3. Tikhonov regularization

Consider a linear ill-posed problem of estimating $x \in \mathbb{R}^n$ from observed $b \in \mathbb{R}^m$,

$$b = Ax + e \quad , \quad A \in \mathbb{R}^{m \times n},$$

and $e \in \mathbb{R}^m$ is unknown noise vector. We assume that A has an SVD,

$$A = U D V^T,$$

where

$$D = \mathrm{diag}(d_1, d_2, \ldots, d_{\min\{n,m\}})$$

and

$$d_1 \gg d_{\min\{n,m\}}.$$

When solving the problem using TSVD, we observed that in the presence of noise, the norm of the estimate, $\| \hat{x}^{(k)} \|$ grows strongly as k grows. The simplest form of Tikhonov regularization seeks to control the norm of the functional.

Define the Tikhonov functional

$$T_\delta(x) = \| Ax - b \|^2 + \delta^2 \| x \|^2 \quad , \quad \delta \geq 0.$$

The parameter δ is the regularization parameter.

We define the Tikhonov regularized solution x_δ by setting

$$x_\delta = \arg\min_{x \in \mathbb{R}^n} T_\delta(x)$$

i.e., x_δ is the vector $x \in \mathbb{R}^n$ that minimizes $T_\delta(x)$.

Before discussing how to solve x_δ, let us analyze the solution:

(1) As $\delta \to 0+$, at the limit

$$T_\delta(x) \to T_0(x) = \|Ax - b\|^2$$

and x_δ becomes the least squares (LSQ) solution

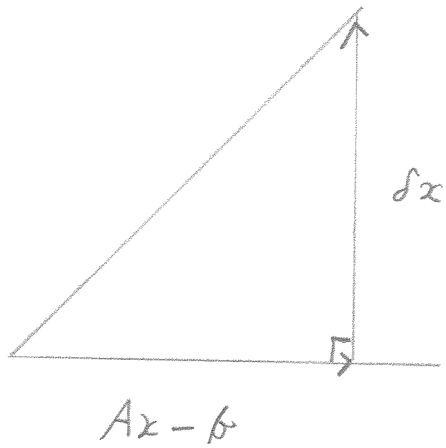
(2) As $\delta \to \infty$, the term $\|Ax - b\|^2$, and consequently the data b, has less and less significance. At the limit $x_\infty = 0$.

In general, we are between these two extreme cases.

To calculate x_δ, notice that

$$T_\delta(x) = \|Ax - b\|^2 + \delta^2\|x\|^2$$

$$= \left\| \begin{bmatrix} A \\ \delta I \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|^2$$



so x_δ is in fact, the LSQ solution of the augmented system

$$\begin{bmatrix} A \\ \delta I \end{bmatrix} x = \begin{bmatrix} b \\ 0 \end{bmatrix}$$

This, in fact, gives the practical means of solving x_δ. Using the pseudoinverse,

$$x_\delta = \begin{bmatrix} A \\ \delta I \end{bmatrix}^+ \begin{bmatrix} y \\ 0 \end{bmatrix}$$

or, in Matlab,

$$x_\delta = \begin{bmatrix} A \\ \delta I \end{bmatrix} \backslash \begin{bmatrix} y \\ 0 \end{bmatrix} \qquad (\text{"}\backslash\text{"} = \text{ml div})$$

To analyse the Tikhonov solution, let us consider the normal equations,

$$\begin{bmatrix} A \\ \delta I \end{bmatrix}^T \begin{bmatrix} A \\ \delta I \end{bmatrix} x = \begin{bmatrix} A \\ \delta I \end{bmatrix}^T \begin{bmatrix} b \\ 0 \end{bmatrix}$$

But

$$\begin{bmatrix} A \\ \delta I \end{bmatrix}^T = [A^T \ \ \delta I]$$

and

$$[A^T \ \delta I] \begin{bmatrix} A \\ \delta I \end{bmatrix} = A^TA + \delta^2 I \ , \quad [A^T \delta I] \begin{bmatrix} b \\ 0 \end{bmatrix} = A^T b$$

so we arrive at

$$(A^TA + \delta^2 I) x = A^T b,$$

which is called the <u>regularized normal equations</u>.

In order to analyze this, introduce the SVD of A. We have

$$A^TA = (UDV^T)^T (UDV^T) = V D^T \underbrace{U^T U}_{=I} D V^T$$

$$= V D^T D V^T.$$

Further, by the orthogonality of V,

$$\underline{I} = V V^T,$$

so the equations can be written as

$$(V D^T D V^T + \delta^2 V V^T) x = V D^T U^T b$$

or

$$V (D^T D + \delta^2 I) V^T x = V D^T U^T b$$

By defining $\tilde{x} = V^T x$, $\tilde{b} = U^T b$, and by multiplying with V^T from the left, we arrive at

$$(D^T D + \delta^2 I)\tilde{x} = D^T \tilde{b},$$

where all the matrices are diagonal. If $m \geq n$, $D \in \mathbb{R}^{m \times n}$ has more rows than columns,

$$D^T D = \begin{bmatrix} d_1^2 & & & \\ & d_2^2 & & \\ & & \ddots & \\ & & & d_n^2 \end{bmatrix},$$

while if $n > m$,

$$D^T D = \begin{bmatrix} d_1^2 & & & & & \\ & d_2^2 & & & & \\ & & \ddots & & & \\ & & & d_m^2 & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix}$$

In the latter case we interpret $d_j^2 = 0$, for $j > m$.

Let $k = \min\{n, m\}$. We have

$$(d_j^2+\delta^2)\tilde{x}_j = \begin{cases} d_j \tilde{b}_j, & 1\le j\le k \\ 0 \end{cases}$$

leading to the solution

$$x_\delta = \sum_{j=1}^{k} \frac{d_j}{d_j^2+\delta^2}(u_j^T b)\, v_j$$

Observe that as δ is small,

$$\frac{d_j}{d_j^2+\delta^2} = \frac{1}{d_j}\left(\frac{1}{1+(\delta/d_j)^2}\right) = \frac{1}{d_j}\left(1+\mathcal{O}\left(\frac{\delta^2}{d_j^2}\right)\right)$$

The Tikhonov regularized solution depends on δ, so the question arises how to choose δ. One possible answer, again, is to use the Morozov discrepancy principle:

Assume that we have an estimate of the noise level.

$$\|e\| \approx \varepsilon$$

The MDP gives a criterion for selection of δ:

Choose it from the condition

$$d(\delta) = \| A x_\delta - b \| = \varepsilon \text{ , i.e.}$$

$$\delta = d^{-1}(\varepsilon).$$

Is this always possible? Let's see:

(a) As $\delta \longrightarrow 0$, x_δ tends to the LSQ solution of

$$Ax = b,$$

and for the LSQ solution,

$$\| Ax - b \| = \| P b \| ,$$

where $P : \mathbb{R}^m \longrightarrow R(A)^\perp$ is the orthogonal projection

(b) As $\delta \longrightarrow \infty$, $x_\delta \longrightarrow 0$, so

$$\| A x_\delta - b \| \longrightarrow \| b \|$$

(c) The mapping $\delta \longmapsto \| A x_\delta - b \| = d(\delta)$ is increasing, which is seen as follows:

$$A x_\delta = \sum_{j=1}^{k} u_j d_j (v_j^T x_\delta)$$

$$= \sum_{j=1}^{k} \frac{d_j^2}{d_j^2+\delta^2} (u_j^T b) u_j$$

and

$$b = \sum_{j=1}^{m} (u_j^T b) u_j ,$$

so

$$A x_\delta - b = \sum_{j=1}^{k} \left(\frac{d_j^2}{d_j^2+\delta^2} - 1 \right) (u_j^T b) u_j \underbrace{- \sum_{j=k+1}^{m} (u_j^T b) u_j}_{\equiv Pb}$$

$$= -\sum_{j=1}^{k} \frac{\delta^2}{d_j^2+\delta^2} (u_j^T b) u_j - Pb$$

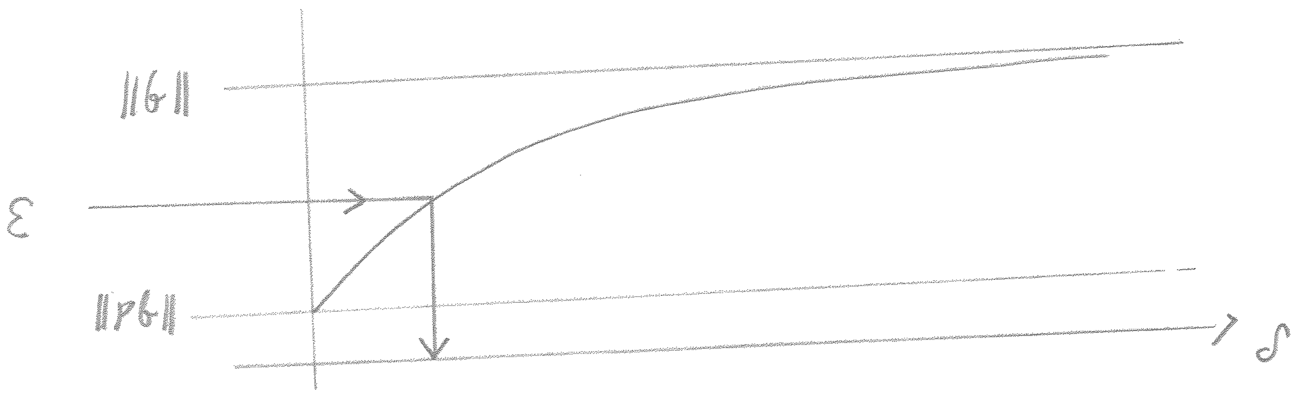
Therefore,

$$d(\delta)^2 = \|A x_\delta - b\|^2 = \sum_{j=1}^{k} \left(\frac{\delta^2}{d_j^2+\delta^2} \right)^2 (u_j^T b)^2 + \|Pb\|^2$$

Now it is not difficult to see that all the functions

$$\delta \longmapsto \frac{\delta^4}{(d_j^2+\delta^2)^2}$$

are increasing (Differentiate!), so $d(\delta)^2$ and therefore $d(\delta)$ is increasing.



We therefore see: if

$$\|Pb\| \leq \varepsilon \leq \|b\|,$$

there is a <u>unique</u> δ such that

$$d(\delta) = \varepsilon.$$

Computing this δ may be a tedious job, if A is of large dimensions.

Generalizations: In the basic form of Tikhonov regularization, the idea is to control simultaneously the discrepancy norm $\|Ax-b\|$ and the solution norm $\|x\|$. This norm $\|x\|$ is often called the penalty: The algorithm penalizes the solution for having a large norm, and the regularization parameter tells how strong the penalty is.

It is possible to replace the penalty by other types of penalties: Suppose, for instance that we seek a solution x for which a functional $J(x)$ should not be too large. We could then seek a solution

$$x_\delta = \operatorname{argmin}\left(\|Ax-b\|^2 + \delta^2 J(x)\right)$$

A typical choice of J is a smoothness penalty: For simplicity, let f be a signal, $f: [0,1] \to \mathbb{R}$, and $x \in \mathbb{R}^n$ a discrete sampling of f,

$$x_j = f\left(\frac{j}{n}\right), \quad 1 \le j \le n.$$

Assume that $f(0) = 0$.

The finite difference approximation for the derivative of f is

$$f'\left(\frac{j}{n}\right) \cong n\left(f\left(\frac{j}{n}\right) - f\left(\frac{j-1}{n}\right)\right)$$

$$= n\,(x_j - x_{j-1}), \quad 1 \le j \le n, \quad x_0 = 0,$$

or in matrix form,

$$f' \cong \underbrace{n \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}}_{L} x = Lx$$

Suppose that we seek to estimate x knowing that the underlying function f is not oscillating rapidly, i.e., its derivative f' must be small over the interval. We may now penalize x for oscillations, and seek a solution x by

$$x_\delta = \arg\min \left(\|Ax - b\|^2 + \delta^2 \|Lx\|^2 \right)$$

The penalty $\|Lx\|^2$ is called the first order smoothness penalty.