Tikhonov Regularization in General Form §8.1

To introduce a more general formulation, let us return to the continuous formulation of the first-kind Fredholm integral equation.

In this setting, the residual norm for the generic problem is

$$R(f) = \left\| \int_0^1 K(s,t)\, f(t)\, dt - g(s) \right\|_2 .$$

In the same setting, we can introduce a *smoothing norm* $S(f)$ that measures the regularity of the solution f. Common choices of $S(f)$ belong to the family given by

$$S(f) = \|f^{(d)}\|_2 = \left(\int_0^1 \left(f^{(d)}(t)\right)^2 dt \right)^{1/2}, \qquad d = 0, 1, 2, \ldots,$$

where $f^{(d)}$ denotes the dth derivative of f.

Then we can write the Tikhonov regularization problem for f in the form

$$\min \left\{ R(f)^2 + \lambda^2 S(f)^2 \right\}, \qquad (1)$$

where λ plays the same role as in the discrete setting.

The previous discrete Tikhonov formulation is merely a special version of this general Tikhonov problem with $S(f) = \|f\|_2$.

We obtain a general version by replacing the norm $\|x\|_2$ with a discretization of the smoothing norm $S(f)$, of the form $\|L\,x\|_2$, where L is a discrete approximation of a derivative operator.

The Tikhonov regularization problem in *general form* is thus

$$\min_x \left\{ \|A\,x - b\|_2^2 + \lambda^2 \|L\,x\|_2^2 \right\} .$$

The matrix L is $p \times n$ with no restrictions on the dimension p.

About the L Matrix

If L is invertible, such that L^{-1} exists, then the solution can be written as

$$x_{L,\lambda} = L^{-1}\bar{x}_\lambda$$

where $\bar{x}_\lambda$ solves the standard-form Tikhonov problem

$$\min_{\bar{x}}\{\|(A\,L^{-1})\,\bar{x} - b\|_2^2 + \lambda^2\|\bar{x}\|_2^2\}.$$

The multiplication with L^{-1} in the back-transformation $x_\lambda = L^{-1}\bar{x}_\lambda$ represents integration, which yields additional smoothness in the Tikhonov solution, compared to $L = I$.

The same is also true for more general rectangular and non-invertible smoothing matrices L.

Similar to the standard-form problem obtained for $L = I$, the general-form Tikhonov solution $x_{L,\lambda}$ is the solution to a linear least-squares problem:

$$\min_x \left\| \begin{pmatrix} A \\ \lambda\,L \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2 .$$

The solution $x_{L,\lambda}$ is unique when the coefficient matrix has full rank, i.e., when the null spaces of A and L intersect trivially:

$$\mathcal{N}(A) \cap \mathcal{N}(L) = \emptyset.$$

Since multiplication with A represents a smoothing operation, it is unlikely that a smooth null vector of L (if L is rank deficient) is also a null vector of A.

Various choices of the matrix L are discussed in §8.2.

Two common choices of L are the rectangular matrices

$$L_1 = \begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{(n-1)\times n}$$

$$L_2 = \begin{pmatrix} 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \end{pmatrix} \in \mathbb{R}^{(n-2)\times n}$$

which represent the first and second derivative operators.

In Regularization Tools use `get_l(n,1)` and `get_l(n,2)` to compute these matrices.

Thus, the discrete smoothing norm $\|L\,x\|_2$, with L given by either I, L_1 or L_2, represents the continuous smoothing norms $S(f) = \|f\|_2$, $\|f'\|_2$, and $\|f''\|_2$, respectively.

To illustrate the improved performance of the general-form formulation, consider a simple ill-posed problem with missing data.

Let x be given as samples of a function, and let the right-hand side be given by a subset of these samples, e.g.,

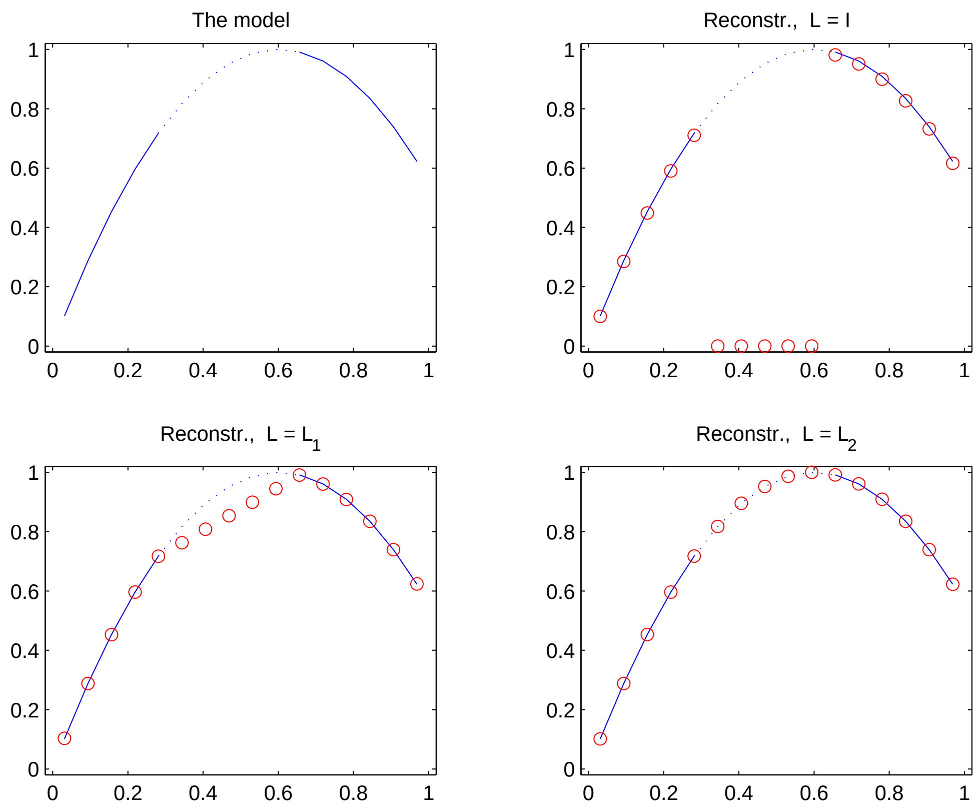
$$b = A\,x \ , \qquad A = \begin{pmatrix} I_{\rm left} & 0 & 0 \\ 0 & 0 & I_{\rm right} \end{pmatrix} \ ,$$

where $I_{\rm left}$ and $I_{\rm right}$ are two identity matrices.

The figure next page shows the solution x (consisting of samples of the sine function), as well as three reconstructions obtained with the three discrete smoothing norms $\|x\|_2$, $\|L_1\,x\|_2$ and $\|L_2\,x\|_2$.

For this problem, the solution is independent of λ.

The first choice is bad: the missing data are set to zero, in order to minimize the 2-norm of the solution. The choice $\|L_1\,x\|_2$ produces a linear interpolation in the interval with missing data, while the choice $\|L_2\,x\|_2$ produces a quadratic interpolation here.



Moving Away From the 2-Norm §8.6

Tikhonov is based on penalizing the 2-norm of the solution:

$$\min_x \left\{ \|A\,x - b\|_2^2 + \alpha^2 \|x\|_2^2 \right\}.$$

The same is true for TSVD, which can also be formulated as

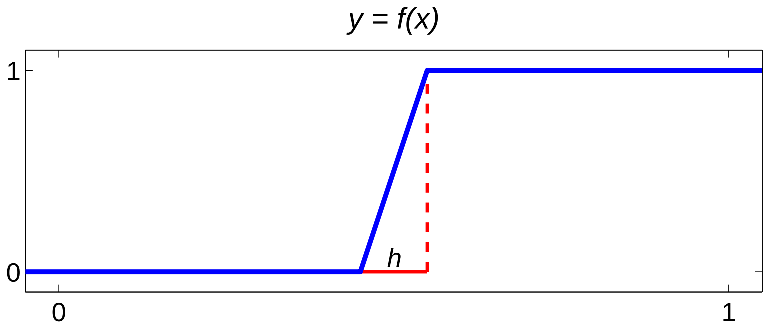
$$\min \|x\|_2 \quad \text{subject to} \quad \|A_k\,x - b\|_2 = \min, \qquad A_k = \sum_{i=1}^{k} u_i\,\sigma_i\,v_i^T.$$

It is the 2-norm penalization, together with the spectral properties of the SVD basis vectors, that cause a bad reconstruction of the edges = discontinuities.

It turns our that it is a better idea to involve the derivative of the solution **and** another norm!

So what is a good smoothing norm $S(f)$?

An Example Using a Continuous Function



Consider the piecewise linear function

$$f(t) = \begin{cases} 0, & 0 \le t < \frac{1}{2}(1-h) \\ \dfrac{t}{h} - \dfrac{1-h}{2h}, & \frac{1}{2}(1-h) \le t \le \frac{1}{2}(1+h) \\ 1, & \frac{1}{2}(1+h) < t \le 1 \end{cases}$$

which increases linearly from 0 to 1 in $[\frac{1}{2}(1-h), \frac{1}{2}(1+h)]$.

Norms of the First Derivative

It is easy to show that the 1- and 2-norms of $f'(t)$ satisfy

$$\begin{aligned} \|f'\|_1 &= \int_0^1 |f'(t)|\, dt = \int_0^h \frac{1}{h}\, dt = 1, \\ \|f'\|_2^2 &= \int_0^1 f'(t)^2\, dt = \int_0^h \frac{1}{h^2}\, dt = \frac{1}{h}. \end{aligned}$$

Note that $\|f'\|_1$ is independent of the slope of the middle part of $f(t)$, while $\|f'\|_2$ penalizes steep gradients (when h is small).

- The 2-norm of $f'(t)$ will not allow any steep gradients and therefore it produces a smooth solution .
- The 1-norm, on the other hand, allows some steep gradients – but not too many – and it is therefore able to produce a less smooth solution, and even a discontinuous solution.

Total Variation (TV) Regularization

The example motivates us to replace Tikhonov's 2-norm with the 1-norm of the first derivative, which is known as the *total variation*.

In the discrete setting:

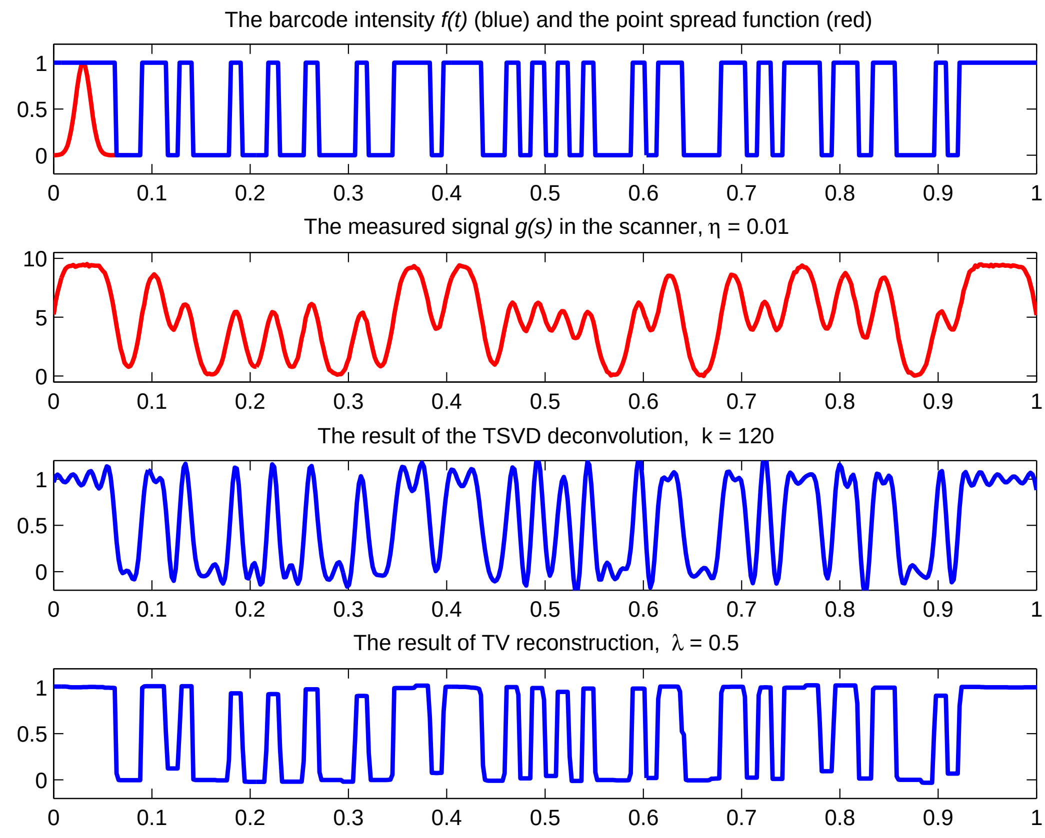
$$\min_x \left\{ \|A\,x - b\|_2^2 + \alpha^2 \|L\,x\|_1 \right\},$$

where

$$L = \begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{(n-1)\times n}$$

such that $\|L\,x\|_1$ approximates the total variation $\|f'\|_1$.

The figure on the next page shows a good TV reconstruction to the barcode problem.



TV in 2D

In two dimensions, given a function $f(\mathbf{t})$ with $\mathbf{t} = (t_1, t_2)$, we use the gradient magnitude define as

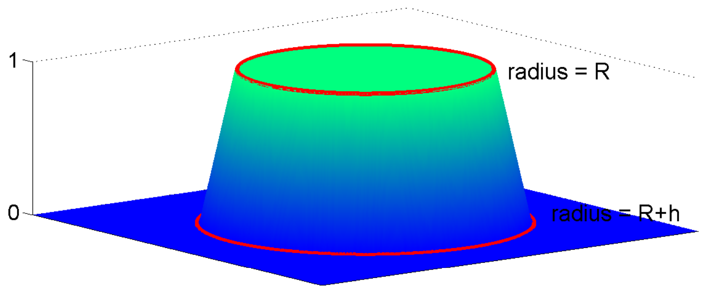
$$|\nabla f| = \left(\left(\frac{\partial f}{\partial t_1} \right)^2 + \left(\frac{\partial f}{\partial t_2} \right)^2 \right)^{\frac{1}{2}},$$

to obtain the 2D version of the total variation $\big\| |\nabla f| \big\|_1$. The relevant norms of $f(\mathbf{t})$ are now

$$\big\| |\nabla f| \big\|_1 = \int_0^1 \int_0^1 |\nabla f| \, dt_1 \, dt_2 = \int_0^1 \int_0^1 \left(\left(\frac{\partial f}{\partial t_1} \right)^2 + \left(\frac{\partial f}{\partial t_2} \right)^2 \right)^{\frac{1}{2}} dt_1 \, dt_2$$

$$\big\| |\nabla f| \big\|_2^2 = \int_0^1 \int_0^1 |\nabla f|^2 \, dt_1 \, dt_2 = \int_0^1 \int_0^1 \left(\frac{\partial f}{\partial t_1} \right)^2 + \left(\frac{\partial f}{\partial t_2} \right)^2 dt_1 \, dt_2.$$

An Example in 2D



To illustrate the difference between these two norms, consider a function $f(\mathbf{t})$ with the polar representation

$$f(r, \theta) = \begin{cases} 1, & 0 \le r < R \\ 1 + \frac{R}{h} - \frac{r}{h}, & R \le r \le R + h \\ 0, & R + h < r. \end{cases}$$

2D Example Continued

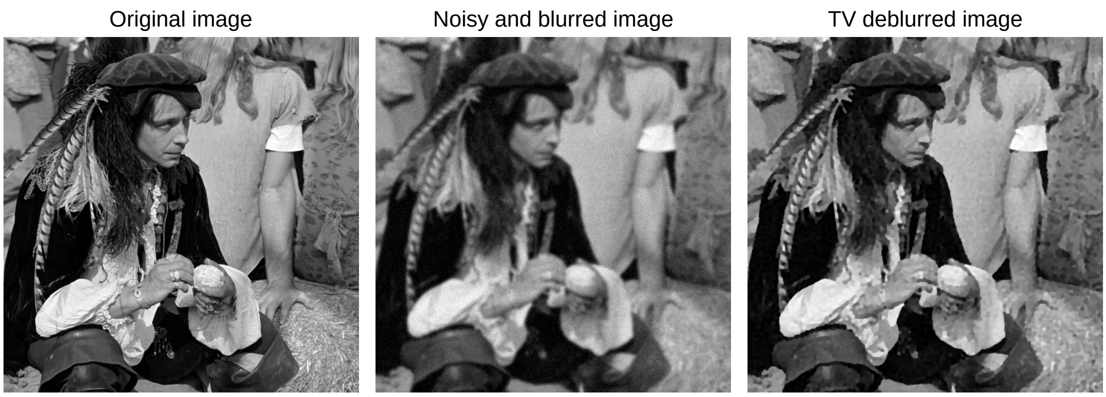
The function f is 1 inside the disk with radius $r = R$, zero outside the disk with radius $r = R + h$, and it has a linear radial slope between 0 and 1. In the area between these two disks the gradient magnitude is $|\nabla f| = 1/h$, and elsewhere it is zero.

$$\left\| |\nabla f| \right\|_1 = \int_0^{2\pi} \int_R^{R+h} \frac{1}{h}\, r\, dr\, d\theta = 2\pi R + \pi h$$

$$\left\| |\nabla f| \right\|_2^2 = \int_0^{2\pi} \int_R^{R+h} \frac{1}{h^2}\, r\, dr\, d\theta = \frac{2\pi R}{h} + \pi .$$

Similar to the one-dimensional example, we see that the total variation smoothing norm is almost independent of the size of the gradient, while the 2-norm penalizes steep gradients. In fact, as $h \to 0$ we see that $\left\| |\nabla f| \right\|_1$ converges to the circumference $2\pi R$.

Total Variation Image Deblurring Example



This example is from the paper:

J. Dahl, P. C. Hansen, S. H. Jensen, and T. L. Jensen, *Algorithms and software for total variation image reconstruction via first-order methods*, Numerical Algorithms, 53 (2010), pp. 67–92.