Topics in Numerical Analysis II
Computational Inverse Problems

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

Chinese University of Hong Kong

September 18, 2023

Outline

1 Tikhonov regularization

Review of previous lecture

The truncated SVD solution: for $\mathbb{N} \ni k \leq \mathrm{rank}(A)$, there exists a unique $x_k \in X$ s.t.

$$Ax_k = P_k y, \quad x_k \perp \ker(A)$$

with $P_k : Y \to \mathrm{span}(u_i)_{i=1}^k$ is the orthogonal projection. This solution can be given as

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

- the method is convergent with proper choice of k
- the index k by Morozov's discrepancy principle

For a matrix $A \in \mathbb{R}^{m \times n}$, the SVD is usually written as

$$A = USV^\top$$

with $S \in \mathbb{R}^{m \times n}$ has nonnegative singular values on its diagonal, and the columns of $V \in \mathbb{R}^{n \times n}$ and $U \in \mathbb{R}^{m \times m}$ are orthonormal basis. The truncated SVD solution of order k is given by

$$x_k = VS_k^\dagger U^\top y, \quad S_k^\dagger = \mathrm{diag}(s_1^{-1}, \ldots, s_k^{-1}, 0, \ldots, 0) \in \mathbb{R}^{n \times m}.$$

pseudo-inverse: $k = \mathrm{rank}(A)$

- an approximate SVD is viable using randomized SVD, if the singular values decay rapidly

Motivation of Tikhonov regularization

The norm of the residual

$$\|Ax - y\|$$

is minimized by the sequence of truncated SVD solutions (x_k) as k tends to $\mathrm{rank}(A)$. Unfortunately, for inverse problems, we typically also have

$$\|x_k\| \to \infty \quad \text{as } k \to \mathrm{rank}(A)$$

and it seems well motivated to minimize the residual and the norm of the solution simultaneously

Tikhonov(-Phillips) regularization

Tikhonov's lemma, 1943

Let $K \subset X$ be compact and $A : X \to Y$ be a *continuous injective* mapping. Then $A : K \to A(K)$ is uniformly continuously invertible.

- injectivity of $K \Rightarrow A|_K$ has a formal inverse, mapping $A(K)$ onto K
- Suppose that for some $\epsilon > 0$ and pairs $\{(x_n, \tilde{x}_n)\} \subset K$ with $\|x_n - \tilde{x}_n\| \geq \epsilon$ for all n, there holds $\delta_n = \|Ax_n - A\tilde{x}_n\| \to 0$ as $n \to \infty$
- Since K is compact, we can find a subsequence $(n_k)_{k=1}^{\infty}$ such that $x_{n_k} \to x$, $\tilde{x}_{n_k} \to \tilde{x}$, and $x, \tilde{x} \in K$.
- $\|x - \tilde{x}\| \geq \epsilon$, while $\|Ax - A\tilde{x}\| = \lim_{k\to\infty} \|Ax_{n_k} - A\tilde{x}_{n_k}\| = 0$, contradicting injectivity of A, i.e., no such ϵ exists
- for every $\epsilon > 0$ and every pair of elements $x, \tilde{x} \in A$ with $\|x - \tilde{x}\| \geq \epsilon$ there exists $\delta(\epsilon) > 0$ s.t. $\|Ax - A\tilde{x}\| > \delta(\epsilon)$
- i.e., if $x, \tilde{x} \in K$ satisfy $\|Ax - A\tilde{x}\| < \delta(\epsilon)$, then necessarily $\|x - \tilde{x}\| < \epsilon$. i.e., $A|_K$ is uniformly continuous.

if $x^\dagger$ is a priori known to belong to some compact set K, then it is sensible to determine the extremum of the problem

$$\min_{x\in K} \|Ax - y^\delta\|$$

Since A is compact, the problem has at least one minimizer x^δ and

$$\|Ax^\delta - y^\delta\| \le \|Ax^\dagger - y^\delta\| = \delta$$

$\Rightarrow$ a general strategy: restrict the solution to a compact set (often with better Sobolev regularity), impose the constraint by a penalty,

Tikhonov regularization A. Tikhonov 1943, 1963, D. Phillips 1962 (linear integral equations)

$$J_\alpha(x) = \|Ax - y\|^2 + \alpha\|x\|^2$$

and take the minimizer (if it exists) as an approximation

Tikhonov regularization

$$J_\alpha(x) = \|Ax - y\|^2 + \alpha\|x\|^2$$

- like k, α has to be specified (a priori / a posteriori)
- intuition: look for a solution with small $\|x\| \Rightarrow$ weak compactness ...
- can incorporate a differential operator L
- the penalty term often induces a (weak) compact subset

A. N. Tikhonov. On the stability of inverse problems. Doklady Akademii Nauk SSSR 39 (5): 195–198, 1943.

A. N. Tikhonov. Doklady Akademii Nauk SSSR 151: 501–504, 1963.

D. L. Phillips. J. ACM 9: 84, 1962

What is nice about Tikhonov regularization?

- J_α is strictly convex

$$J_\alpha(tx_1 + (1-t)x_2) \leq tJ_\alpha(x_1) + (1-t)J_\alpha(x_2) \quad \forall t \in [0,1]$$

  with strict inequality when $t \in (0,1)$

- J_α amounts to solve a linear system (if a solution exists)

$$(A^tA + \alpha I)x = A^ty$$

  hence it avoids SVD ... $\Rightarrow$ (hopefully) cheaper than SVD
  efficient solvers for dense SPD linear systems

- can handle box constraint easily ...

$$\min_{x \in \mathcal{C}} J_\alpha(x)$$

  nearly impossible with truncated SVD

There are a number of basic questions on the approach

existence Is there a solution to the Tikhonov model?

stability Does it depend stably on the perturbation of model parameters etc so that it is numerically tractable ?

accuracy How good is it as an approximation ?

ra. choice How to choose the regularization parameter α ?

solver Is it efficiently solvable ? (Not easy for large linear systems!)

discret. (for PDE related): Is the discretization convergent, and can quantify the reconstruction error ?

...

existence of Tikhonov minimizer:

let $A : X \to Y$ be linear and bounded, X and Y are Hilbert spaces. Consider the functional

$$J_\alpha(x) = \|Ax - y\|_Y^2 + \alpha \|x\|_X^2$$

existence/uniqueness

For any $\alpha > 0$, there exists one and only one solution to J_α.

proof by direct method of calculus of variations (later)

A Tikhonov regularized solution $x_\alpha \in X$ is a minimizer of the functional

$$J_\alpha(x) = \|Ax - y\|^2 + \alpha\|x\|^2$$

where $\alpha > 0$ is called the regularization parameter.

Theorem

A Tikhonov regularized solution exists, is unique and is given by

$$x_\alpha = (A^*A + \alpha I)^{-1}A^*y = \sum_i \frac{s_i}{s_i^2 + \alpha}(y, u_i)v_i$$

simple case: $X = \mathrm{R}^n$ and $Y = \mathbb{R}^m$. The general case follows the same idea but requires some functional analysis. Note that

$$x^\top (A^\top A + \alpha I)x = \|Ax\|^2 + \alpha\|x\|^2 \geq \alpha\|x\|^2 > 0$$

if $x \neq 0$. In particular, $A^\top A + \alpha I \in \mathbb{R}^{n\times n}$ is injective, i.e., it is invertible due to fundamental theorem of linear algebra. Hence

$$x_\alpha = (A^\top A + \alpha I)^{-1} A^\top y \in X$$

is well defined.

Let $(s_j)_{j=1}^r$ be the positive singular values of A, and $(v_j)_{j=1}^r$ and $(u_j)_{j=1}^r$ the corresponding sets of singular vectors that span $\ker(A)^\perp$ and $\mathrm{range}(A)$, respectively. Then we expand x_α by

$$x_\alpha = \sum_j (x_\alpha, v_j) v_j + Q x_\alpha,$$

where $Q : \mathbb{R}^n \to \ker(A)$ is an orthogonal projection. Then

$$(A^\top A + \alpha I) x_\alpha = \sum_{j=1}^r (s_j^2 + \alpha)(x_\alpha, v_j) v_j + \alpha Q x_\alpha$$

and similarly

$$A^\top y = \sum_{j=1}^r s_j (y, u_j) v_j$$

equating these two expression results in

$$(x_\alpha, v_j) = \frac{s_j}{s_j^2 + \alpha}(y, u_j), \quad j = 1, \ldots, r$$

and $Qx_\alpha = 0$, i.e.,

$$x_\alpha = \sum_{j=1}^{r} \frac{s_j}{s_j^2 + \alpha}(y, u_j) v_j$$

Finally, consider $x = x_\alpha + z$, where z is arbitrary. Then

$$\begin{aligned}
J_\alpha(x) &= \|(A(x_\alpha + z) - y\|^2 + \alpha\|x_\alpha + z\|^2 \\
&= \|Ax_\alpha - y\|^2 + 2(Ax_\alpha - y, Az) + \|Az\|^2 \\
&\quad + \alpha\|x_\alpha\|^2 + 2\alpha(x, z) + \alpha\|z\|^2 \\
&= J_\alpha(x_\alpha) + \|Az\|^2 + \alpha\|z\|^2 + 2(z, (A^\top A + \alpha I)x_\alpha - A^\top y) \\
&= J_\alpha(x_\alpha) + \|Az\|^2 + \alpha\|z\|^2 \geq J_\alpha(x_\alpha)
\end{aligned}$$

where the equality holds iff $z = 0$. Thus, $x_\alpha = (A^\top A + \alpha I)^{-1}A^\top y$ is the unique minimizer of the Tikhonov functional.

What does SVD say about Tikhonov ?

$$(V\Sigma^\top U^\top U\Sigma V^\top + \alpha I)x_{tikh} = V\Sigma U^\top y,$$

i.e.,

$$x_{tikh} = \sum_{s_i>0} \frac{s_i(y, u_i)}{s_i^2 + \alpha} v_i \quad \text{v.s.} \quad x_{lsq} = \sum_{s_i>0} \frac{(y, u_i)}{s_i} v_i$$

so we approximate $1/s_i$ by $s_i/(s_i^2 + \alpha)$

$$\begin{cases} \dfrac{s_i}{s_i^2 + \alpha} \approx \dfrac{1}{s_i}, & \text{for } s_i \gg \alpha, \\ \dfrac{s_i}{s_i^2 + \alpha} \approx \dfrac{s_i}{\alpha}, & \text{for } s_i \ll \alpha \end{cases}$$

it damps high-freq. modes like the TSVD (almost similar manner) practical implementation: $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$. The Tikhonov functional can be written as

$$J_\alpha(x) = \left\| \begin{bmatrix} A \\ \sqrt{\alpha} I \end{bmatrix} x - \begin{bmatrix} y \\ 0 \end{bmatrix} \right\|, \quad I \in \mathbb{R}^{n\times n}, 0 \in \mathbb{R}^n$$

The normal equation corresponding to this least-squares problem is

$$\begin{bmatrix} A \\ \sqrt{\alpha} I \end{bmatrix}^\top \begin{bmatrix} A \\ \sqrt{\alpha} I \end{bmatrix} x = \begin{bmatrix} A \\ \sqrt{\alpha} I \end{bmatrix}^\top \begin{bmatrix} y \\ 0 \end{bmatrix}$$

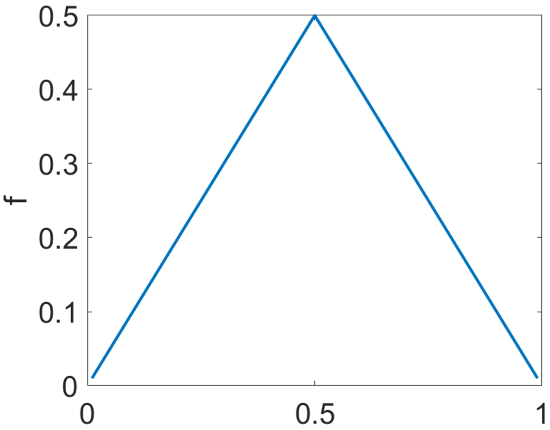
i.e.

$$(A^\top A + \alpha I)x = A^\top y$$

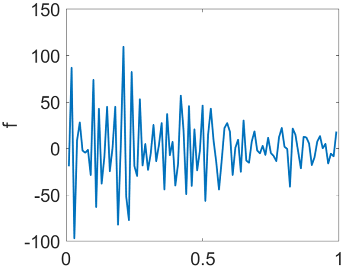
One actually does not need to form the normal equation when using Tikhonov regularization. Let

$$K = \begin{bmatrix} A \\ \sqrt{\alpha} I \end{bmatrix} \in \mathbb{R}^{(n+m)\times n} \quad \text{and} \quad z = \begin{bmatrix} y \\ 0 \end{bmatrix}$$

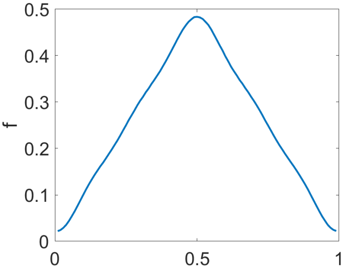
the command $x = K \backslash z$ computes the Tikhonov regularized solution



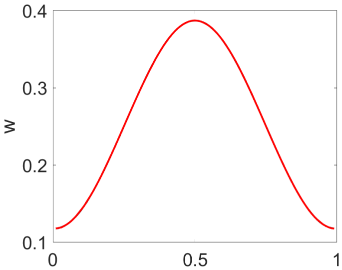
backward heat problem with wedge initial data



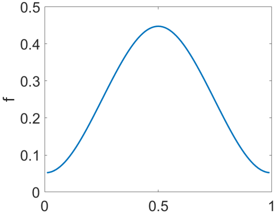
least-squares
reconstruction for exact data



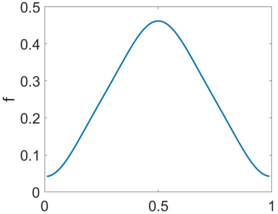
$\alpha =$ 1e-13



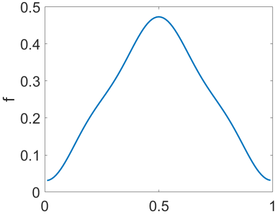
noisy data, with a small amount of noise ($\epsilon = 1$e-4)



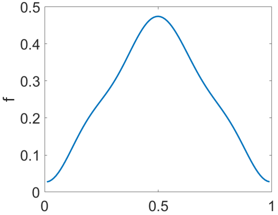
$\alpha =$ 1e-2



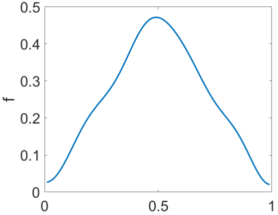
$\alpha =$ 1e-3



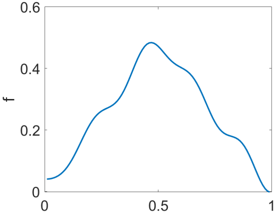
$\alpha =$ 1e-4



$\alpha =$ 1e-5



$\alpha =$ 1e-6



$\alpha =$ 1e-7

main messages:

- with *proper* α, the Tikhonov solution is good.
- The solution is generally smooth (and almost nowhere zero)
- tend to have big errors near the boundary / nonsmooth points (kinks) ...

proof technique (using functional analysis)

- $J_\alpha(x) \geq 0 \Rightarrow$ there exists a minimizing sequence $\{x^n\}$

$$\inf_{x \in X} J_\alpha(x) = \lim_{n \to \infty} J_\alpha(x^n)$$

- The minimizing seq. is uniformly bdd: $\{x^n\}$ is uniformly bdd. in X
- There exists a subseq. $\{x^{n_k}\}$ converges weakly to x^* in X
- $Ax^{n_k} \to Ax^*$ weakly (weak continuous)

$$\langle Ax^{n_k}, \bar{y} \rangle_{Y,Y^*} = \langle x^{n_k}, A^*\bar{y} \rangle_{X,X^*} \to \langle x^*, A^*\bar{y} \rangle_{X,X^*} = \langle Ax^*, \bar{y} \rangle_{Y,Y^*}$$

- weak lower semicontinuity $\|x^*\| \leq \liminf_{k \to \infty} \|x^{n_k}\|$
- x^* satisfies $J_\alpha(x^*) \leq \liminf_{k \to \infty} J_\alpha(x^{n_k}) = \inf_x J_\alpha(x)$

ingredients: coercivity, weak continuity, weak lower semicontinuity

comments

- the proof is valid for any reflexive Banach space
- also valid for nonreflexive spaces with minor changes
- for nonlinear operators, the weak continuity of A can be delicate
- the w.l.s.c. is challenging for nonconvex functionals ...

consistency

setting: given a sequence of noisy data $\{y^\delta\}_{\delta>0}$, with

$$\|y^\delta - y^\dagger\| = \delta$$

for each y^δ, construct an approx. $x^\delta_{\alpha(\delta)}$ by Tikhonov regularization

$$x^\delta_\alpha = \arg\min \|Ax - y^\delta\|^2 + \alpha\|x\|^2.$$

Question: Does $x^\delta_{\alpha(\delta)}$ converge to $x^\dagger$?

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

essentially the same argument $\Rightarrow$

consistency

$\alpha(\delta) \to 0 + \frac{\delta^2}{\alpha(\delta)} \to 0 \Rightarrow x^{\delta}_{\alpha(\delta)} \to x^{\dagger}$ as $\delta \to 0$.

recover the exact solution as the data tends to the exact one ...

The choice balances the approx. error with data error !

characterization of least-squares solution $x^{\dagger}$ (for exact data $y^{\dagger}$):

$$x^{\dagger} = \arg \min_{x \in X : Ax = y^{\dagger}} \|x\|$$

The minimum norm solution $x^\dagger$ exists and is unique.

- $y^\dagger \in \mathrm{range}(A) \Rightarrow$ the set $\mathcal{S} = \{x : Ax = y^\dagger\}$ is nonempty
- The set $\mathcal{S}$ is (weakly) closed.
- The functional $\|x\|$ is nonnegative, there exists a minimizing sequence $(x^n)_n$

$$\lim_{n\to\infty} \|x^n\| = \inf_{x\in\mathcal{S}} \|x\|$$

- a subsequence converges weakly $(x^n)_n$ to some x^*
- weak lower semicontinuity

$$\|x^*\| \leq \liminf \|x^n\| = \inf_{x\in\mathcal{S}} \|x\|$$

- existence of a minimum-norm solution
- uniqueness via strict convexity of $\|\cdot\|^2$

By the minimizing property of x_α^δ: $J_\alpha(x_\alpha^\delta) \leq J_\alpha(x^\dagger)$:

$$\begin{aligned}\|Ax_\alpha^\delta - y^\delta\|^2 + \alpha\|x_\alpha^\delta\|^2 &\leq \|Ax^\dagger - y^\delta\|^2 + \alpha\|x^\dagger\|^2 \\ &\leq \delta^2 + \alpha\|x^\dagger\|^2\end{aligned}$$

implication:

$$\begin{aligned}\|x_\alpha^\delta\|^2 &\leq \alpha^{-1}\delta^2 + \|x^\dagger\|^2, \\ \|Ax_\alpha^\delta - y^\delta\|^2 &\leq \delta^2 + \alpha\|x^\dagger\|^2\end{aligned}$$

- condition $\alpha(\delta)^{-1}\delta^2 \to 0 \Rightarrow$ the sequence $(x_{\alpha(\delta)}^\delta)_\delta$ is uniformly bdd
- $\exists$ a subsequence $(x_{\alpha(\delta)}^\delta)_\delta$ converges weakly to x^* in X
- by weak lower semi-continuity of norms

$$\|x^*\|^2 \leq \liminf \|x_{\alpha(\delta)}^\delta\|^2 \leq \lim_{\delta\to 0^+} \alpha(\delta)^{-1}\delta^2 + \|x^\dagger\|^2 = \|x^\dagger\|^2$$

- $Ax^{\delta}_{\alpha(\delta)} \to Ax^*$ weakly in Y
- $Ax^{\delta}_{\alpha(\delta)} - y^{\delta} \to Ax^* - y^{\dagger}$ weakly
- weak lower semi-continuity of norms

$$\begin{aligned}\|Ax^* - y^{\dagger}\|^2 &\leq \liminf \|Ax^{\delta}_{\alpha(\delta)} - y^{\delta}\|^2 \\ &\leq \lim_{\delta \to 0^+} \delta^2 + \alpha(\delta)\|x^{\dagger}\|^2 = 0\end{aligned}$$

under the condition $\lim_{\delta \to 0^+} \alpha(\delta) = 0$

- the limit x^* satisfies

$$\|x^*\| \leq \|x^\dagger\|, \quad \|Ax^* - y^\dagger\| = 0$$

x^* is the unique least-squares solution $x^\dagger$

- $\|x^\dagger\| \leq \|x^*\| \leq \limsup \|x_\alpha^\delta\| \leq \|x^\dagger\|$, i.e.

$$\lim_{\delta \to 0^+} \|x_\alpha^\delta\| = \|x^\dagger\|$$

+ weak convergence of $x_\alpha^\delta \Rightarrow$

$$\lim_{\delta \to 0^+} \|x_\alpha^\delta - x^\dagger\| = 0$$

- the standard subsequence argument implies the whole sequence converges weakly to $x^\dagger$

stability

- For any $\alpha > 0$, $x_\alpha(y^n) \to x_\alpha(y)$ in X as $y^n \to y$ in Y.
- Let $\alpha > 0$, for any $y \in Y$, $x_{\alpha_n} \to x_\alpha$ in X as $\alpha_n \to \alpha$.

the regularized scheme is stable $\Rightarrow$ numerically feasible ...
proof is left as an exercise.

What about the quality of approximation ? e.g., for some $\gamma > 0$

$$\|x^{\delta}_{\alpha(\delta)} - x^{\dagger}\| \leq c\delta^{\gamma} \quad \text{as } \delta \to 0?$$

- No extra conditions on $x^{\dagger}$, there can be no convergence rate !
- Under suitable conditions, we have *sublinear* rate, $0 < \gamma < 1$
  by ill-posedness, we ALWAYS lose something !

The first point is not surprising: FEM/FDM have no convergence rate if no extra regularity is available.

numerical differentiation: numerically differ. a function $g \in C^1[-1,1]$

$$g'(x) \approx D_h^+ g(x) = (g(x+h) - g(x))/h$$

for any $x \in (-1,1)$, small $h > 0$: $\lim_{h\to 0} D_h^+(x) = g'(x)$

- noisy data g^δ: $\|g^\delta - g\|_{C[-1,1]} \le \delta$
- approximation: $D_h^+ g^\delta = (g^\delta(x+h) - g^\delta)/h$
- error bound:

$$\begin{aligned} |D_h^+ g^\delta(x) - g'(x)| &\le |D_h^+ g(x) - g'(x)| + |D_h^+ g(x) - D_h^+ g^\delta| \\ &\le |D_h^+ g(x) - g'(x)| + 2\delta/h \end{aligned}$$

- Ex: $g(x) = |x|^{1+a}$, $a > 0$

$$D_h^+ g(0) = h^a$$

error bound: $h^a + 2h^{-1}\delta$, optimal error $c\delta^{a/(1+a)}$

canonical source condition

$$x^\dagger = A^* w, \quad w \in Y$$

Theorem

Under canonical source condition and $\alpha \sim \delta$, $\|x^\dagger - x_\alpha^\delta\| \leq c\delta^{1/2}$.

proof follows by means of completing squares

By the minimizing property $J_\alpha(x_\alpha^\delta) \le J_\alpha(x^\dagger)$

$$\|Ax_\alpha^\delta - y^\delta\|^2 + \alpha\|x_\alpha^\delta\|^2 \le \|Ax^\dagger - y^\delta\|^2 + \alpha\|x^\dagger\|^2$$

completing the square

$$\begin{aligned}\|Ax_\alpha^\delta - y^\delta\|^2 + \alpha\|x_\alpha^\delta - x^\dagger\|^2 &\le \|Ax^\dagger - y^\delta\|^2 - 2\alpha\langle x^\dagger, x_\alpha^\delta - x^\dagger\rangle \\ &= \|y^\dagger - y^\delta\|^2 - 2\alpha\langle A^*w, x_\alpha^\delta - x^\dagger\rangle \\ &= \|y^\dagger - y^\delta\|^2 - 2\alpha\langle w, A(x_\alpha^\delta - x^\dagger)\rangle\end{aligned}$$

completing the square again

$$\begin{aligned}&\|Ax_\alpha^\delta - y^\delta + \alpha w\|^2 + \alpha\|x_\alpha^\delta - x^\dagger\|^2 \\ \le\ &\|y^\delta - y^\dagger\|^2 - 2\alpha\langle w, y^\delta - y^\dagger\rangle + \alpha^2\|w\|^2 \\ \le\ &(\alpha\|w\| + \delta)^2\end{aligned}$$

the role of source condition is to link X and Y spaces

the choice $\alpha \sim \delta \Rightarrow O(\delta^{1/2})$ rate

- there are more general conditions,
  e.g., $(A^*A)^\mu w = x^\dagger$, or logarithmic type ...

  H.W. Engl, M. Hanke, A. Neubauer. Regularization of Inverse Problems. Kluwer, 1996
- these are related to conditional stability estimates (in PDEs)

  Cheng-Yamamoto Inverse Problems 2000

  $\Rightarrow$ Carlemann estimates M Yamamoto. Inverse Problems 2009
- variational inequality approach T. Schuster, B. Kaltenbacher, B. Hofmann and K. Kazimierski, Regularization Methods in Banach spaces, De Gruyter, Berlin, 2012.
- Kurdyka–Lojasiewicz inequality (very popular in optimization)

  Gerth-Kindermann 2019

Generally, the verification of such a condition is very hard !

Morozov's discrepancy principle

How to determine α ?
Given the data $y^\delta \in Y$ is a noisy version of the exact data $y^\dagger \in Y$ and

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

For Tikhonov regularization, Morozov's discrepancy principle chooses α s.t.

$$\|y^\delta - Ax_\alpha^\delta\| = \delta.$$

Such a regularization parameter exists if

$$\|y^\delta - Py^\delta\| < \delta < \|y^\delta\|$$

well-definedness of discrepancy principle

The function $\phi(\alpha) = \|Ax_\alpha^\delta - y^\delta\|^2$ is continuous in α, strictly monotone in α, and

$$\lim_{\alpha\to 0^+} \phi(\alpha) = \|y^\delta - Py^\delta\|^2 \quad \text{and} \quad \lim_{\alpha\to\infty} \phi(\alpha) = \|y^\delta\|^2.$$

Implication: there is a unique α satisfying the discrepancy principle, if

$$\|y^\delta - Py^\delta\| < \delta < \|y^\delta\|$$

By the representation

$$x_\alpha^\delta = \sum_i \frac{s_i}{s_i^2 + \alpha}(y^\delta, u_i)v_i$$

and thus, the residual

$$\begin{aligned} Ax_\alpha^\delta - y^\delta &= \sum_i \frac{s_i}{s_i^2 + \alpha}(y^\delta, u_i)Av_i - y^\delta \\ &= \sum_i \frac{s_i^2}{s_i^2 + \alpha}(y^\delta, u_i)u_i - y^\delta \\ &= \sum_i \left(\frac{s_i^2}{s_i^2 + \alpha} - 1\right)(y^\delta, u_i)u_i - (I - P)y^\delta \end{aligned}$$

Hence,

$$\phi(\alpha) = \|Ax_\alpha^\delta - y^\delta\|^2 = \sum_i \frac{\alpha^2}{(s_i^2 + \alpha)^2}(y^\delta, u_i)^2 + \|(I - P)y^\delta\|^2$$

strictly monotone:

$$\phi'(\alpha) = \sum_i \frac{2\alpha s_i^2}{(s_i^2+\alpha)^3}(y^\delta, u_i)^2 > 0$$

limit relation

$$\begin{aligned}\lim_{\alpha\to 0^+}\phi(\alpha) &= \sum_i \lim_{\alpha\to 0^+}\frac{\alpha^2}{(s_i^2+\alpha)^2}(y^\delta, u_i)^2 + \|(I-P)y^\delta\|^2 \\ &= \|(I-P)y^\delta\|^2 \\ \lim_{\alpha\to\infty}\phi(\alpha) &= \sum_i \lim_{\alpha\to\infty}\frac{\alpha^2}{(s_i^2+\alpha)^2}(y^\delta, u_i)^2 + \|(I-P)y^\delta\|^2 \\ &= \sum_i (y^\delta, u_i)^2 + \|(I-P)y^\delta\|^2 = \|y^\delta\|^2.\end{aligned}$$

regularizing property of the discrepancy principle

Theorem

- *Tikhonov regularization equipped with DP is consistent!*
- *Under the canonical source condition,*

$$\|x^{\delta}_{\alpha(\delta)} - x^{\dagger}\| \leq c\delta^{\frac{1}{2}}.$$

(i.e., the approximation $x^{\delta}_{\alpha(\delta)}$ converges at a rate $O(\delta^{\frac{1}{2}})$)

By the minimizing property of x_α^δ: $J_\alpha(x_\alpha^\delta) \le J_\alpha(x^\dagger)$:

$$\begin{aligned}\|Ax_\alpha^\delta - y^\delta\|^2 + \alpha\|x_\alpha^\delta\|^2 &\le \|Ax^\dagger - y^\delta\|^2 + \alpha\|x^\dagger\|^2 \\ &\le \delta^2 + \alpha\|x^\dagger\|^2\end{aligned}$$

discrepancy principle: $\|Ax_\alpha^\delta - y^\delta\|^2 = \delta^2 \quad \Rightarrow \quad \|x_\alpha^\delta\|^2 \le \|x^\dagger\|^2$

- convergent subsequence to x^*, with wlsc: $\|x^*\| \le \|x^\dagger\|$
- $Ax_\alpha^\delta - y^\delta$ converges weakly to $AX^* - y^\dagger$
- the residual also converges to zero

$$\|Ax^* - y^\dagger\| \le \liminf \|Ax_\alpha^\delta - y^\delta\| = \lim_{\delta\to 0} \delta = 0$$

x^* is the minimum norm solution

By the minimizing property $J_\alpha(x_\alpha^\delta) \le J_\alpha(x^\dagger)$

$$\|Ax_\alpha^\delta - y^\delta\|^2 + \alpha\|x_\alpha^\delta\|^2 \le \|Ax^\dagger - y^\delta\| + \alpha\|x^\dagger\|^2$$

+ DP $\Rightarrow \|x_\alpha^\delta\|^2 \le \|x^\dagger\|^2$ + source condition $\Rightarrow$

$$\begin{aligned}\|x_\alpha^\delta - x^\dagger\|^2 &\le 2\langle x^\dagger, x_\alpha^\delta - x^\dagger\rangle = -2\langle A^* w, x_\alpha^\delta - x^\dagger\rangle \\ &= -2\langle w, A(x_\alpha^\delta - x^\dagger)\rangle \le 2\|w\|\|Ax_{\alpha(\delta)}^\delta - y^\dagger\|\end{aligned}$$

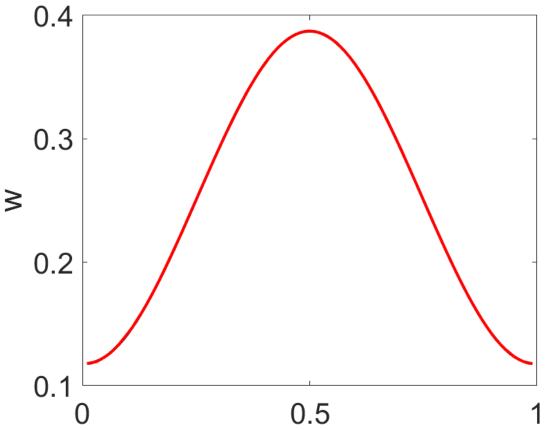
DP + triangle inequality $\Rightarrow$

$$\|Ax_{\alpha(\delta)}^\delta - y^\dagger\| \le \|Ax_{\alpha(\delta)}^\delta - y^\delta\| + \|y^\delta - y^\dagger\| \le 2\delta$$

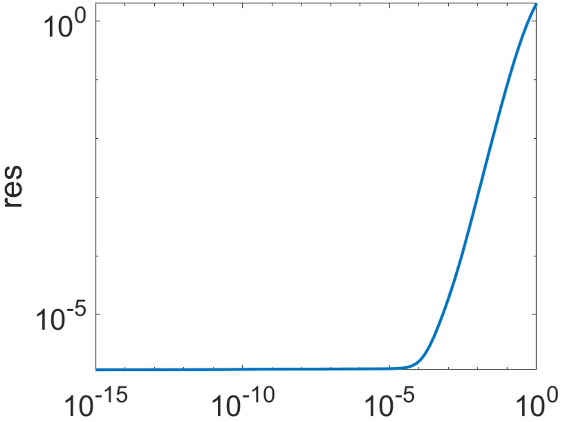
convergence rate

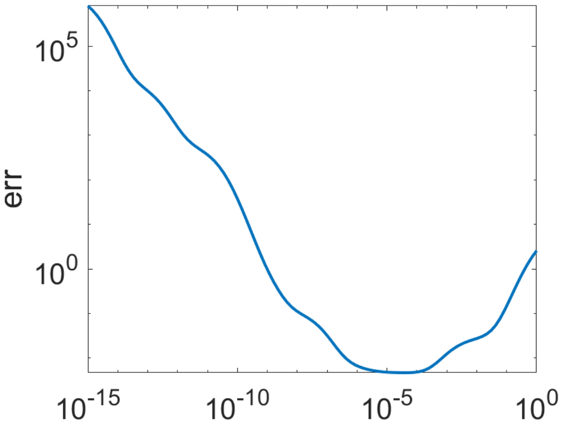
$$\|x_\alpha^\delta - x^\dagger\| \le 2\|w\|^{\frac{1}{2}}\delta^{\frac{1}{2}}.$$

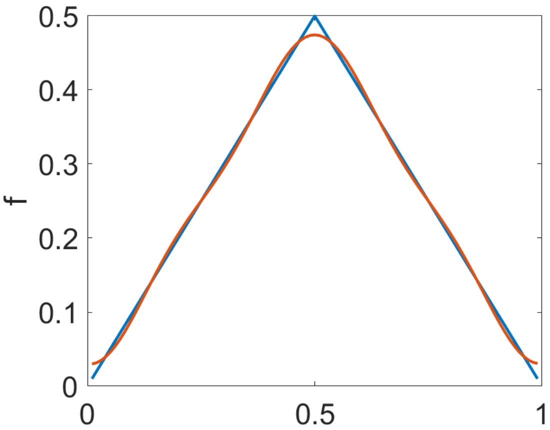
discrepancy principle in action



noisy data, with a small amount of noise ($\epsilon = 1$e-4)


10^0
10^-5
res
10^-15
10^-10
10^-5
10^0

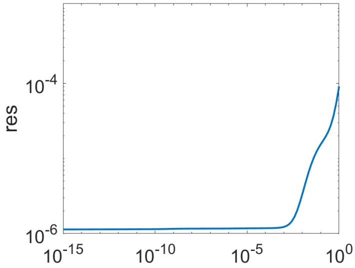

err
10^5
10^0
10^{-15}
10^{-10}
10^{-5}
10^0


0.5
0.4
0.3
0.2
0.1
0
f
0
0.5
1

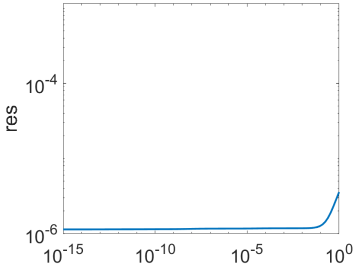
(generalized) Tikhonov regularization

$$J_\alpha(x) = \|Ax - y\|^2 + \alpha\|Lx\|^2$$

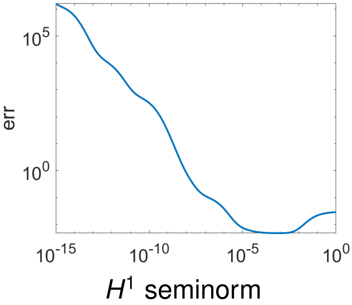
- L: differential operator
- intuition: look for a solution with small derivative $\|x\| \Rightarrow$ weak compactness ...
- the penalty term often induces a (weak) compact subset in Sobolev spaces

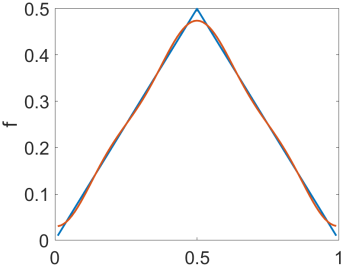


H^1 seminorm

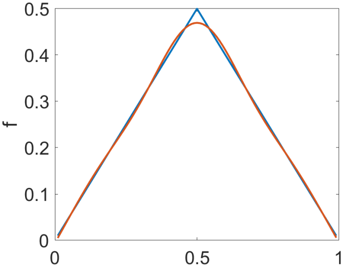


H^2 seminorm


err
10^5
10^0
10^{-15}
10^{-10}
10^{-5}
10^0
H^1 seminorm
err
10^5
10^0
10^{-15}
10^{-10}
10^{-5}
10^0
H^2 seminorm



H^1 seminorm



H^2 seminorm

Bakushinskii's veto

Let $A : X \to Y$ be injective with dense range $\mathcal{R}(A)$ in Y, and A fails to have a bounded inverse in $\mathcal{L}(Y, X)$. Then there is no parameter choice rule $\alpha = \alpha(y^\delta)$, which depends only on the data y^δ, such that for every $y^\dagger \in \mathcal{R}(A)$ and every family $\{y^\delta\}_{\delta>0}$ of approximate data satisfying $y^\delta \to y^\dagger$ as $\delta \to 0$ there holds

$$x^\delta_{\alpha(y^\delta)} \to A^\dagger y^\dagger, \quad \text{as } \delta \to 0, \qquad (*)$$

where $x^\delta_{\alpha(y^\delta)}$ is given by Tikhonov regularization with $\alpha = \alpha(y^\delta)$.

A. Bakushinskii. Remarks on the choice of regularization parameter from quasioptimality and relation tests. Zh. Vychisl. Mat. i Mat. Fiz. 24(8), 1258–1259 (1984)

- For any $y^\dagger \in \mathcal{R}(A)$, define approx.$\{y^\delta\}_{\delta>0}$ with $y^\delta = y^\dagger$, $\forall\delta > 0$

$$(*) \Rightarrow x_{\alpha(y^\dagger)} = A^\dagger y^\dagger$$

- then $(*) \Rightarrow$

$$A^\dagger y^\delta \to A^\dagger y^\dagger, \quad y^\delta \to y^\dagger,$$

  for any (other) set $\{y^\delta\}_{\delta>0} \subset \mathcal{R}(A)$ of approx. of $y^\dagger$
- $\Rightarrow A^\dagger : \mathcal{R}(A) \to X$ is continuous on the dense subset $\mathcal{R}(A)$ of Y and hence has a continuous extension to Y
- contradiction ! $\Rightarrow$ no such parameter choice rule can exist