Topics in Numerical Analysis II
Computational Inverse Problems

Lecturer: Bangti Jin (b.jin@cuhk.edu.hk)

Chinese University of Hong Kong

September 9, 2024

Outline

- Truncated SVD (spectral cutoff)

Review: model setting

model problem: find $x \in X$ s.t.

$$Ax = y,$$

- $A : X \to Y$ a linear compact operator:
  bounded set in $X \to$ relatively compact set in Y
  limits of operators of finite rank
- $y \in Y$: given data, often contains noise

Examples

- backward heat problem: $F = F$, $X = Y = L^2(\Omega)$
- Euclidean case: $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$

Review: singular system

characterization of compact operators: There exists a set of (possibly countably infinite) vectors $(v_n)_n \subset X$ and $(u_n)_n \in Y$ and a sequence of positive numbers $(s_n)_n$, ordered nonincreasingly and $\lim_{n\to\infty} s_n = 0$ (if the rank is not finite) such that

$$Ax = \sum_n s_n(x, v_n)u_n, \quad \forall x \in X$$

or

$$Av_n = s_n u_n, \quad n = 1, \dots \quad \text{or} \quad A = \sum_{n=1}^{\infty} s_n u_n \otimes v_n$$

and

$$\overline{\mathrm{range}(A)} = \overline{\mathrm{span}(u_n)}, \quad (\ker(A))^{\perp} = \overline{\mathrm{span}(v_n)}$$

The system $(s_n, u_n, v_n)_n$ is called a singular system of A, and the expansion is called the singular value decomposition (SVD) of A.

Review: solvability condition

Picard's criterion 1909

The equation $Ax = y$ has a solution iff

$$y = Py \quad \text{and} \quad \sum_n s_n^{-2} |(y, u_n)|^2 < \infty$$

Under this condition, all solutions of $Ax = y$ are of the form

$$x = x_0 + \sum_n s_n^{-1} (y, u_n) v_n$$

for some $x_0 \in \ker(A)$

truncated singular value decomposition

Define a family of finite-dimensional orthogonal projections:

$$P_k : Y \to \mathrm{span}(u_i)_{i=1}^k, \quad y \mapsto \sum_{i=1}^k (y, u_i) u_i.$$

Due to the orthonormality of (u_n),

$$P(P_k y) = \sum_{n=1}^\infty (P_k y, u_n) u_n = \sum_{n=1}^k (y, u_n) u_n = P_k y,$$

and moreover

$$\sum_{n=1}^k s_n^{-2} |(P_k y, u_n)|^2 = \sum_{n=1}^k s_n^{-2} (y, u_n)^2 < \infty$$

(for any $k \leq \mathrm{rank}(A)$ if the latter is finite).

Thus, the problem

$$Ax = P_k y$$

satisfies Picard's criterion. The corresponding solutions are given by

$$x = x_0 + \sum_{n=1}^{k} s_n^{-1}(y, u_n)v_n \in X \qquad (*)$$

By the truncated SVD solution of $Ax = y$ for given $k \geq 1$, we mean $x_k \in X$ that satisfies $(*)$ and is orthogonal to the subspace $\ker(A)$ Since (v_n) span $\ker(A)^\perp$, x_k is unique and and has the smallest norm of the solutions, and is given by

$$x_k = \sum_{n=1}^{k} s_n^{-1}(y, u_n)v_n.$$

Convergence issue

Setting:

$$Ax^\dagger = y^\dagger$$

(i) with noisy data y^δ with $\|y^\dagger - y^\delta\| = \delta$
(ii) construct approximation by truncated SVD:

$$x^\delta_{k(\delta)} = \sum_{n=1}^{k(\delta)} s_n^{-1}(y^\delta, u_n)v_n$$

Question:

$$\lim_{\delta\to 0} \|x^\delta_{k(\delta)} - x^\dagger\| = 0?$$

by choosing properly $k(\delta)$

triangle inequality $\Rightarrow$

$$\|x^{\delta}_{k(\delta)} - x^{\dagger}\| \leq \|x^{\delta}_{k(\delta)} - x_{k(\delta)}\| + \|x_{k(\delta)} - x^{\dagger}\|$$

data error

$$x^{\delta}_{k(\delta)} - x_{k(\delta)} = \sum_{n=1}^{k(\delta)} s_n^{-1}(y^{\delta} - y^{\dagger}, u_n)v_n = \sum_{n=1}^{k(\delta)} s_n^{-1}(\xi, u_n)v_n$$

$\lim_{\delta\to 0} \|x^{\delta}_{k(\delta)} - x_{k(\delta)}\| = 0$ if $s^{-1}_{k(\delta)}\delta \to 0$ as $\delta \to 0$

approximation error

$$x_{k(\delta)} - x^{\dagger} = \sum_{n=k(\delta)+1}^{\infty} s_n^{-1}(y^{\dagger}, u_n)v_n$$

$\lim_{\delta\to 0} \|x_{k(\delta)} - x^{\dagger}\| = 0$ if $k(\delta) \to \infty$ as $\delta \to 0$

a priori choice of stopping rule $k(\delta)$:
if

$$\lim_{\delta\to 0} s_{k(\delta)}^{-1}\delta = 0 \quad \text{and} \quad \lim_{\delta\to 0} k(\delta) = \infty$$

then

$$\lim_{\delta\to 0} \|x_{k(\delta)}^{\delta} - x^{\dagger}\| = 0.$$

- The convergence also holds for the discrepancy principle (later).
- What about the convergence rate ? (optimal in some sense)

DEPARTMENT OF MATHEMATICS
THE CHINESE UNIVERSITY OF HONG KONG

TSVD is a classical technique, but in the presence of random noise, it is still relatively new

Further reading: G Blanchard, M Hoffmann, M Reiß. Early stopping for statistical inverse problems via truncated SVD estimation. Electronic Journal of Statistics 2018; 12(2), 3204–3231

Example: heat conduction

$$\begin{aligned} u_t &= u_{xx}, && \text{in } \Omega \times \mathbb{R}_+, \\ u_x(0,\cdot) = u_x(1,\cdot) &= 0, && \text{on } \mathbb{R}_+, \\ u(\cdot,0) &= f, && \text{in } \Omega. \end{aligned}$$

The forward operator:

$$F : f \mapsto u(\cdot, T), \quad X = L^2(\Omega) \to L^2(\Omega) = Y$$

is characterized by

$$F : v_n \mapsto s_n v_n$$

with $(v_n) = \{1\} \cup (\sqrt{2}\cos n\pi x)_{n=1}^{\infty}$ form an orthonormal basis of $L^2(\Omega)$, and $s_n = e^{-n^2\pi^2 T} > 0$ converges to zero as $n \to \infty$.

Thus,

$$Ff = \sum_{n=0}^{\infty} s_n (f, v_n) v_n$$

where the inner product in $L^2(\Omega)$ is defined by

$$(f, g) = \int_0^1 fg dx, \quad f, g \in L^2(\Omega).$$

$u_n = v_n$ (since F is self-adjoint). Since $(v_n)_{n=0}^{\infty}$ are an orthornormal basis for $L^2(\Omega)$, we have

$$(\ker(F))^{\perp} = \overline{\text{range}(F)} = L^2(\Omega)$$

i.e., F is injective and has a dense range. In particular, the projection P into the closure of the range of F is the identity operator.

Picard criterion: there exists $f \in L^2(\Omega)$ s.t.

$$Ff = w$$

for a given $w \in L^2(\Omega)$ iff

$$\sum_{n=0}^{\infty} s_n^{-2}(w, v_n)^2 = \sum_{n=0}^{\infty} e^{2n^2\pi^2 T}(w, v_n)^2 < \infty$$

which is very restrictive, indicating that the problem is very ill-posed. The truncated SVD solution is given by

$$f_k = \sum_{n=0}^{k} s_n^{-1}(w, v_n)v_n = \sum_{n=0}^{k} e^{n^2\pi^2 T}(w, v_n)v_n, \quad k \geq 0.$$

Euclidean case

Euclidean case: $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$, i.e., a linear system

$$Ax = y$$

Since all operators of finite rank, i.e., with finite-dimensional range, are compact, we have the representation

$$Ax = \sum_{j=1}^{r} s_j (x, v_j) u_j, \quad r \le \min(m, n)$$

where $(v_j)_{j=1}^r \subset \mathbb{R}^n$ and $(u_j)_{j=1}^r \subset \mathbb{R}^m$ are sets of orthonormal vectors and $(s_j)_{j=1}^r$ are positive numbers such that $s_j \ge s_{j+1}$, and $r = \mathrm{rank}(A)$.

Gram-Schmidt process for computing the complementary sets of orthonormal vectors $(v_j)_{j=r+1}^{n}$ and $(u_j)_{j=r+1}^{m}$, such that the completed systems $(v_j)_{j=1}^{n}$ and $(u_j)_{j=1}^{m}$ are orthonormal basis for $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Moreover, we set $s_j = 0$, $j = r+1, \ldots, \min(n, m)$
now define

$$\begin{aligned} V &= [v_1 \ v_2 \ \ldots \ v_n] \in \mathbb{R}^{n \times n}, \\ U &= [u_1 \ u_2 \ \ldots \ u_m] \in \mathbb{R}^{m \times m}, \\ S &= \mathrm{diag}(s_1, \ldots, s_{\min(n,m)}) \in \mathbb{R}^{m \times n} \end{aligned}$$

where S is a diagonal matrix, with s_i on the diagonal.
Due to the orthonormality of (v_j) and (u_j), the matrices V and U are orthogonal

$$V^\top V = VV^\top = I, \quad U^\top U = UU^\top = I$$

A simple computation shows that

$$USV^\top x = \sum_{j=1}^{r} s_j u_j (v_j^\top x) = Ax, \quad \forall x \in \mathbb{R}^n$$

hence we have the decomposition

$$A = USV^\top$$

This is called SVD for matrices in $\mathbb{R}^{m\times n}$ (in MATLAB: `svd`)
computational cost: $O(\min(mn^2, nm^2))$

Note that the singular values $(s_j)_{j=1}^{\min(n,m)}$ are just non-negative, which were assumed to be positive, and

$$\begin{aligned}\operatorname{range}(A) &= \operatorname{span}(u_j)_{j=1}^{r}\\ \ker(A) &= \operatorname{span}(v_j)_{j=r+1}^{n}\\ (\operatorname{range}(\mathrm{A}))^{\perp} &= \operatorname{span}(u_j)_{j=r+1}^{m}\\ (\ker(A))^{\perp} &= \operatorname{span}(v_j)_{j=1}^{r}\end{aligned}$$

truncated SVD for a matrix $A \in \mathbb{R}^{m\times n}$
The truncated SVD solution, i.e., the solution of

$$Ax = P_k y, \quad x \in \ker(A), \quad k \in \{1, \ldots, r\}$$

with $P_k \to \mathrm{span}(u_j)_{j=1}^k$ is an orthogoal projection, is given by

$$x_k = \sum_{j=1}^{k} s_j^{-1}(y, u_j)v_j = VS_k^{\dagger}U^{\top}y,$$

where S_k^{-1} is given by

$$S_k^{\dagger} = \mathrm{diag}(s_1^{-1}, \ldots, s_k^{-1}, 0, \ldots, 0)$$

For the largest possible cut-ff $k = r$, the matrix

$$A^\dagger := A_r^\dagger = VS_r^\dagger U^\top =: VS^\dagger U^\top$$

is called Moore-Penrose pseudoinverse. It follows from the discussions that $x^\dagger = A^\dagger y$ is the solution of the projected equation

$$Ax = P_r y = Py$$

where $P : \mathbb{R}^m \to \mathbb{R}^m$ is, once again, the orthogonal projection onto $\mathrm{range}(A)$. However, since the smallest nonzero singular values s_r is often very small for inverse problems, the use of pseudoinverse is often sensitive to the noise in the data y

Example: heat conduction revisited

$$\begin{aligned} u_t &= u_{xx}, && \text{in } \Omega \times \mathbb{R}_+, \\ u_x(0,\cdot) = u_x(1,\cdot) &= 0, && \text{on } \mathbb{R}_+, \\ u(\cdot,0) &= f, && \text{in } \Omega. \end{aligned}$$

discretize the spatial variable x, and investigate the properties of the inverse problem numerically

discretization: $h = 1/K$, grid points $x_j = jh$, $j = 0, \ldots, K$, and let $u_j(t) = u(x_j, t)$

we approximate the second-derivative of u w.r.t. x at the point (x_j, t) by the central difference

$$u_{xx}(x_j, t) = h^{-2}(u_{j+1}(t) - 2u_j(t) + u_{j-1}(t)), \quad j = 1, \ldots, K-1$$

discretize the boundary conditions by

$$\begin{aligned} u_x(0, t) &\approx h^{-1}(u_1(t) - u_0(t)) = 0, \\ u_x(1, t) &\approx h^{-1}(u_K(t) - u_{K-1}(t)) = 0 \end{aligned}$$

By solving this for $u_0(t)$ and $u_K(t)$, and substituting them into the preceding finite difference approximation, we obtain

$$\begin{aligned} u_{xx}(x_1, t) &= h^{-2}(-u_1(t) + u_2(t)) \\ u_{xx}(x_j, t) &= h^{-2}(u_{j-1}(t) - 2u_j(t) + u_{j+1}(t)), \quad j = 2, \ldots, K-2 \\ u_{xx}(x_{K-1}, t) &= h^{-2}(u_{K-2}(t) - u_{K-1}(t)) \end{aligned}$$

Let $U = (u_1, \dots, u_{K-1})^\top$ and $F = (f(x_1), \dots, f(x_{K-1}))^\top$ and substituting them into the heat equation, we obtain

$$\begin{aligned} U'(t) &= BU(t), \quad t \in \mathbb{R}_+ \\ U(0) &= F, \end{aligned}$$

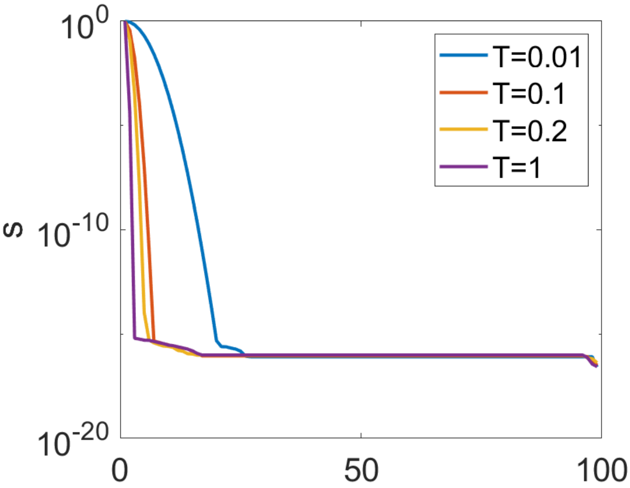
(B is a certain tridiagonal matrix)
discrete forward map: the matrix exponential function (with $T > 0$)

$$U(T) = AF, \quad \text{with } A = e^{TB}$$

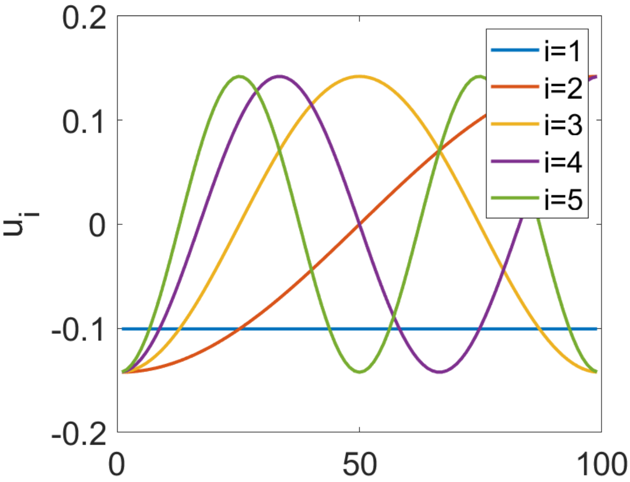
In MATLAB, the matrices B and $A = e^{TB}$ can be formed concisely

```
T = 0.1;
N = 100;
h = 1/N;
B = diag(ones(N-2,1),-1) - 2*eye(N-1) ...
    + diag(ones(N-2,1),1);
B(1,1) = -1; B(N-1,N-1)=-1;
B = B/h^2;
A = expm(T*B);
[U,S,V]=svd(A);
semilogy(diag(S),'linewidth',2)
```

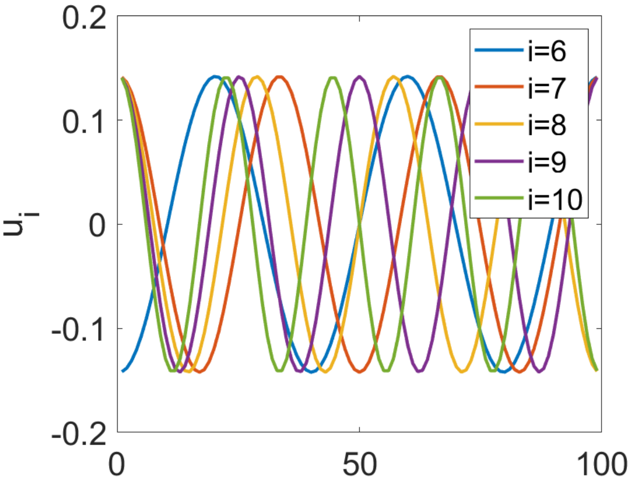
singular value distribution



singular vectors

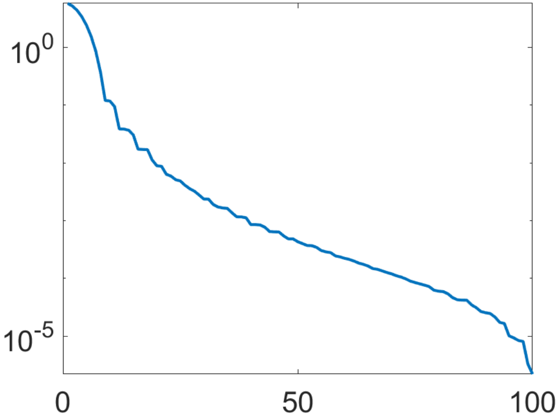


singular vectors

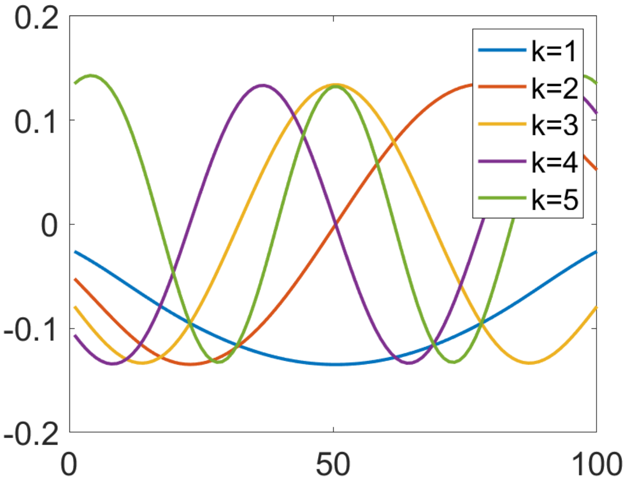


phillips: linear integral equation with kernel $k(s,t) = \phi(s-t)$

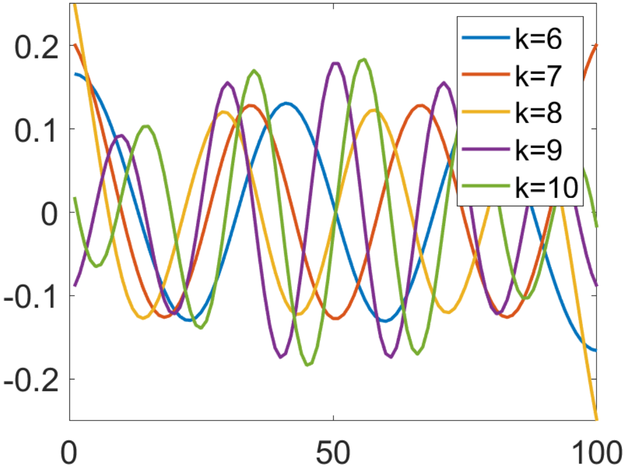
$$\phi(x) = 1 + \cos(\tfrac{x}{3}\pi)\chi_{|x|\leq 3}$$



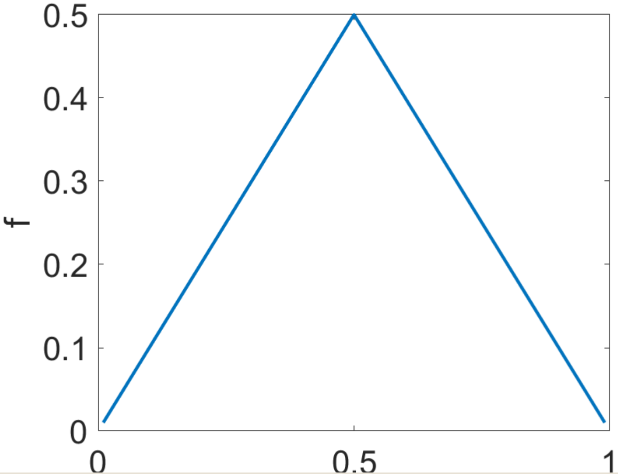
singular vectors



singular vectors



backward heat with nonsmooth initial condition, wedge, and compute the terminal observation at $T = 0.01$

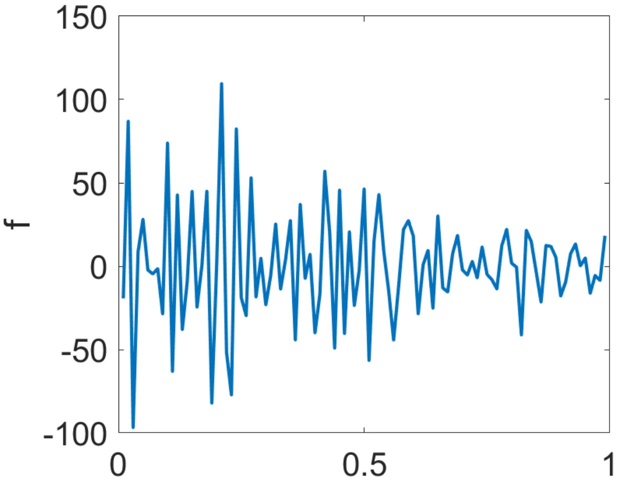


naive solution: recover the initial data by inverting A

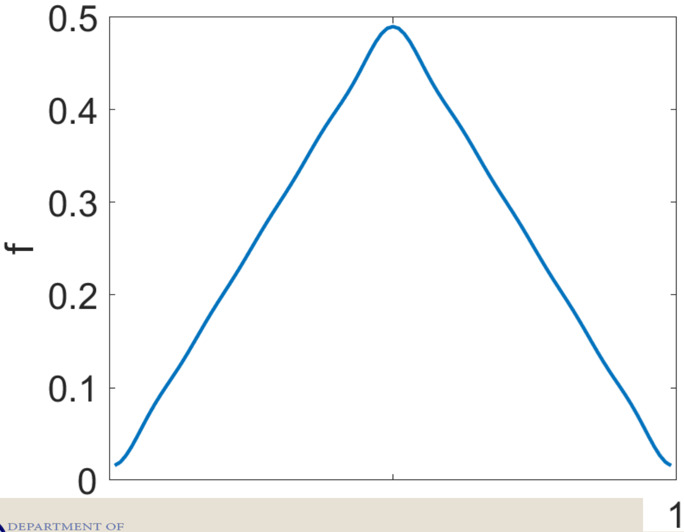
$$f^{\dagger} = A \backslash w$$

which gives a catastrophe. This is not surprising since $\mathrm{rank}(A)$ (in MATLAB) gives the value 19. Hence, A is not numerically invertible!

least-squares solution



clever solution by means of truncated SVD for $k = 19$



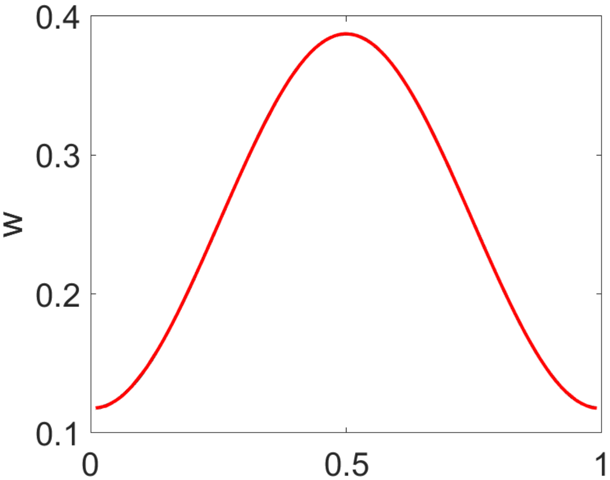
```
k = 19;
d = diag(S);
fk  = V(:,1:k)*((U(:,1:k)'*w)./diag(S(1:k,1:k)));
plot(x,f,x,fk,'k','linewidth',2)
```

inverse crime

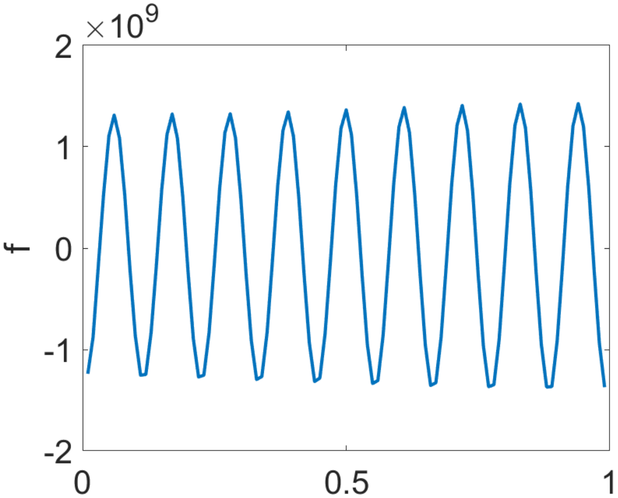
the experiment committed a severe **inverse crime**: if an inverse problem is solved using the same discretization with which the data is generated, the results are overly optimistic. This problem could be circumvented, e.g., by interpolating onto a sparser grid before the inversion. The inverse crime effect can also be reduced by adding noise.

In practice, the measurement is always inaccurate! We add a small amount of noise (1e-4), so tiny that it is barely perceptible with naked eye. Frustratingly, this approach does not work any more: the inverse of the 18th singular value is approximately $3.15 \cdot 10^{12}$, which means that component of the noise vector in the direction of v_{18} is hugely magnified.

noisy v.s. exact data



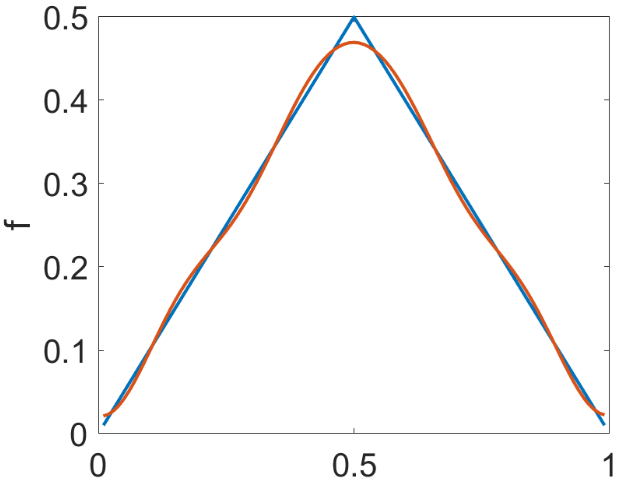
naive solution for noisy data



by trial and error, we decide to take the largest $k = 9$ singular values into account when computing truncated SVD solution

This is the best one can do without additional information about the initial data.

regularized solution



Morozov's discrepancy principle

To make the truncated SVD a more useful tool, one needs some rule for choosing the spectral cut-off index $k \geq 1$ in the truncated SVD:

$$Ax = P_k y^\delta \quad \text{and} \quad x \perp \ker(A)$$

unfortunately it is difficult to invent a reliable general scheme for choosing k

However, there exists a widely used rule of thumb called the Morozov discrepancy principle

Assume that the measurement $y^\delta \in Y$ is a noisy version of some underlying exact data $y^\dagger \in Y$. Furthermore, suppose that we have some estimate on the discrepancy between y^δ and $y^\dagger$:

$$\|y^\delta - y^\dagger\| \approx \delta > 0$$

commonly assumed noise model:

$$y^\delta = y^\dagger + \xi$$

where ξ is a realization of some random variables with known probability distribution. Knowledge of the statistics of ξ could be calibrated for some measurement devices.

The idea of Morozov's discrepancy principle is to choose the smallest $k = k(\delta)$ such that the residual satisfies

$$\|y^\delta - Ax^\delta_{k(\delta)}\| \leq \delta$$

intuition: one cannot expect the approximate solution to yield a smaller residual than the measurement error, otherwise we fit the solution to the noise
Question: Does such $k(\delta)$ exist ?
Yes, it does, if $\delta > \|Py^\delta - y^\delta\|$!

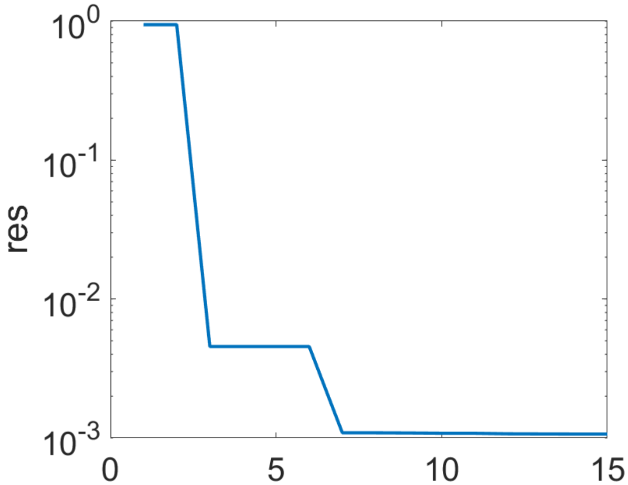
If $\operatorname{rank}(A) = \infty$, it follows from $\overline{\operatorname{range}(A)} = \operatorname{range}(P) \perp \operatorname{range}(I - P)$ that

$$\begin{aligned} \|Ax_k^\delta - y^\delta\|^2 &= \|(Ax_k^\delta - Py^\delta) + (Py^\delta - y^\delta)\|^2 \\ &= \|Ax_k^\delta - Py^\delta\|^2 + \|(P - I)y^\delta\|^2 \\ &= \sum_{n=k+1}^{\infty} (y^\delta, u_n)^2 + \|(P - I)y^\delta\|^2 \\ &\to \|Py^\delta - y^\delta\|^2 \quad \text{as } k \to \infty. \end{aligned}$$

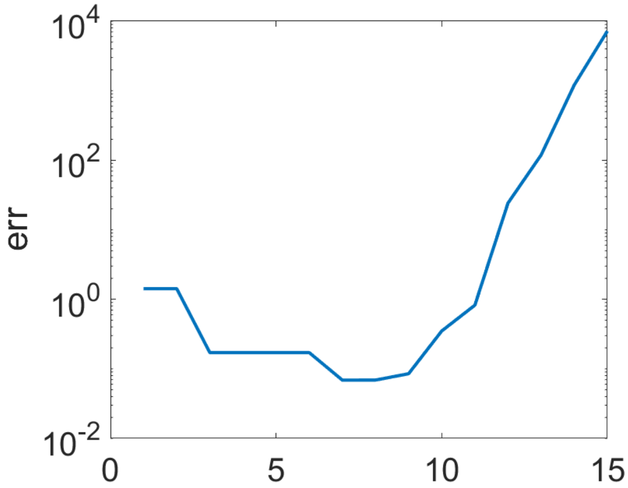
(however, there is no guarantee that x_k would not explode as $k \to \infty$)
If $r = \operatorname{rank}(A) < \infty$

$$\|Ax_r^\delta - y^\delta\| = \|P_r y^\delta - y^\delta\| = \|Py^\delta - y^\delta\|$$

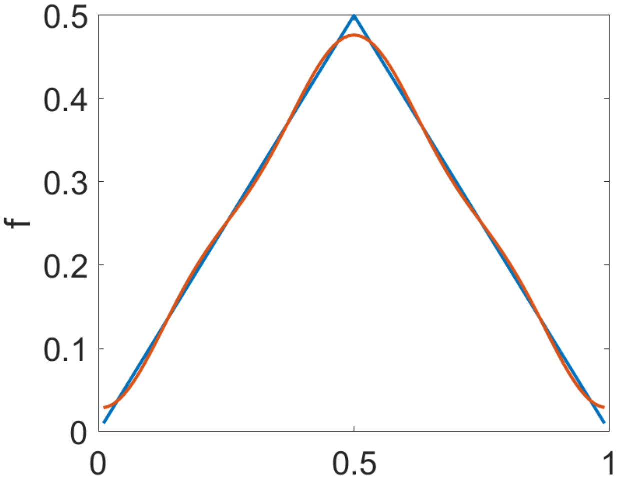
(usually one should not choose the largest spectral cutoff in practice) residual change with the stopping index



error change with the stopping index



TSVD solution with discrepancy principle, $k^* = 7$



ard deviation)

general remarks on TSVD

- it gives insight into regularization directly (removing high-freq. modes)
- it requires specifying a scalar (truncation number k)
  with optimal k, it gives a sublinear error estimate
- the method extends to general Hilbert space, compact operators
- it requires singular value decomposition $\Rightarrow$ expensive

  One can employ the randomized SVD ...
- BUT hard to incorporate other a prior knowledge

Make SVD useful for large-scale problems

complexity : computing SVD in $O(\min(n^2 m, m^2 n))$ ops
$\Rightarrow$ very expensive for large n, m (okay if $m, n \sim 1000$)

Take advantage of being ill-posed .....

$$\begin{aligned} \text{intrinsic ill-posedness} &\approx \text{low-rank approximation} \\ &\approx \text{effective low-dim column space} \end{aligned}$$

randomized SVD algorithm P.G. Martinsson, V. Rokhlin, and M. Tygert, ACHA 2006; N. Halko, P. G. Martinsson, J. A. Tropp, SIAM Review 2011 randomized SVD

1: Generate a Gaussian matrix $\Omega \in \mathbb{R}^{n\times k}$
2: Form the matrix $Y = A\Omega \in \mathbb{R}^{m\times k}$
3: Compute an orthonormal matrix $Q \in \mathbb{R}^{m\times k}$ via $Y = QR$
4: Compute the matrix $B = Q^t A \in \mathbb{R}^{k\times n}$
5: Compute the SVD of B: $B = W\Sigma V^t$
6: Form the matrix $U = QW \in \mathbb{R}^{n\times r}$, then $A \approx U\Sigma V^t$

The randomization step approx. the range of the matrix A well ...

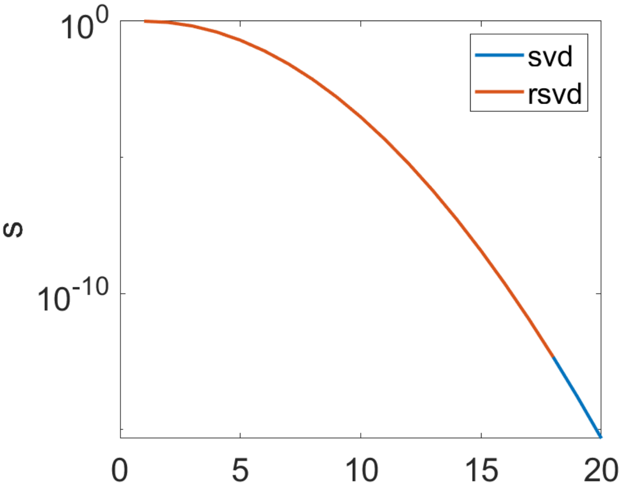
This algorithm works well if the singular values decay fast !

recall that the data is noisy ...

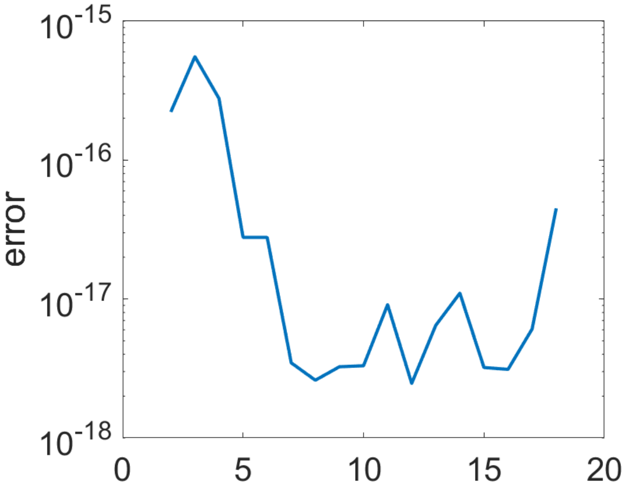
short algorithm

```
Omega = randn(n,k);
Y = A*Omega;
[Q,R] = qr(Y);
B = Q'*A;
[Uhat,S,V] = svd(B);
U = Q*Uhat;
```

randomized SVD approximation of heat example



the error of randomized approximation



low-rank approximation

optimality of SVD (in $\|\cdot\|$ or $\|\cdot\|_F$)

Theorem (Eckart-Young-Mirsky theorem)

$$\arg \min_{D \in \mathbb{R}^{m\times n}, \operatorname{rank}(D) \le r} \|A - D\|_2$$

is given by

$$D = \sum_{i=1}^{r} s_i u_i v_i^\top$$

Let $A_k = \sum_{i=1}^k s_i u_i v_i^\top$. Then

$$\|A - A_k\|_2 = \Big\| \sum_{i=1}^n s_i u_i v_i^\top - \sum_{i=1}^k s_i u_i v_i^\top \Big\|_2 = \Big\| \sum_{i=k+1}^n s_i u_i v_i^\top \Big\|_2 = s_{k+1}$$

For any $B_k = XY^\top$ with X, Y having k columns. Since Y has k columns, there exists a unit vector $w \in \mathrm{span}(v_i)_{i=1}^{k+1}$ s.t. $Y^\top w = 0$:

$$w = \sum_{i=1}^{k+1} \gamma_i v_i, \quad \text{with} \sum_{i=1}^{k+1} \gamma_i^2 = 1.$$

Then

$$\|A - B_k\|_2^2 \geq \|(A - B_k)w\|_2^2 = \|Aw\|_2^2 = \sum_{i=1}^{k+1} \gamma_i^2 s_i^2 \geq s_{k+1}^2.$$

error $e_k = \|A - \hat{A}_k\|_2$ v.s. the smallest error $s_{k+1} = \|A - A_k\|_2$

Theorem N. Halko, P. G. Martinsson, J. A. Tropp, SIAM Review 2011

If p is a small integer (e.g., $p = 5$), then

$$\mathbb{E}\|A - \hat{A}_{k+p}\|_2 \leq \left(1 + \left(\frac{k}{p-1}\right)^{\frac{1}{2}}\right) s_{k+1} + \frac{e(k+p)^{\frac{1}{2}}}{p} \left(\sum_{j=k+1}^{n} s_j^2\right)^{\frac{1}{2}}$$

- singular values decay rapidly: $\left(\sum_{j=k+1}^{n} s_j^2\right)^{\frac{1}{2}} \sim s_{k+1}$
- singular values decay slowly: $\left(\sum_{j=k+1}^{n} s_j^2\right)^{\frac{1}{2}} \sim (n-k)^{\frac{1}{2}} s_{k+1}$