Inverse Problems

Sommersemester 2023

Vesa Kaarnioja
vesa.kaarnioja@fu-berlin.de

FU Berlin, FB Mathematik und Informatik

Fifth lecture, May 15, 2023

Tikhonov regularization

Tikhonov regularization

The sequence of TSVD solutions $\{x_k\}$ minimizes the norm of the residual

$$\|Ax - y\|$$

as k tends to $\mathrm{rank}(A)$. Unfortunately, when inverse/ill-posed problems are considered, it may also happen that

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

In consequence, it appears reasonable to try minimizing the residual and the norm of the solution *simultaneously*.

Definition

A *Tikhonov regularized solution* $x_\delta \in H_1$ is a minimizer of the Tikhonov functional

$$F_\delta(x) := \|Ax - y\|^2 + \delta\|x\|^2,$$

where $\delta > 0$ is called the *regularization parameter*.

Theorem

Let $A\colon H_1 \to H_2$ be a compact linear operator with the singular system (λ_n, v_n, u_n). Then the Tikhonov regularized solution exists, is unique, and is given by the formula

$$x_\delta = (A^*A + \delta I)^{-1}A^*y = \sum_{n=1}^{p} \frac{\lambda_n}{\lambda_n^2 + \delta}\langle y, u_n\rangle v_n,$$

where $p = \mathrm{rank}(A)$.

Remark. The Tikhonov regularized solution can be obtained without knowing the SVD of A by solving x_δ from $(A^*A + \delta I)x_\delta = A^*y$.

Proof. We make use of the *Lax–Milgram lemma*:

Lemma (Lax–Milgram)

Let H be a Hilbert space, and let $B\colon H \times H \to \mathbb{R}$ be a bilinear quadratic form such that

$$|B(x,y)| \leq C\|x\|\|y\| \quad \textit{for all } x, y \in H,$$
$$B(x,x) \geq c\|x\|^2 \quad \textit{for all } x \in H$$

for some constants $0 < c \leq C < \infty$. Then there exists a unique linear boundedly invertible operator $T\colon H \to H$ such that

$$B(x,y) = \langle x, Ty\rangle \quad \textit{for all } y \in H,$$
$$\|T\| \leq C \quad \textit{and} \quad \|T^{-1}\| \leq \frac{1}{c}.$$

In our case, we define the bilinear operator $B(x,y) := \langle x, (A^*A + \delta I)y\rangle$ and observe that $|B(x,y)| \leq (\|A\|^2 + \delta)\|x\|\|y\|$ (boundedness) and $B(x,x) = \langle x, (A^*A + \delta I)x\rangle = \|Ax\|^2 + \delta\|x\|^2 \geq \delta\|x\|^2$ (coercivity). $\therefore$ $(A^*A + \delta I)^{-1}$ exists such that $\|(A^*A + \delta I)^{-1}\| \leq \frac{1}{\delta}$. In particular, $x_\delta = (A^*A + \delta I)^{-1}A^*y$ is well-defined.

Recall that $Ax = \sum_n \lambda_n \langle x, v_n \rangle u_n$ and $A^* y = \sum_n \lambda_n \langle y, u_n \rangle v_n$. Especially,

$$A^* A x = \sum_n \lambda_n^2 \langle x, v_n \rangle v_n.$$

Since $H_1 = \mathrm{Ker}(A) \oplus \mathrm{Ker}(A)^\perp$, we can write

$$x_\delta = P x_\delta + Q x_\delta = \sum_n \langle x_\delta, v_n \rangle v_n + Q x_\delta,$$

where $P \colon H_1 \to \mathrm{Ker}(A)^\perp = \overline{\mathrm{span}\{v_n\}}$ and $Q \colon H_1 \to \mathrm{Ker}(A)$ are orthogonal projections. Thus

$$(A^* A + \delta I) x_\delta = A^* y \quad \Leftrightarrow \quad \sum_n (\lambda_n^2 + \delta) \langle x_\delta, v_n \rangle v_n + Q x_\delta = \sum_n \lambda_n \langle y, u_n \rangle v_n.$$

Equating terms yields that $Q x_\delta = 0$ and

$$(\lambda_n^2 + \delta) \langle x_\delta, v_n \rangle = \lambda_n \langle y, u_n \rangle \quad \Leftrightarrow \quad \langle x_\delta, v_n \rangle = \frac{\lambda_n}{\lambda_n^2 + \delta} \langle y, u_n \rangle,$$

as desired.

Finally, to show that x_δ minimizes the quadratic functional $F_\delta(x) = \|Ax - y\|^2 + \delta\|x\|^2$, consider

$$x = x_\delta + z,$$

where $z \in H_1$ is arbitrary. Now

$$\begin{aligned} F_\delta(x) &= F_\delta(x_\delta + z) = F_\delta(x_\delta) + \langle z, (A^*A + \delta I)x_\delta - A^*y\rangle + \langle z, (A^*A + \delta I)z\rangle \\ &= F_\delta(x_\delta) + \langle z, (A^*A + \delta I)z\rangle, \end{aligned}$$

by definition of x_δ.The last term is nonnegative and vanishes only if $z = 0$. This proves the claim. $\square$

Morozov discrepancy principle for Tikhonov regularization

Suppose that the measurement $y \in H_2$ is a noisy version of some underlying "exact" data $y_0 \in H_2$, and that

$$\|y - y_0\| \approx \varepsilon > 0.$$

In the framework of Tikhonov regularization, the Morozov discrepancy principle tells us to choose the regularization parameter $\delta > 0$ so that the residual satisfies

$$\|y - Ax_\delta\| = \varepsilon.$$

It turns out that there is a unique regularization parameter satisfying this condition if

$$\|y - Py\| < \varepsilon < \|y\|,$$

where $P\colon H_2 \to \overline{\mathrm{Ran}(A)}$ is an orthogonal projection.

Properties of the Tikhonov regularized solution

Theorem

Let $A\colon H_1 \to H_2$ be a compact linear operator with the singular system (λ_n, v_n, u_n). Let $P\colon H_2 \to \overline{\mathrm{Ran}(A)}$ be an orthogonal projection. Then we have the following:

(i) $\delta \mapsto \|Ax_\delta - y\|$ *is a strictly increasing function of* $\delta > 0$.

(ii) $\|Py - y\| = \lim_{\delta\to 0+} \|Ax_\delta - y\| \le \|Ax_\delta - y\| \le \lim_{\delta\to\infty} \|Ax_\delta - y\| = \|y\|$.

(iii) *If* $Py \in \mathrm{Ran}(A)$, *then* x_δ *converges to the solution of the problem*

$$Ax = Py \quad \text{and} \quad x \perp \mathrm{Ker}(A)$$

as $\delta \to 0+$.

Corollary

The equation $\|Ax_\delta - y\| = \varepsilon$ *has a unique solution* $\delta = \delta(\varepsilon)$ *iff* $\|(I - P)y\| < \varepsilon < \|y\|$.

Interpretation: $\|(I-P)y\| < \varepsilon$ means that any component in the data y orthogonal to the range of A must be due to noise; $\varepsilon < \|y\|$ means that the error level should not exceed the signal level.

Proof. Suppose that the operator A has the SVD

$$Ax = \sum_n \lambda_n \langle x, v_n \rangle u_n.$$

Then $Av_n = \lambda_n u_n$, the orthogonal projection $P \colon H_2 \to \overline{\mathrm{Ran}(A)}$ is

$$Py = \sum_n \langle y, u_n \rangle u_n,$$

and the Tikhonov regularized solution x_δ and its image under A are

$$x_\delta = \sum_n \frac{\lambda_n}{\lambda_n^2 + \delta} \langle y, u_n \rangle v_n \quad \Rightarrow \quad Ax_\delta = \sum_n \frac{\lambda_n^2}{\lambda_n^2 + \delta} \langle y, u_n \rangle u_n.$$

(i) It follows that

$$\begin{aligned}
\|Ax_\delta - y\|^2 &= \|Ax_\delta - Py\|^2 + \|(I-P)y\|^2 \\
&= \sum_n \left(\frac{\lambda_n^2}{\lambda_n^2 + \delta} - 1 \right)^2 |\langle y, u_n \rangle|^2 + \|(I-P)y\|^2 \\
&= \sum_n \left(\frac{\delta}{\lambda_n^2 + \delta} \right)^2 |\langle y, u_n \rangle|^2 + \|(I-P)y\|^2.
\end{aligned}$$

We arrived at

$$\|Ax_\delta - y\|^2 = \sum_n \left(\frac{\delta}{\lambda_n^2 + \delta}\right)^2 |\langle y, u_n\rangle|^2 + \|(I-P)y\|^2.$$

For each term of the sum,

$$\frac{\mathrm{d}}{\mathrm{d}\delta}\left(\frac{\delta}{\lambda_n^2 + \delta}\right)^2 = \frac{2\delta\lambda_n^2}{(\lambda_n^2+\delta)^3} > 0,$$

implying that the mapping $\delta \mapsto \|Ax_\delta - y\|^2$ is strictly increasing.

(ii) It is easy to see that

$$\|Ax_\delta - y\|^2 = \sum_n \left(\frac{\delta}{\lambda_n^2 + \delta}\right)^2 |\langle y, u_n\rangle|^2 + \|(I-P)y\|^2 \xrightarrow{\delta\to 0+} \|(I-P)y\|^2,$$

$$\|Ax_\delta - y\|^2 = \sum_n \left(\frac{\delta}{\lambda_n^2 + \delta}\right)^2 |\langle y, u_n\rangle|^2 + \|(I-P)y\|^2$$

$$\xrightarrow{\delta\to\infty} \|Py\|^2 + \|(I-P)y\|^2 = \|y\|^2.$$

(iii) Let $Py \in \operatorname{Ran}(A)$. This implies that there exists $x \in \operatorname{Ker}(A)^{\perp}$ such that $Ax = Py$; this is the minimum norm solution

$$x = \sum_n \frac{1}{\lambda_n} \langle y, u_n \rangle v_n,$$

for which it can be shown that

$$x_\delta = \sum_n \frac{\lambda_n}{\lambda_n^2 + \delta} \langle y, u_n \rangle v_n \xrightarrow{\delta \to 0+} \sum_n \frac{1}{\lambda_n} \langle y, u_n \rangle v_n = x. \quad \square$$

Remark. In parts (ii) and (iii), one should take care when interchanging the order of the limit and the summation, i.e., justifying the steps

$$\lim_{\lambda \to 0+} \sum_n = \sum_n \lim_{\lambda \to 0+} \quad \text{and} \quad \lim_{\lambda \to \infty} \sum_n = \sum_n \lim_{\lambda \to \infty}.$$

Standard techniques involve the monotone convergence theorem and the dominated convergence theorem (note that these apply to infinite series as well as integrals). In part (iii), it is helpful to observe that $x_\delta \xrightarrow{\delta \to 0+} x$ iff $\langle x_\delta, \phi \rangle \xrightarrow{\delta \to 0+} \langle x, \phi \rangle$ for all $\phi \in H_1$ and $\|x_\delta\| \xrightarrow{\delta \to 0+} \|x\|$.

Tikhonov regularization with matrices

Consider the special case $H_1 = \mathbb{R}^n$ and $H_2 = \mathbb{R}^m$ corresponding to the matrix equation $y = Ax$. The Tikhonov functional takes the special form

$$F_\delta(x) = \left\| \begin{bmatrix} A \\ \sqrt{\delta} I \end{bmatrix} x - \begin{bmatrix} y \\ 0 \end{bmatrix} \right\|^2, \quad I \in \mathbb{R}^{n \times n},\ 0 \in \mathbb{R}^n.$$

The minimizer can be found by solving the least squares problem

$$\begin{bmatrix} A \\ \sqrt{\delta} I \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} A \\ \sqrt{\delta} I \end{bmatrix} x = \begin{bmatrix} A \\ \sqrt{\delta} I \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} y \\ 0 \end{bmatrix}$$

or, equivalently,

$$(A^{\mathrm{T}} A + \delta I) x = A^{\mathrm{T}} y.$$

In MATLAB, this can be implemented simply as follows:

```
K = [A;sqrt(delta)*eye(n)];
z = [y; zeros(n,1)];
xdelta = K\z;
```

In Python, e.g., `scipy.linalg.lstsq` can be used to obtain the least squares solution. For sparse matrices, e.g.,

```
xdelta =
scipy.sparse.linalg.lsqr(A,y,damp=numpy.sqrt(delta))[0].
```

Numerical example: backward heat equation

Let us revisit the backward heat equation from earlier:

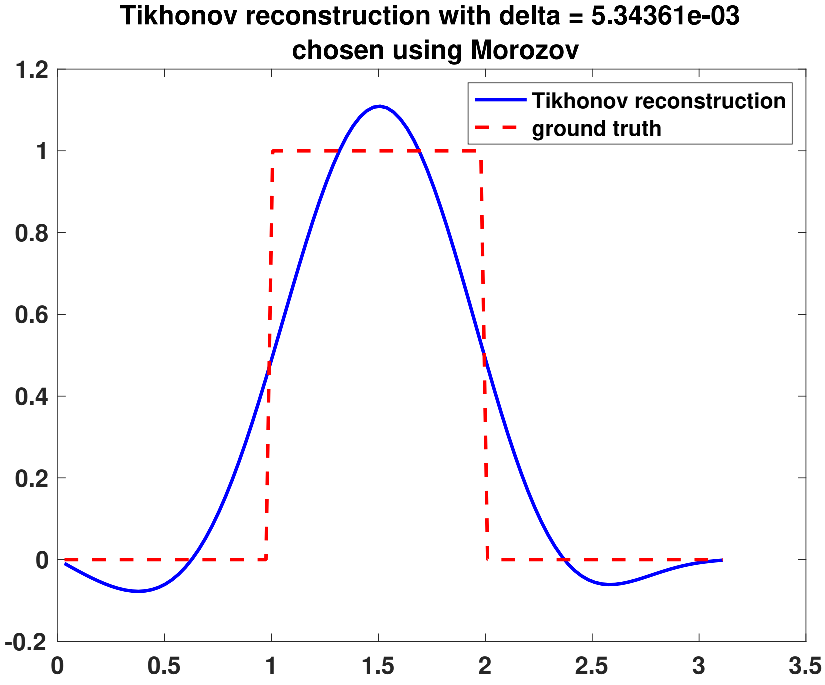
$$\begin{cases} \partial_t u(x,t) = \partial_x^2 u(x,t) & \text{for } (x,t) \in (0,\pi) \times \mathbb{R}_+, \\ u(0,\cdot) = u(\pi,\cdot) = 0 & \text{on } \mathbb{R}_+, \\ u(\cdot,0) = f & \text{on } (0,\pi), \end{cases}$$

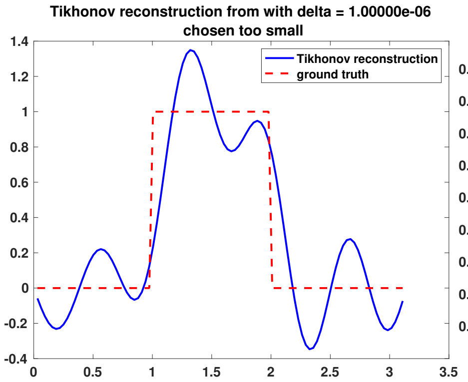
where $f\colon (0,\pi) \to \mathbb{R}$ is the initial heat distribution.

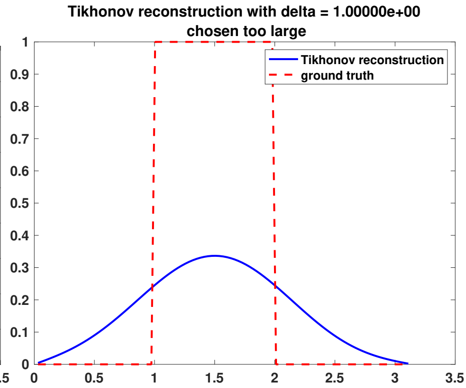
We reconstruct the initial state f based on noisy measurements of $u(\cdot, T)$ at time $T > 0$ using Tikhonov regularization.

We assume that the data $U(T) \in \mathbb{R}^{99}$ at time $T = 0.1$ is contaminated with mean-zero Gaussian noise with standard deviation 0.01, and that the discrepancy between the measured data and the underlying "exact" data equals the square root of the expected value of the squared norm of the noise vector, i.e.,

$$\varepsilon = \sqrt{99 \cdot 0.01^2} \approx 0.0995.$$


Tikhonov reconstruction with delta = 5.34361e-03
chosen using Morozov
Tikhonov reconstruction
ground truth
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
0
0.5
1
1.5
2
2.5
3
3.5





See the files `heateq_tikhonov.py` / `heateq_tikhonov.m` on the course webpage!

Tikhonov regularization for nonlinear problems

Unlike the TSVD, Tikhonov regularization can be generalized to nonlinear problems as well. Consider a *nonlinear* operator $A\colon H_1 \to H_2$ and the problem

$$y = A(x).$$

A standard way of solving such a problem is via sequential linearizations, which leads to solving a set of linear problems involving the *Fréchet derivative* of operator A.

Definition

The function $A\colon H_1 \to H_2$ is called *Fréchet differentiable* at $x_0 \in H_1$ if there exists a continuous linear operator $A'_{x_0}\colon H_1 \to H_2$ such that

$$A(x+h) = A(x) + A'_{x_0} h + W_{x_0}(z),$$

where $\|W_{x_0}(z)\| \leq \varepsilon(x_0, z)\|z\|$ and the functional $z \mapsto \varepsilon(x_0, z)$ tends to zero as $z \to 0$.

The linear operator A'_{x_0} is called the *Fréchet derivative* of A at x_0.

We are interested in minimizing

$$F_\delta(x) = \|A(x) - y\|^2 + \delta\|x\|^2, \quad \delta > 0.$$

Since F_δ is no longer quadratic, it is unclear whether a unique minimizer exists and typically the minimizer cannot be given by an explicit formula even it exists.

Let A be Fréchet differentiable. The linearization of A around a given point x_0 leads to the approximation of the functional F_δ,

$$\begin{aligned} F_\delta(x) \approx \widetilde{F}_\delta(x; x_0) &= \|A(x_0) + A'_{x_0}(x - x_0) - y\|^2 + \delta\|x\|^2 \\ &= \|A'_{x_0}(x) - g(y, x_0)\|^2 + \delta\|x\|^2, \end{aligned}$$

where $g(y, x_0) := y - A(x_0) + A'_{x_0}(x_0)$.

From the previous discussion on the linear case, we know that the minimizer of $\widetilde{F}_\delta(x; x_0)$ is given by

$$x = ((A'_{x_0})^* A_{x_0} + \delta I)^{-1}(A'_{x_0})^* g(y, x_0).$$

Minimization strategy with step size control

It may happen that the solution of the linearized problem does not reflect adequately the nonlinearities of the original function. A better strategy is to implement some form of step size control. For example, we might design the following iterative method.

1. Pick an initial guess x_0 and set $k = 0$.

   Repeat:

   2. Calculate the Fréchet derivative A'_{x_0}.
   3. Determine

      $$x = ((A'_{x_k})^* A'_{x_k} + \delta I)^{-1} (A'_{x_k})^* g(y, x_k), \quad g(y, x_k) = y - A(x_k) + A'_{x_k} x_k,$$

      and define $\Delta x = x - x_k$.
   4. Find step size $s > 0$ by minimizing the function

      $$f(s) = \|A(x_k + s\Delta x) - y\|^2 + \|x_k + s\Delta x\|^2.$$

   5. Set $x_{k+1} = x_k + s\Delta x$ and increase $k \leftarrow k + 1$.

   until convergence.

Remarks on nonlinear Tikhonov regularization

- In practice, evaluating A'_{x_k} is often the most difficult part.
- For finite-dimensional operators, the Fréchet derivative is simply the Jacobi matrix.
- Depending on the nature of the nonlinearity, one might also consider more "specialized" optimization methods (e.g., Gauss–Newton algorithm, Levenberg–Marquardt algorithm…).

More general penalty terms

A more general way of defining the Tikhonov functional is

$$F_\delta(x) = \|Ax - y\|^2 + \delta G(x),$$

where $G \colon H_1 \to \mathbb{R}_{\geq 0}$ takes non-negative values. The existence of a unique minimizer for this kind of functional depends on the properties of G, as does the workload needed for finding it.

One typical way of defining G is

$$G(x) = \|L(x - x_0)\|^2,$$

where $x_0 \in H_1$ is a given reference vector and L is some linear operator. The choice of x_0 and L reflects our prior knowledge about "feasible" solutions: Lx is some property that is known to be relatively close to the reference value Lx_0 for all reasonable solutions. (In the standard case $x_0 = 0$ and $L = I$, the solutions are "known" to lie relatively close to the origin.)

The numerical implementation of Tikhonov regularization with $G(x) = \|L(x - x_0)\|^2$ is approximately as easy as for the standard penalty term.

In the case where $H_1 = \mathbb{R}^n$ and $H_2 = \mathbb{R}^m$, the operator L is some matrix in $\mathbb{R}^{\ell+n}$ and the Tikhonov functional can be given as

$$F_\delta(x) = \left\| \begin{bmatrix} A \\ \sqrt{\delta} L \end{bmatrix} x - \begin{bmatrix} y \\ \sqrt{\delta} L x_0 \end{bmatrix} \right\|^2.$$

Assuming that the singular values of K are bounded suitably far away from zero, the Tikhonov solution can be computed in MATLAB as

```
K = [A; sqrt(delta)*L];
z = [y; sqrt(delta)*L*x0];
xdelta = K\z;
```

In Python, e.g., `scipy.linalg.lstsq` can be used to solve the equivalent least squares problem $\begin{bmatrix} A \\ \sqrt{\delta} L \end{bmatrix} x = \begin{bmatrix} y \\ \sqrt{\delta} L x_0 \end{bmatrix}$. For sparse matrices, e.g.,

```
K = scipy.sparse.vstack((A,np.sqrt(delta)*L))
z = np.hstack((y,np.sqrt(delta)*L@x0))
xdelta = scipy.sparse.linalg.lsqr(K,z)[0]
```