Computational inverse problems

Nuutti Hyvönen and Pauliina Hirvi

nuutti.hyvonen@aalto.fi, pauliina.hirvi@aalto.fi

Third lecture

Summary of the previous lecture

The truncated SVD solution: For $\mathbb{N} \ni k \leq \mathrm{rank}(A)$, there exist unique $x_k \in H_1$ such that

$$Ax_k = P_k y \qquad \text{and} \qquad x_k \perp \mathrm{Ker}(A).$$

where $P_k : H_2 \to \mathrm{span}\{u_1, \ldots, u_k\}$ is an orthogonal projection. This solution can be given as

$$x_k = \sum_{n=1}^{k} \frac{1}{\lambda_n} \langle y, u_n \rangle v_n.$$

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

$$A = U\Lambda V^{\mathrm{T}},$$

where $\Lambda \in \mathbb{R}^{m\times n}$ has the (non-negative!) 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 composed of the (extended!) orthonormal basis $\{v_j\}_{j=1}^{n}$ and $\{u_j\}_{j=1}^{m}$, respectively.

The truncated SVD solution for $1 \le k \le p := \mathrm{rank}(A)$ is given by

$$x_k = V\Lambda_k^{\dagger}U^{\mathrm{T}}y$$

where $\Lambda_k^{\dagger} \in \mathbb{R}^{n\times m}$ has the elements $1/\lambda_1, \ldots, 1/\lambda_k, 0, \ldots, 0$ on its diagonal. The matrix $A^{\dagger} = V\Lambda_p^{\dagger}U^{\mathrm{T}}$ is called the Moore–Penrose pseudoinverse of A.

Morozov discrepancy principle

(Let us return to the case where H_1 and H_2 are general separable real Hilbert spaces, and $A : H_1 \to H_2$ is a compact linear operator.)

To make the truncated SVD a more useful tool, one should come up with some rule for choosing the spectral cut-off index $k \geq 1$ appearing in the truncated SVD problem

$$Ax = P_k y \qquad \text{and} \qquad x \perp \mathrm{Ker}(A).$$

Unfortunately, it is difficult (if not impossible) to invent a reliable general scheme of doing this.

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

Assume that the measurement $y \in H_2$ is a noisy version of some underlying 'exact' data vector $y_0 \in H_2$. Furthermore, suppose that we have some estimate on the discrepancy between y and y_0, i.e.,

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

For example, it may be known that

$$y = y_0 + n,$$

where the vector $n \in H_2$ is a realization of some random variable with known probability distribution. Knowledge about the statistics of n could be due to, e.g., calibrations of the measurement device.

The idea of the Morozov discrepancy principle is to choose the smallest $k \geq 1$ such that the *residual* satisfies

$$\|y - Ax_k\| \leq \epsilon.$$

Intuitively this means that we cannot expect the approximate solution to yield a smaller residual than the measurement error — otherwise we would be fitting the solution to noise.

Does such k exist?

Yes, it does if $\epsilon > \|Py - y\|$, as explained below.

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

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

which is the best one can do since $\inf_{z \in \mathrm{Ran}(A)} \|z - y\| = \|Py - y\|$ by virtue of the *projection theorem*. (However, there is no guarantee that $\|x_k\|$ would not explode as $k \to \infty$.)

On the other hand, if $p = \mathrm{rank}(A) < \infty$,

$$\|Ax_p - y\| \;=\; \|P_p y - y\| \;=\; \|Py - y\|.$$

(Usually, one should not choose this large *spectral cut-off* in practice.)

2.3 Tikhonov regularization

Motivation of Tikhonov regularization

As pointed out on the previous slide, 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, when inverse/ill-posed problems are considered, we typically also have

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

(If $\mathrm{rank}(A) = \infty$, this can be understood literally; if $\mathrm{rank}(A) = p < \infty$, this should be understood in the sense that the x_p is usually rubbish — especially, if the measurement y is noisy.)

As a consequence, it seems well-motivated to try minimizing the residual and the norm of the solution *simultaneously*.

Tikhonov regularized solution

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

$$x_\delta = (A^*A + \delta I)^{-1}A^*y = \sum_{j=1}^{p} \frac{\lambda_j}{\lambda_j^2 + \delta}\langle y, u_j\rangle v_j,$$

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

Proof: Let us prove this claim only in the case that $H_1 = \mathbb{R}^n$ and $H_2 = \mathbb{R}^m$; the general result follows from the same ideas, but requires some more sophisticated functional analysis.

To begin with, note that

$$x^{\mathrm{T}}(A^{\mathrm{T}}A + \delta I)x \; = \; \|Ax\|^2 + \delta\|x\|^2 \; \geq \; \delta\|x\|^2 > 0$$

if $x \neq 0$. In particular, $A^{\mathrm{T}}A + \delta I \in \mathbb{R}^{n\times n}$ is injective, which means that it is invertible due to the *fundamental theorem of linear algebra*.

Hence,

$$x_\delta \; := \; (A^{\mathrm{T}}A + \delta I)^{-1}A^{\mathrm{T}}y \in H_1$$

is well-defined.

Let $\{\lambda_j\}_{j=1}^p$ be the positive singular values of A, and $\{v_j\}_{j=1}^p$ and $\{u_j\}_{j=1}^p$ the corresponding sets of singular vectors that span $\mathrm{Ker}(A)^\perp$ and $\mathrm{Ran}(A)$, respectively.

We expand $x_\delta = \sum (v_j^{\mathrm{T}} x_\delta) v_j \; + \; Q x_\delta$, where $Q : \mathbb{R}^n \to \mathrm{Ker}(A)$ is an orthogonal projection. According to the first exercise of the first exercise session,

$$(A^{\mathrm{T}} A + \delta I) x_\delta \; = \; \sum_{j=1}^{p} (\lambda_j^2 + \delta)(v_j^{\mathrm{T}} x_\delta) v_j + \delta \, Q x_\delta .$$

Similarly,

$$A^{\mathrm{T}} y \; = \; \sum_{j=1}^{p} \lambda_j (u_j^{\mathrm{T}} y) v_j .$$

Equating these two expressions results in

$$(v_j^{\mathrm{T}} x_\delta) \;=\; \frac{\lambda_j}{\lambda_j^2 + \delta}(u_j^{\mathrm{T}} y), \qquad 1 \le j \le p,$$

and $Qx_\delta = 0$, which altogether means that

$$x_\delta \;=\; \sum_{j=1}^{p} \frac{\lambda_j}{\lambda_j^2 + \delta}(u_j^{\mathrm{T}} y) v_j \;=\; \sum_{j=1}^{p} \frac{\lambda_j}{\lambda_j^2 + \delta}\langle y, u_j\rangle v_j .$$

Finally, consider $x = x_\delta + z$, where $z \in \mathbb{R}^n$ is arbitrary. We have

$$
\begin{aligned}
F_\delta(x) &= \|(Ax_\delta - y) + Az\|^2 + \delta\|x_\delta + z\|^2 \\
&= \|Ax_\delta - y\|^2 + 2\,(Az)^{\mathrm{T}}(Ax_\delta - y) + \|Az\|^2 \\
&\quad + \delta\left(\|x_\delta\|^2 + 2z^{\mathrm{T}}x_\delta + \|z\|^2\right) \\
&= F_\delta(x_\delta) + \|Az\|^2 + \delta\|z\|^2 \\
&\quad +2z^{\mathrm{T}}\left((A^{\mathrm{T}}A + \delta I\right)x_\delta - A^{\mathrm{T}}y) \\
&= F_\delta(x_\delta) + \|Az\|^2 + \delta\|z\|^2 \;\geq\; F_\delta(x_\delta),
\end{aligned}
$$

where the equality holds if and only if $z = 0$. This shows that $x_\delta = (A^{\mathrm{T}}A + \delta I)^{-1}A^{\mathrm{T}}y$ is the unique minimizer of the Tikhonov functional. □

Properties of the Tikhonov regularized solution

The Tikhonov regularized solution has the following intuitive properties. The proof of this theorem is omitted.

Theorem. *Let* $P : H_2 \to \overline{\mathrm{Ran}(A)}$ *be an orthogonal projection. The residual* $\|Ax_\delta - y\|$ *is strictly increasing as a function of* δ *and it satisfies*

$$\lim_{\delta \to 0} \|Ax_\delta - y\| = \|Py - y\| \qquad \text{and} \qquad \lim_{\delta \to \infty} \|Ax_\delta - y\| = \|y\|.$$

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

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

as $\delta \to 0$*. On the other hand, if* $Py \notin \mathrm{Ran}(A)$*, then the norm* $\|x_\delta\|$ *tends to infinity as* δ *goes to zero.*

The Morozov principle for Tikhonov regularization

Assume once again that the measurement $y \in H_2$ is a noisy version of some underlying 'exact' data vector $y_0 \in H_2$, and that

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

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

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

Such a regularization parameter exists if

$$\|y - Py\| < \epsilon < \|y\|.$$

This follows from the above theorem because the residual $\|y - Ax_\delta\|$ is continuous with respect to δ.

Tikhonov regularized solution for matrices

Assume once again that $H_1 = \mathbb{R}^n$ and $H_2 = \mathbb{R}^m$. In this case, the Tikhonov functional can be given as

$$F_\delta(x) = \left\| \begin{bmatrix} A \\ \sqrt{\delta} I \end{bmatrix} x - \begin{bmatrix} y \\ 0 \end{bmatrix} \right\|^2, \qquad I \in \mathbb{R}^{n \times n},\ 0 \in \mathbb{R}^n. \tag{5}$$

It is interesting to notice that the normal equation corresponding to this *least squares problem* is (see 3. problem of 1. exercise session)

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

or equivalently

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

Bear in mind that one does not, actually, need to form this normal equation in Matlab when using Tikhonov regularization: After defining

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

the command

```
xdelta = K\z
```

computes the Tikhonov regularized solution.

Explanation: For non-square matrices the `mldivide` command of Matlab tries to solve the corresponding least squares problem. As unique minimizer is known to exist, this corresponds to multiplying z from the left by the Moore–Penrose pseudoinverse of K (see 3. exercise of 1. session). As all n singular values of K are larger than $\sqrt{\delta}$ (see 1. exercise of 2. session) this pseudoinverse is well-behaved.

An example: Heat distribution in a rod (revisited)

Recall the discretized inverse heat conduction problem that was discussed during the second and third lectures. Let `w` be the simulated heat distribution at `T=0.1` with the ‘wedge function’ as the initial data, and `A` the corresponding propagation matrix `A=expm(TB)`. We add the same small amount of noise as previously and compute the Tikhonov regularized solution:

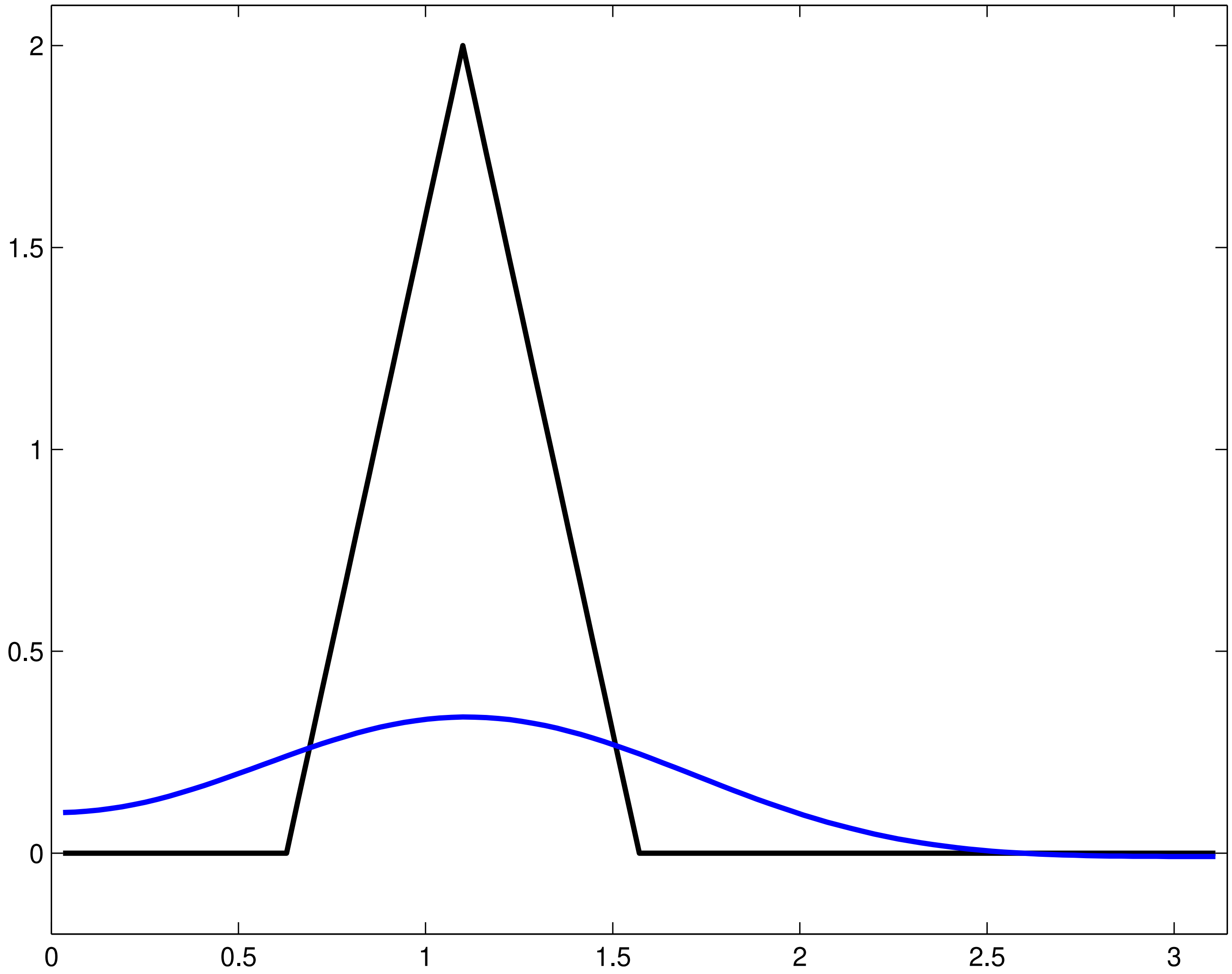
```
wn = w + 0.001*randn(N-1,1);
zn = [wn; zeros(N-1,1)]; %  augmented data vector
K = [A; sqrt(delta)*eye(N-1)]; % augmented system matrix
fdelta = K\zn; % Tikhonov regularized solution
```

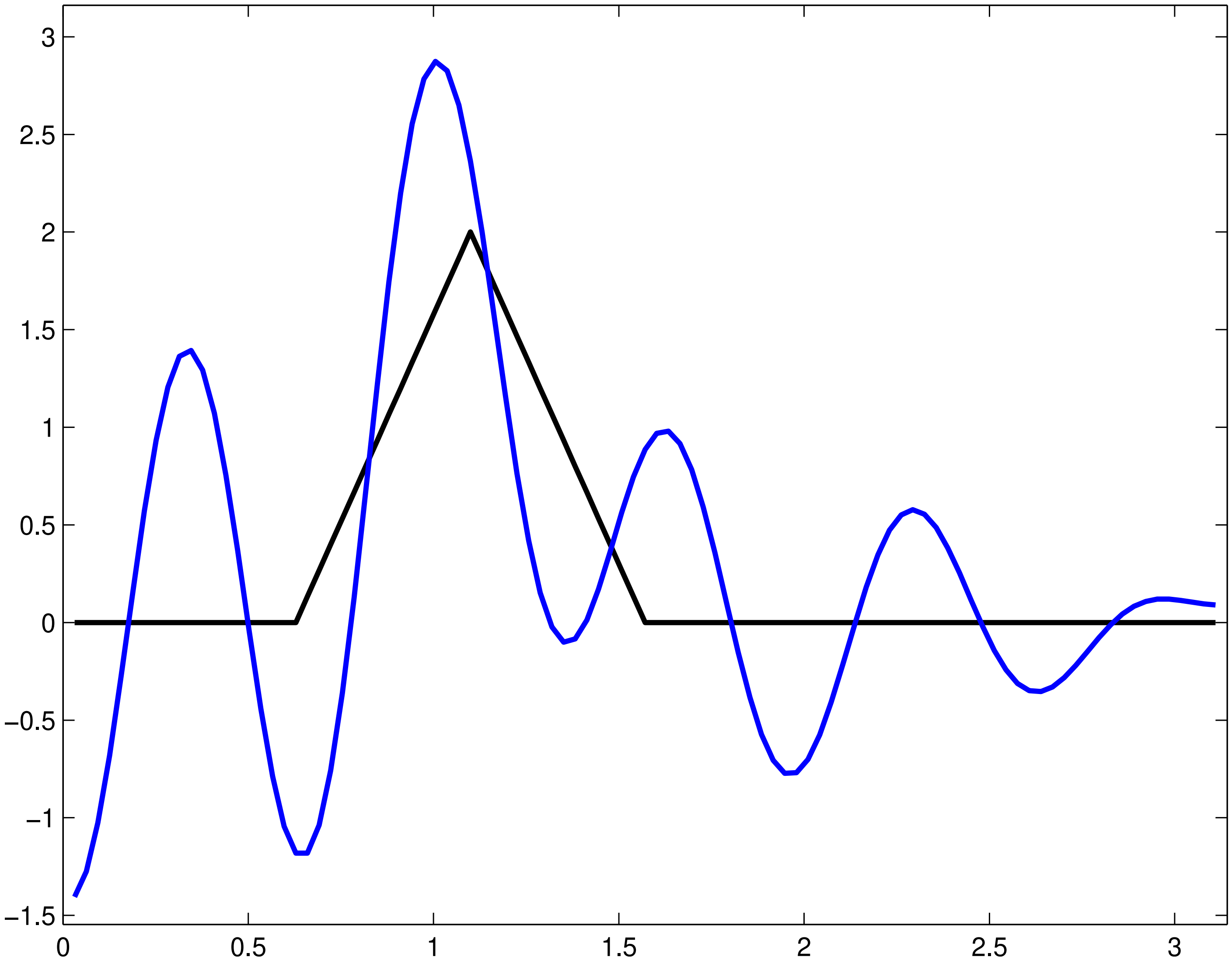
We do this for three different values of the regularization parameter $\delta = 1$ (too large), $\delta = 10^{-8}$ (too small), and $\delta = 5.95 \cdot 10^{-5}$, which corresponds to the Morozov discrepancy principle: We assume here that the discrepancy between the measured data and the underlying 'exact' data equals the square root of the expectation value of the squared norm of the noise vector, i.e.,

$$\epsilon \; = \; \sqrt{99 \cdot 0.001^2} \; \approx \; 9.95 \cdot 10^{-3}.$$

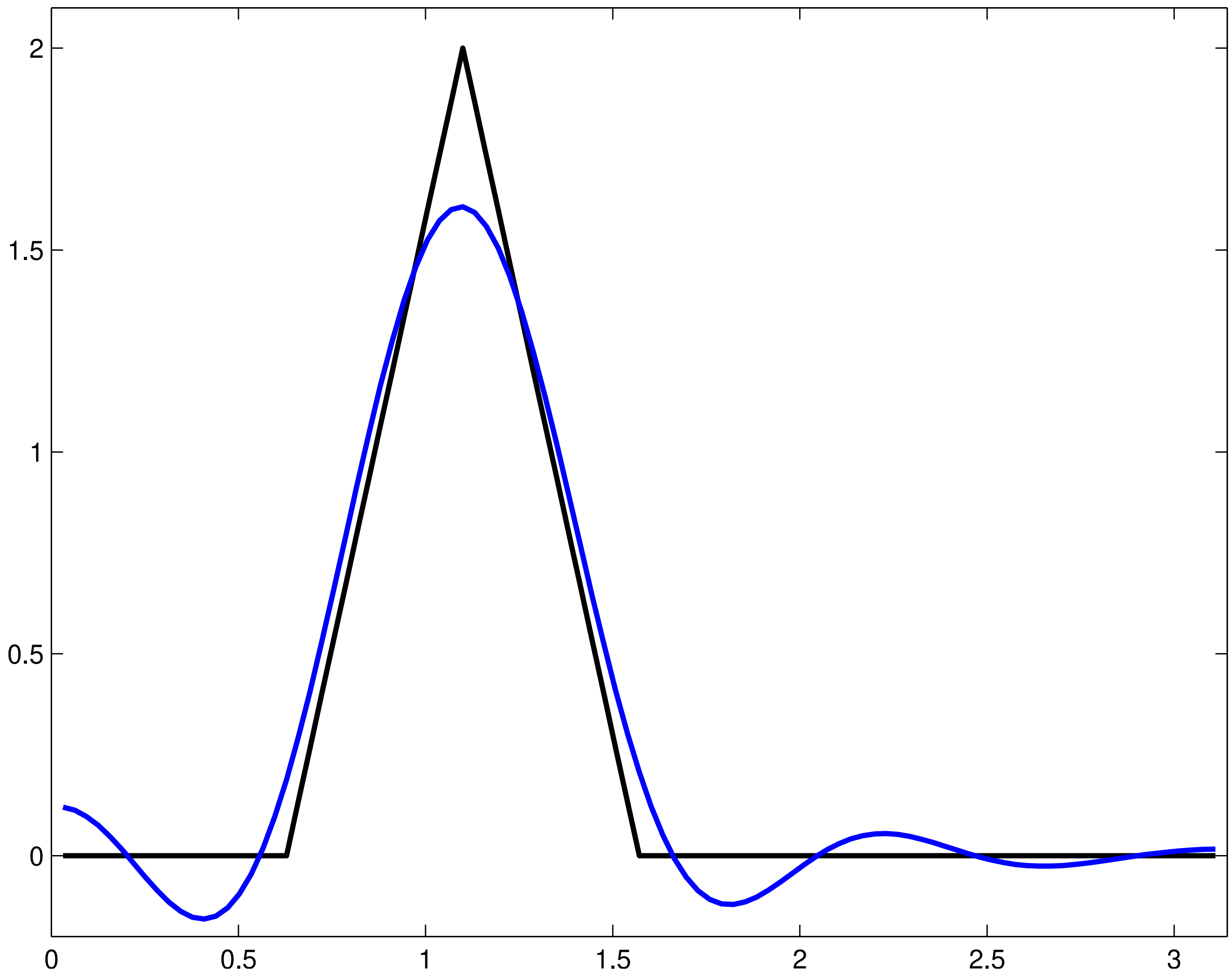
Note that the value of δ given by the discrepancy principle depends on the particular realization of the noise vector even though ϵ does not.

The expectation value of the norm of the noise vector would be as — if not more — logical choice for ϵ, but it is more difficult to write down explicitly. (Luckily, these two choices do not differ that much in the considered case: numerical tests suggest that the latter gives $\epsilon \approx 9.92 \cdot 10^{-3}$.)


2
1.5
1
0.5
0
0
0.5
1
1.5
2
2.5
3



3
2.5
2
1.5
1
0.5
0
−0.5
−1
−1.5
0
0.5
1
1.5
2
2.5
3


2
1.5
1
0.5
0
0
0.5
1
1.5
2
2.5
3

Generalizations of Tikhonov regularization

Tikhonov regularization for nonlinear problems

Let us briefly consider the nonlinear case, where $A : H_1 \to H_2$ is a nonlinear operator and the examined equation is of the form

$$A(x) = y.$$

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

As an example, in Newton's method one would first pick an initial guess $x_0 \in H_1$ and then try to produce the $(j+1)$th iterate by solving the linearized problem

$$A(x_j) + A'(x_j)(x_{j+1} - x_j) = y, \qquad j = 0, 1, \ldots,$$

recursively for x_{j+1}. (In the general setting A' is the *Fréchet derivative* of A, but for finite-dimensional operators it is just the Jacobian matrix.)

Unfortunately, if large alterations of x produce only small changes in $A(x)$, i.e., if the original equation is ill-posed, there is no guarantee that the corresponding linearized problems can be solved as such — not even in the least squares sense. Hence, regularization is needed.

Unlike the truncated SVD method, Tikhonov regularization generalizes easily to this nonlinear framework. Now, it amounts to searching for $x_\delta \in H_1$ that minimizes the functional

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

Since F_δ is no longer quadratic in x, it is not clear that a unique minimizer exists. Furthermore, even if a Tikhonov regularized solution exists, it cannot usually be given by an explicit formula.

Be that as it may, one can try to minimize $F_\delta(x)$ by using some nonlinear optimization technique. One — but probably not the best — way of doing this, is to pick an initial guess $x_{\delta,0} \in H_1$ and then recursively define the $(j+1)$th iterate $x_{\delta,j+1} \in H_1$ to be the unique minimizer of the $x_{\delta,j}$-dependent Tikhonov functional

$$\begin{aligned}\tilde{F}_{\delta,j}(x) &= \left\|A(x_{\delta,j}) + A'(x_{\delta,j})(x - x_{\delta,j}) - y\right\|^2 + \delta\|x\|^2 \\ &= \left\|A'(x_{\delta,j})x - [y - A(x_{\delta,j}) + A'(x_{\delta,j})x_{\delta,j}]\right\|^2 + \delta\|x\|^2,\end{aligned}$$

where the dependence of A on x has been linearized with $x_{\delta,j}$ as the base point. Since this Tikhonov functional is of the 'standard form', $x_{\delta,j+1}$ can be given explicitly with the help of $A'(x_{\delta,j})$, $A(x_{\delta,j})$, $x_{\delta,j}$, y and δ. (In practice, evaluating $A'(x_{\delta,j})$ is often the most difficult part.)

Combining this with some reasonable stopping criterion does indeed give reasonable solutions for many nonlinear inverse problems.

More general penalty terms

A more general way of defining the Tikhonov functional is

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

where the penalty function $G : H_1 \to \mathbb{R}$ 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

$$G(x) \;=\; \|L(x - x_0)\|^2, \tag{6}$$

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 the 'feasible' solutions: Lx is some property that is known to be relatively close to the reference value Lx_0 for all reasonable solutions. (In 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 of (6) is approximately as easy as for the standard penalty term:

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

$$F_\delta(x) \;=\; \|Kx - z\|^2 \tag{7}$$

where

$$K \;=\; \begin{bmatrix} A \\ \sqrt{\delta} L \end{bmatrix} \qquad \text{and} \qquad z = \begin{bmatrix} y \\ \sqrt{\delta} L x_0 \end{bmatrix}.$$

Assuming that the matrix L is chosen so cleverly that all n singular values of K are (well) larger than zero, the Tikhonov regularized solution can be computed in Matlab by applying the pseudoinverse of K on z by the command

```
xdelta = K\z
```

Explanation: As shown in 3. exercise of 1. session, all minimizers of (7) satisfy the normal equation

$$K^{\mathrm{T}} K x = K^{\mathrm{T}} z.$$

On the other hand, it was proved in 1. exercise of 1. session that the symmetric matrix $K^{\mathrm{T}} K \in \mathbb{R}^{n \times n}$ has n positive eigenvalues that are the squares of the singular values of K. In particular, this means that $K^{\mathrm{T}} K$ is invertible, and thus there is exactly one minimizer for (7). This is given by $K^{\dagger} z$ due to 3. exercise of 1. session.

(The fact that a symmetric matrix with nonzero eigenvalues is invertible follows, e.g., from the eigenvalue decomposition.)