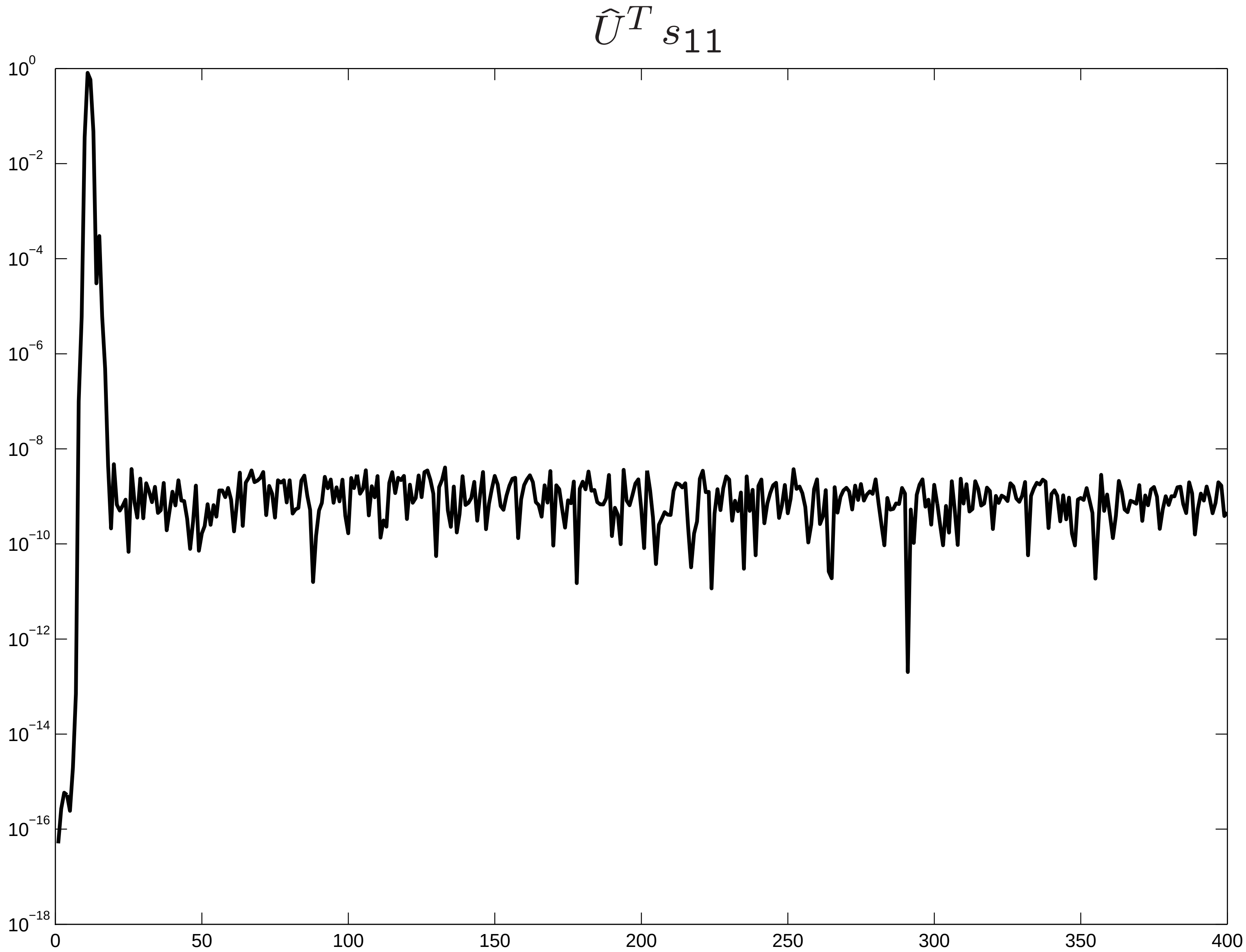
At some moment – for some index j – the low frequencies information is projected out from s_j by orthogonalization against the previous vectors, and the noise will be revealed.

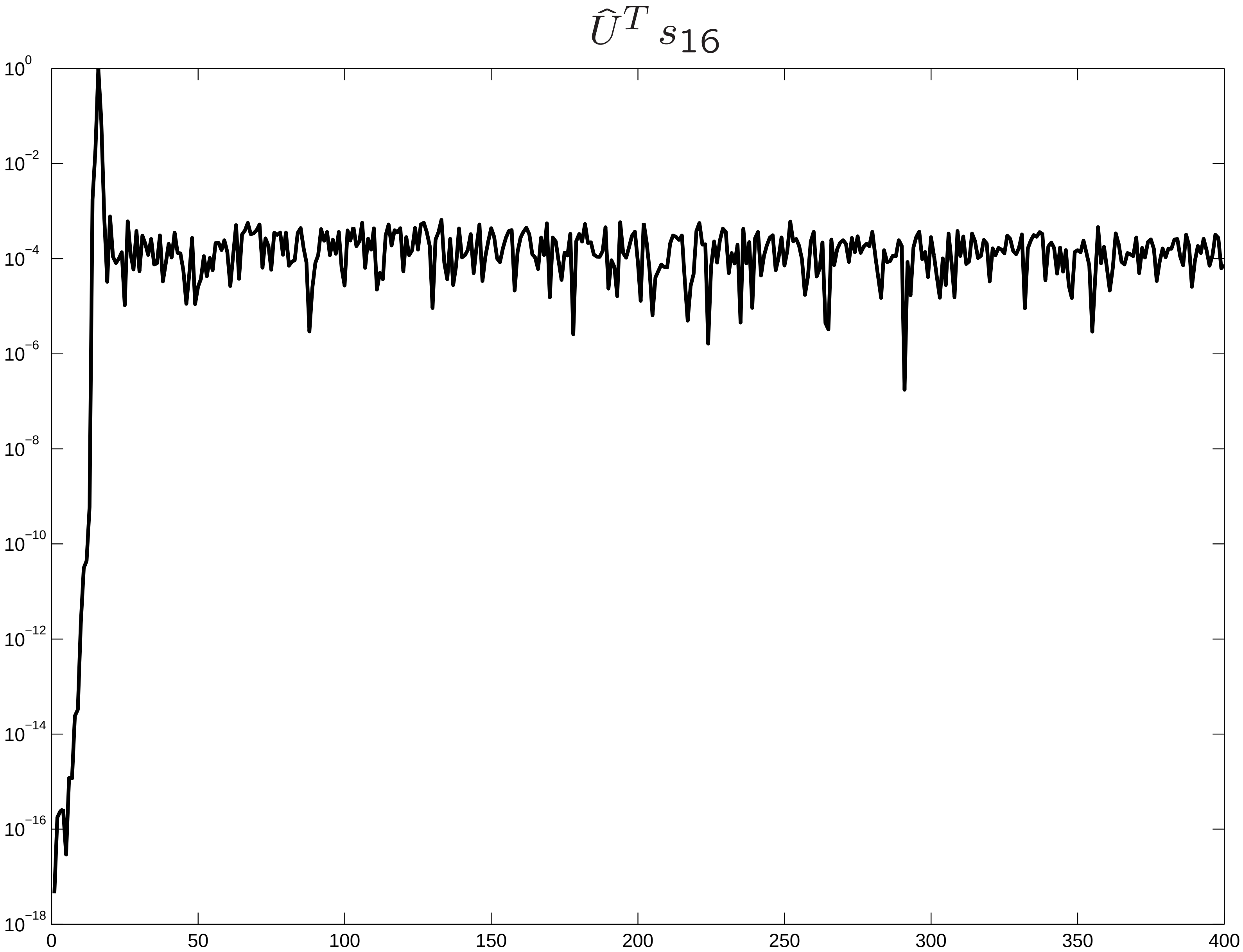


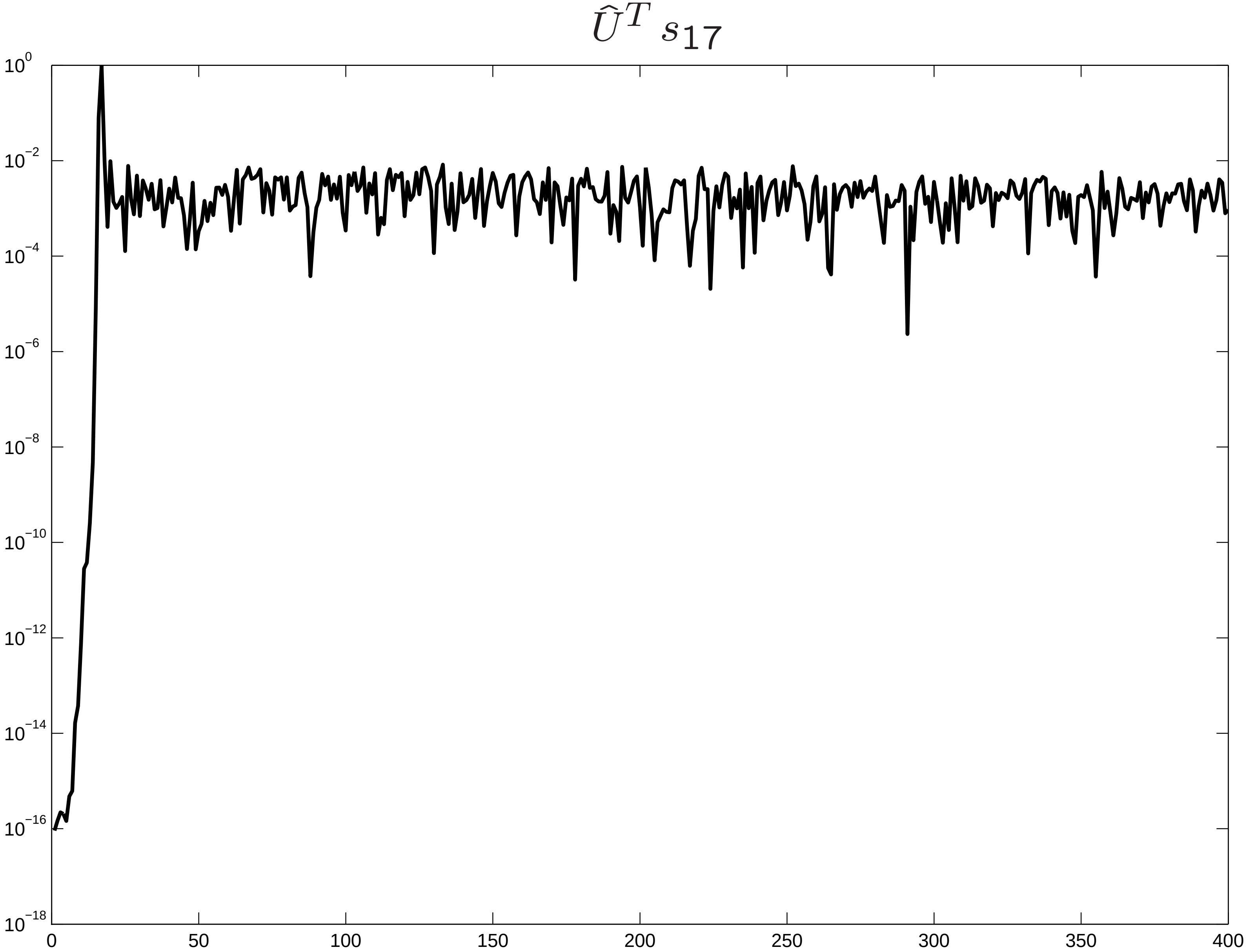
$\widehat{U}^T s_1$
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
10^{-20}
0
50
100
150
200
250
300
350
400



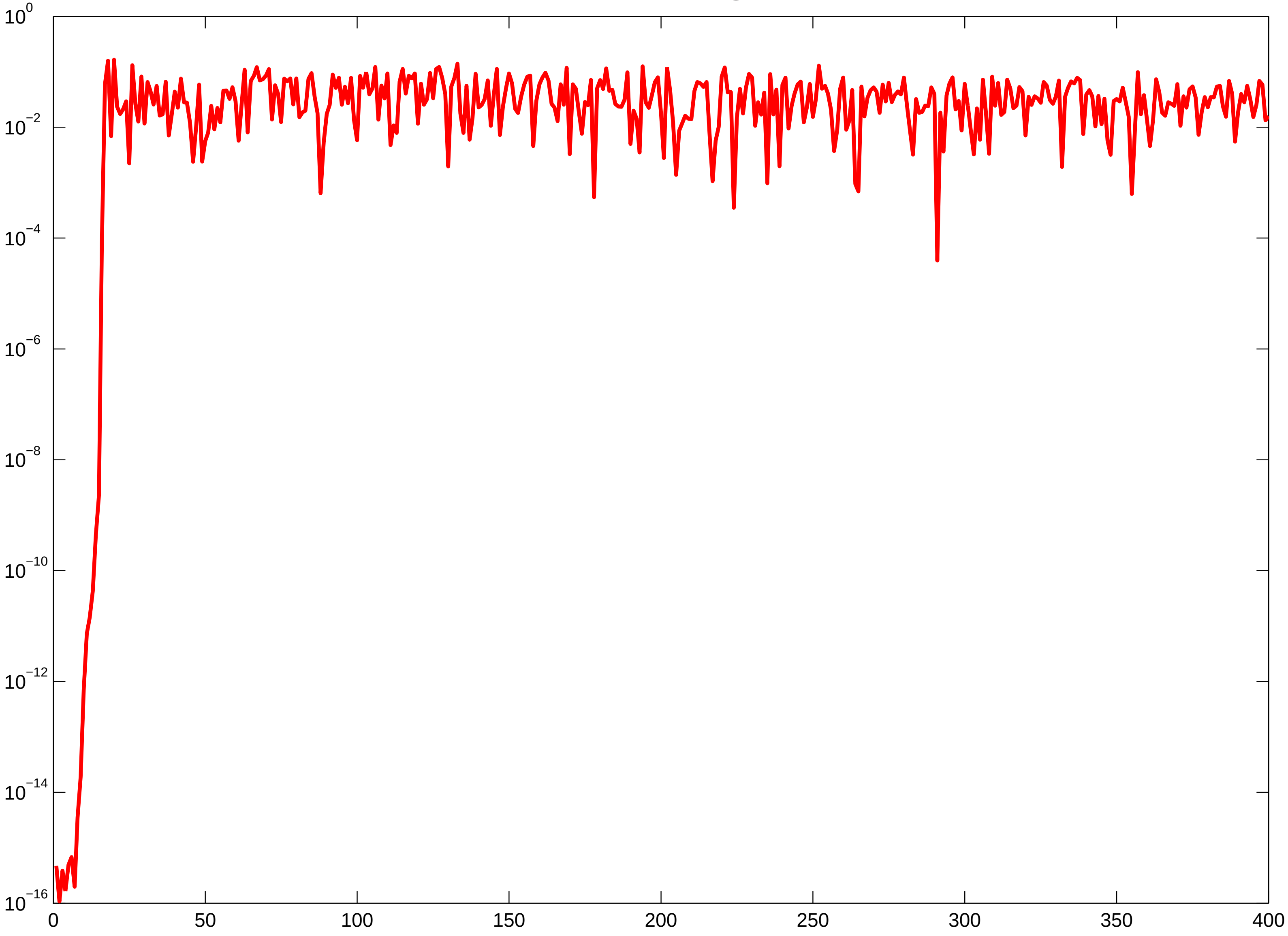
$\widehat{U}^T s_6$
10^0
10^−2
10^−4
10^−6
10^−8
10^−10
10^−12
10^−14
10^−16
10^−18
0
50
100
150
200
250
300
350
400


$\widehat{U}^T s_{11}$
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
0
50
100
150
200
250
300
350
400

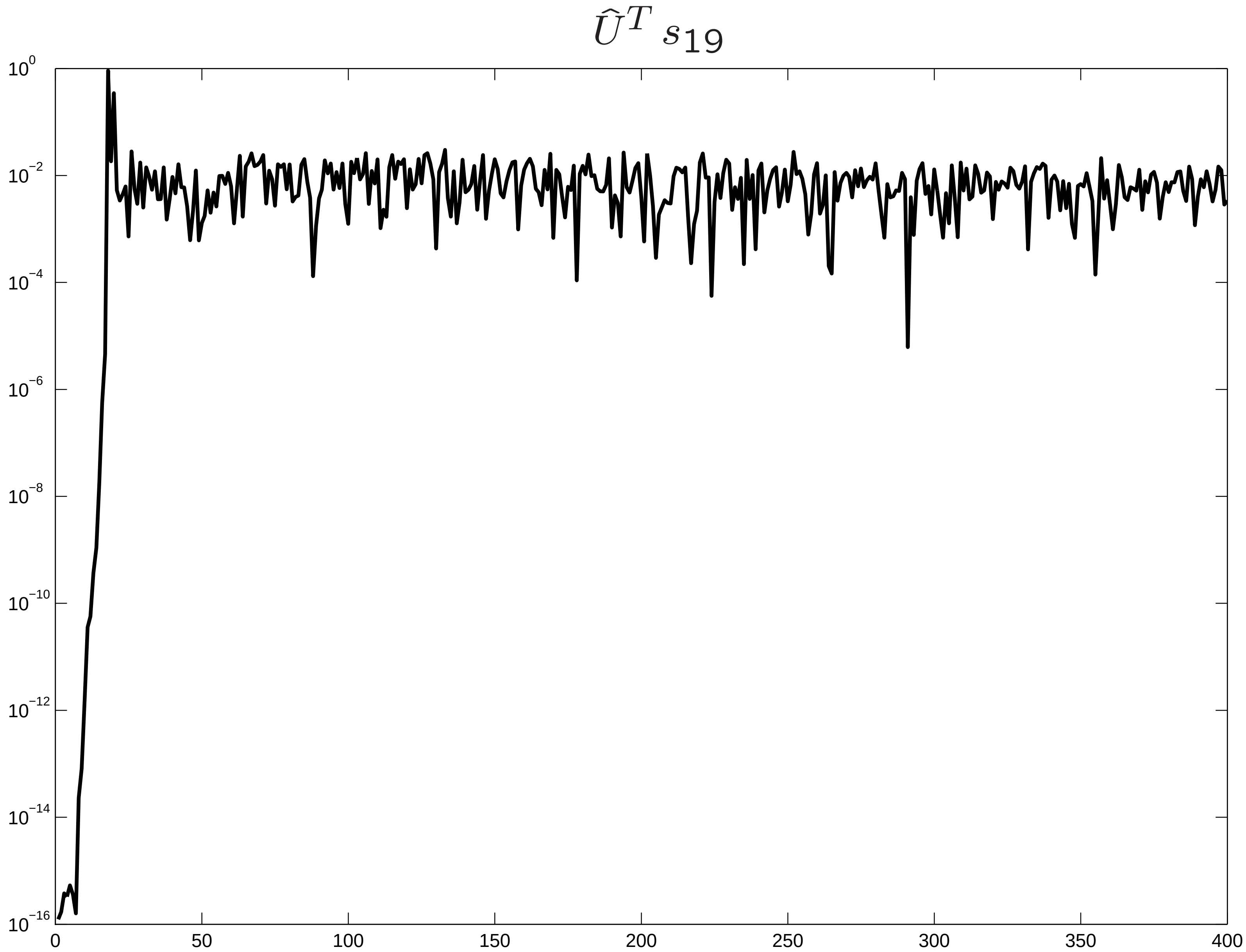

$\widehat{U}^T s_{16}$
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
0
50
100
150
200
250
300
350
400


$\widehat{U}^T s_{17}$
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
10^{-18}
0
50
100
150
200
250
300
350
400

$\hat{U}^T s_{18}$



10^0
10^−2
10^−4
10^−6
10^−8
10^−10
10^−12
10^−14
10^−16
0
50
100
150
200
250
300
350
400


$\widehat{U}^T s_{19}$
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
10^{-16}
0
50
100
150
200
250
300
350
400

Vector $(\widehat{U}^T s_{18})$ is fully dominated by noise –

the **noise level is revealed**.

Thus we get an explicit information when the noise begin to cover a useful information in the data. The solution of original problem $Ax = b$ computed through bidiagonal problem

$$L_l\, y = \beta_1\, e_1\,,$$

can for $l > 18$ be significantly polluted by the noise.

In the next step the noise is partially projected out, because vectors s_j has to be mutually orthonormal.

Direct truncation

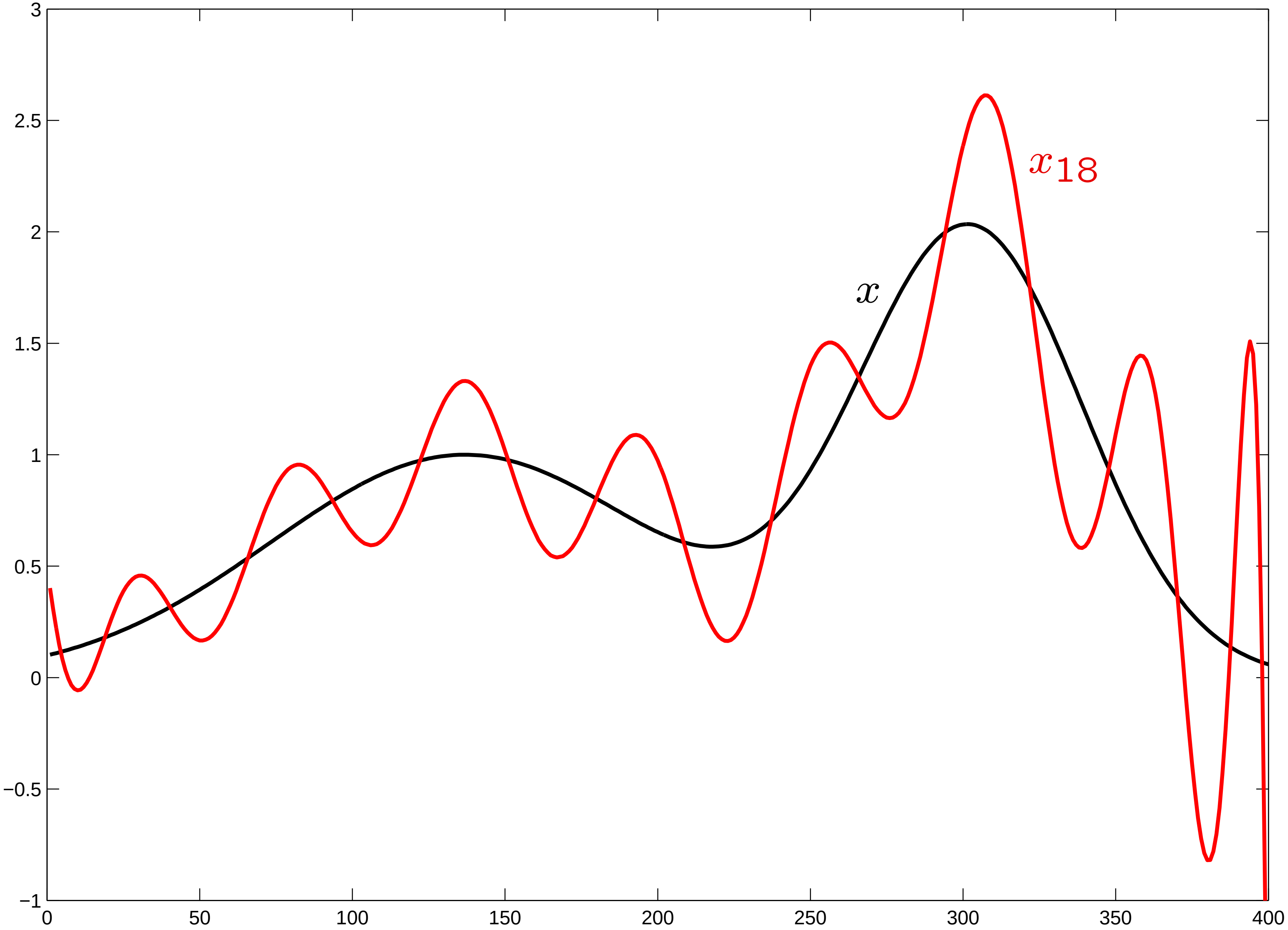
We try to solve the original problem by stopping the Golub-Kahan algorithm directly in the 18$^{\text{th}}$ iteration and compute

$$x_q = (W_q L_q^{-1} S_q^T) b, \qquad q = 18,$$

then

$$\frac{\| x - x_{18} \|_2}{\| x \|_2} \doteq 0.81902.$$

Exact and computed solution



Source of the problem:

The singular value

$$\sigma_{\mathsf{min}}(L_q) \doteq 4.8029 \times 10^{-14}$$

is very small.

We have to regularize the solution!

TSVD-regularized solution

Consider the SVD of the bidiagonal matrix

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

and denote

$$U_q \equiv S_q U_{11}, \quad V_q \equiv W_q V_{11} \quad \in \mathcal{R}^{n \times q},$$

then

$$U_q^T A V_q = \Sigma_1 .$$

Moreover we denote

$$U_q = [u_1, \ldots, u_q], \quad V_q = [v_1, \ldots, v_q], \quad \Sigma_1 = \mathrm{diag}\,(\sigma_1, \ldots, \sigma_q) .$$

We define a q-p-TSVD solution of $Ax = b$ as

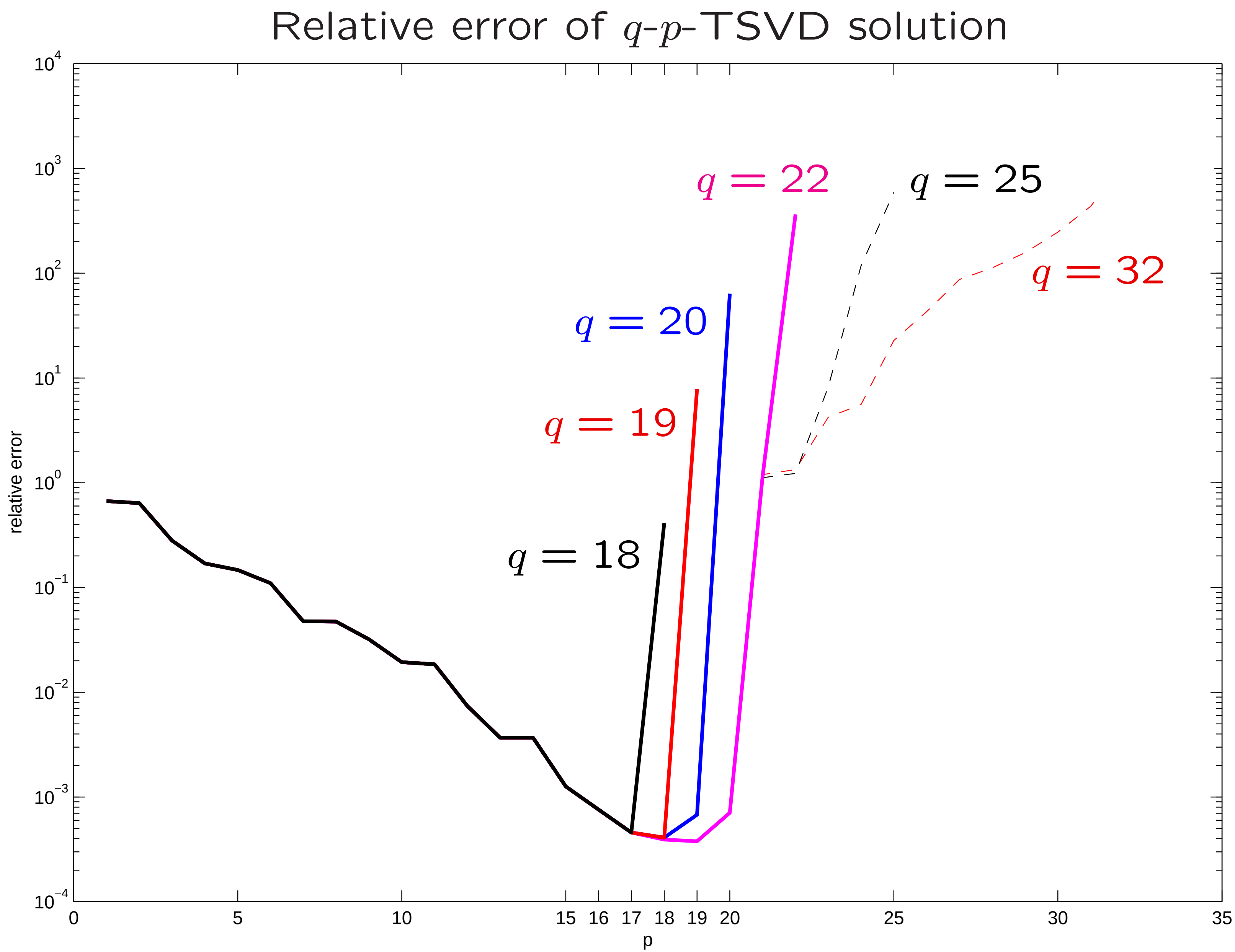
$$x_q^{\mathsf{TSVD},p} = \sum_{j=1}^{p} \frac{u_j^T b}{\sigma_j} v_j\,, \quad p \le q\,.$$

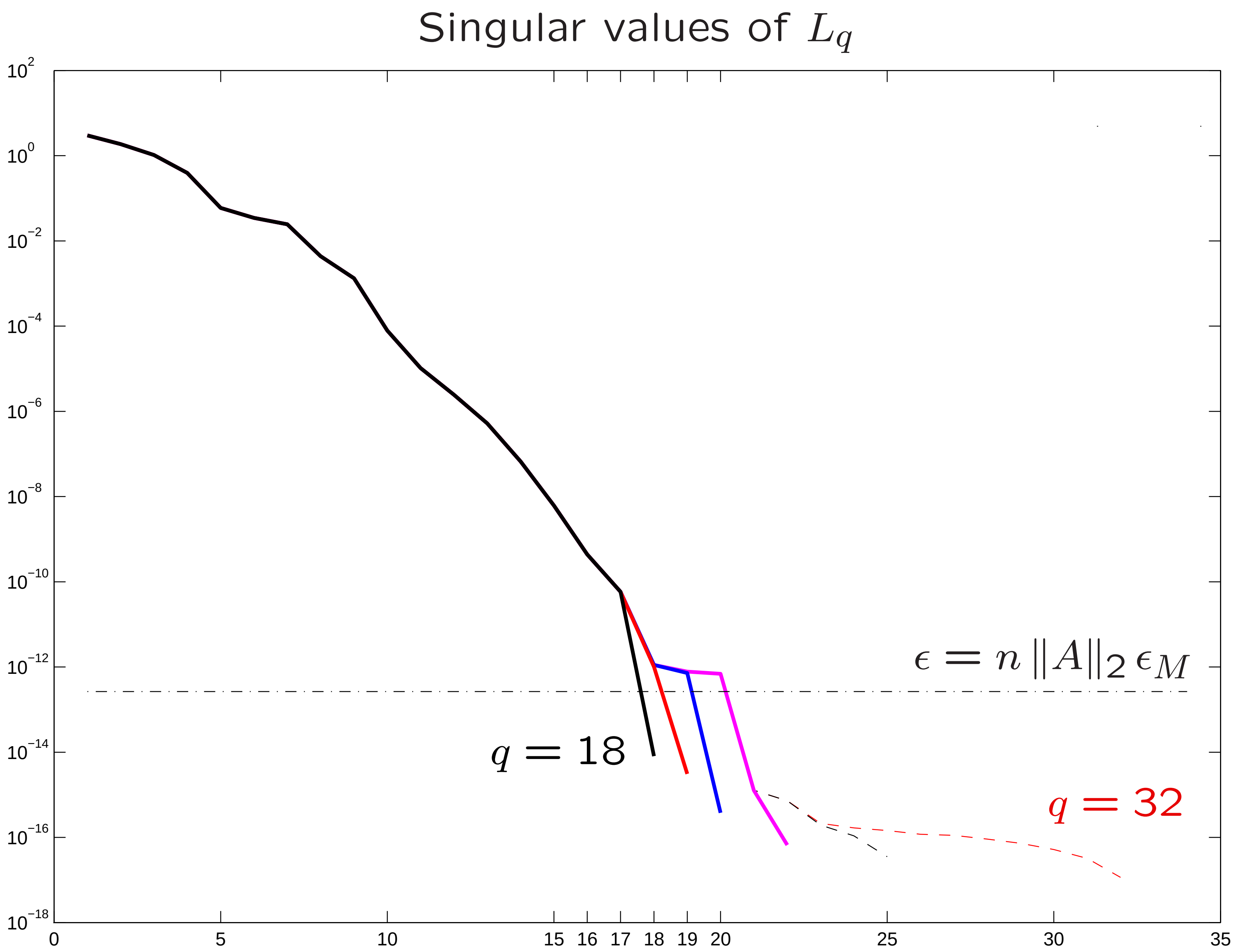
and study the relative error

$$\frac{\| x - x_q^{\mathsf{TSVD},p} \|_2}{\| x \|_2}$$

and the singular values of L_q .

Remember q is dimension of bidiagonal problem, p is the parameter of TSVD applied on the bidiagonal problem.


Relative error of q-p-TSVD solution
relative error
p
$q = 18$
$q = 19$
$q = 20$
$q = 22$
$q = 25$
$q = 32$


Singular values of L_q
$\epsilon = n\,\|A\|_2\,\epsilon_M$
$q = 18$
$q = 32$
10^2
10^0
10^−2
10^−4
10^−6
10^−8
10^−10
10^−12
10^−14
10^−16
10^−18
0
5
10
15
16
17
18
19
20
25
30
35

For $q = 18$, 19, the smallest singular value σ_q is affected by stopping. Later the small singular values are “real” small singular values of the problem. In both cases these singular values destroy the solution.

We have to stop the bidiagonalization later and then regularize the problem to attain a better relative error.
[O'Leary, Simmons – 1981] [Hansen – 1998] [Kilmer, O'Leary – 2001]

However, it is obvious, that about the 18^{th} step something essential is happening.

Summary

GK algorithm stopped at	$q =$	16	17	…
minimum relative error is at	$p =$	16	17	…
relative error is	$10^{-4} \times$	7.5009	5.4003	…

…	18	19	20	22	25	32
…	17	18	18	19	19	19
…	4.5860	4.1007	4.0969	3.7850	**3.7281**	**3.7281**

There are three stages with different behavior.

1. **Stopping GK too early** ($q < 17$): The direct truncation may be good way to approach the solution, but some useful information is lost and thus the approximation is not optimal.

   There is no extremely small singular value.

2. **Stopping GK exactly when noise reveals** ($q = 18$ or a few steps later): The direct solution is already inapplicable. If we regularize, approximation is not optimal too. The information is lost by regularization.

   There is one very small singular value – the effect of stopping.

3. **Stopping GK later** ($q > 22$): We have to regularize, because the noise was absorbed in problem, but all useful information to compute good approximation of solution is captured.

   There are some very small but "real" singular values.

Open questions

Presented ideas are only the analysis of the problem and may not be applicable directly to computation of the stopping moment.

Thus there are many opened questions:

- Relationship to common stopping criteria in hybrid methods?
- How to implement this idea?

References

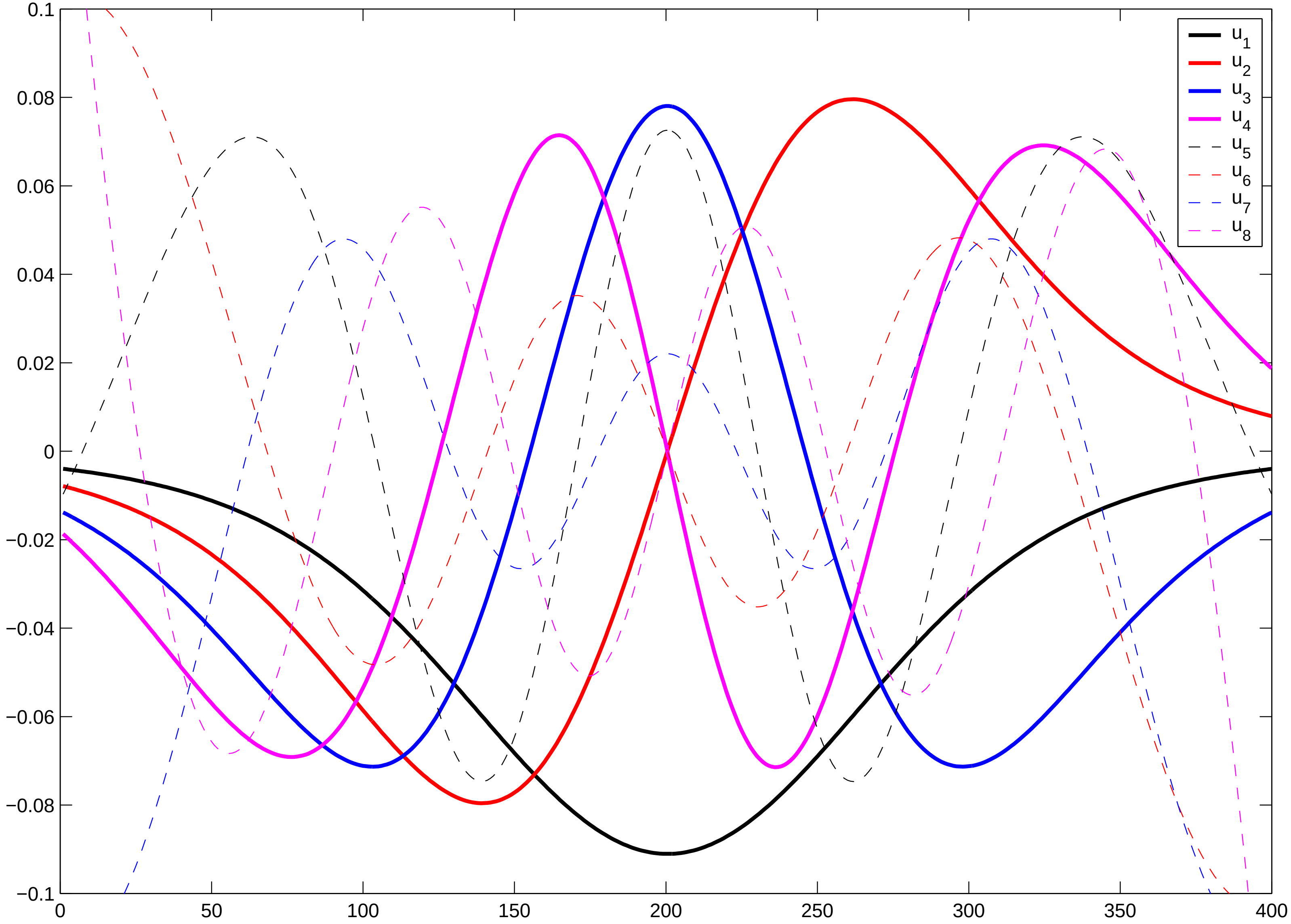
- C. C. Paige, Z. Strakoš: **Core problem in linear algebraic systems**, SIAM J. Matrix Anal. Appl., 27, pp. 861–875, 2006.
- P. C. Hansen: **Rank-Deficient and Discrete Ill-Posed Problems**, SIAM Monographs Math. Modeling Comp., 1998. ISBN 0-89871-403-6.
- P. C. Hansen: **Matlab package: REGULARIZATION TOOLS 3.2**, http://www2.imm.dtu.dk/~pch/Regutools/index.html.
- I. Hnětynková, Z. Strakoš: **Lanczos tridiagonalization and core problems**, to appear in Linear Alg. Appl.
- I. Hnětynková, M. Plešinger, Z. Strakoš: **Lanczos tridiagonalization, Golub-Kahan bidiagonalization and core problem**, to appear in PAMM.
- M. E. Kilmer, D. P. O'Leary: **Choosing Regularization Parameters in Iterative Methods for Ill-Posed Problems**, SIAM J. Matrix Anal. Appl., 22, pp. 1204–1221, 2001.
- D. P. O'Leary, J. A. Simmons: **A bidiagonalization-regularization procedure for large scale discretizations of ill-posed problems**, SIAM J. Sci. Stat. Comput., 2, pp. 474–489.
- D. M. Sima: **Regularization Techniques in Model Fitting and Parameter Estimation**, PhD. Thesis, 2006.
- D. M. Sima, S. Van Huffel: **Using Core Formulations for Ill-posed Linear Systems**, in Proc. of the 76th GAMM Annual Meeting, Luxembourg, pp. 795–796, 2005.

THANK YOU

FOR YOUR ATTENTION

Appendix, $\widehat{u}_j$ and s_j vectors

Left singular vectors $\hat{u}_j$ of SHAW(400)



Left GK vectors s_j of SHAW(400)