MTH 410/510 Inverse Problems & DA — week 4

Tikhonov Regularization (Andrey Tikhonov, 1963)

→ one of the most popular / successful regularization methods for ill-posed problems

→ related with the probabilistic formulation of an inverse problem (maximum likelihood, Bayesian estimation, minimum variance unbiased estimator)

Given $A \in \mathbb{R}^{m \times n}$, $m \geq n$ and noisy data vector $b \in \mathbb{R}^m$, provide an approximate solution to $Ax = b$

as the solution x_λ to the minimization problem

$$(*) \quad \min_{x \in \mathbb{R}^n} \left\{ \|Ax - b\|^2 + \lambda^2 \|x\|^2 \right\}$$

where $\lambda > 0$ is a <u>regularization parameter</u> controls the weight between the residual $\|Ax-b\|^2$ and Euclidean norm of the solution $\|x\|^2$.

Denoting $J_\lambda : \mathbb{R}^n \to \mathbb{R}$,

$$J_\lambda(x) = \|Ax - b\|^2 + \lambda^2 \|x\|^2$$

we have the following result of existence and uniqueness of the regularized solution

<u>Theorem</u> Let $\lambda > 0$. The minimization problem

$$\min_{x \in \mathbb{R}^n} J_\lambda(x)$$

has a unique solution x_λ. The solution x_λ satisfies (solves) the system

$$\boxed{(A^T A + \lambda^2 I)\, x_\lambda = A^T b} \qquad (**)$$

<u>Proof</u> Notice that $J_\lambda(x)$ may be written as

$$J_\lambda(x) = \left\| \begin{pmatrix} A \\ \lambda I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|^2$$

where $I \in \mathbb{R}^{n \times n}$ identity matrix

and therefore x_λ is minimizer to J_λ if and only if x_λ

is the solution to the normal equations

$$\begin{pmatrix} A \\ \lambda I \end{pmatrix}^T \begin{pmatrix} A \\ \lambda I \end{pmatrix} x = \begin{pmatrix} A \\ \lambda I \end{pmatrix}^T \begin{pmatrix} b \\ 0 \end{pmatrix}$$

thus if and only if x_λ solves $(**)$.

<u>Remark</u> The functional J_λ is strictly convex with Hessian matrix positive definite

$$\nabla^2 J_\lambda = A^T A + \lambda^2 I$$

The solution is

$$x_\lambda = (A^T A + \lambda^2 I)^{-1} A^T b$$

Notice that

if $\lambda \to 0$ then $x_\lambda \to x_{LS}$: $A^T A x_{LS} = A^T b$

if $\lambda \to \infty$ then $x_\lambda \to 0$ (over-smoothed)

The main challenge is to properly <u>select the regularitation parameter λ</u>.

Tikhonov solution as a filtered SVD

The Tikhonov solution x_λ may be expressed using the SVD of A

$$A = USV^T$$

as
$$x_\lambda = \sum_{i=1}^{n} f_i(\lambda) \frac{u_i^T \cdot b}{\sigma_i} v_i \qquad (***)$$

where
$$f_i(\lambda) = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \approx \begin{cases} 1 & \text{if } \sigma_i \gg \lambda \\ \frac{\sigma_i^2}{\lambda^2} & \text{if } \sigma_i \ll \lambda \end{cases}$$

→ filter factors are very small (close to 0) for singular values much smaller than λ.

$$\sigma_i \ll \lambda \Rightarrow f_i(\lambda) \approx 0 .$$

<u>Proof</u> (use "thin" version of SVD $S \in \mathbb{R}^{n \times n}$)

$$x_\lambda = (VS^2V^T + \lambda^2 VV^T)^{-1} VSU^Tb$$
$$= V(S^2 + \lambda^2 I)^{-1} SU^Tb \qquad \text{thus } (***)$$

Remark The Tikhonov solution x_λ may be also interpreted as the solution to the constrained optimization problem

$$\min_{x \in \mathbb{R}^n} \|Ax-b\|^2 \text{ subject to } \|x\|^2 \le \delta^2$$

where $|\delta| < \|(A^TA)^{-1}A^Tb\|$

Lagrangian function

$$\mathcal{L}(x,\gamma) = \|Ax-b\|^2 + \gamma(\|x\|^2 - \delta^2)$$

optimality condition:

$$\nabla_x \mathcal{L} = 0 \iff (A^TA + \gamma I)x = A^Tb$$

$$\gamma \geq 0$$

$$\gamma(\|x\|^2 - \delta^2) = 0$$

then $\gamma = \lambda^2 > 0$ (the constraint is active)

Monotonic behavior of the norms

For the TSVD method $f_i = 1,\ i = 1:k$, $f_i = 0,\ i = k+1:n$, such that

$$\|x_{k+1}\|^2 = \|x_k\|^2 + \left(\frac{u_{k+1}^T \cdot b}{\sigma_{k+1}}\right)^2 \geq \|x_k\|^2$$

and $\|Ax_{k+1} - b\|^2 = \|Ax_k - b\|^2 - (u_{k+1}^T \cdot b)^2 \leq \|Ax_k - b\|^2$

thus we have that

$\|x_k\|$ increases with k

$\|Ax_k - b\|$ decreases with k

For the Tikhonov method the filter factors are $f_i = \dfrac{\sigma_i^2}{\sigma_i^2 + \lambda^2}$

and the norms $\|x_\lambda\|_2^2$ and $\|Ax_\lambda - b\|_2^2$ are monotonic w.r.t. λ

Denoting
$$\xi(\lambda) = \|x_\lambda\|^2$$
$$\rho(\lambda) = \|Ax_\lambda - b\|^2$$

Choosing the regularization parameter

The L-curve criteria relies on the norm of the solution $\|x_\lambda\|$ and the norm of the residual $\|Ax_\lambda - b\|$ to select λ.

Notice that in general, if

$$x = \sum_{i=1}^{n} f_i \frac{u_i^T \cdot b}{\sigma_i} v_i$$

Then
$$\|x\|^2 = \sum_{i=1}^{n} \left(f_i \frac{u_i^T \cdot b}{\sigma_i} \right)^2$$

and
$$\|Ax - b\|^2 = \left\| \sum_{i=1}^{n} f_i (u_i^T \cdot b) u_i - b \right\|^2$$

$$= \left\| \sum_{i=1}^{n} f_i (u_i^T \cdot b) u_i - \sum_{i=1}^{m} (u_i^T \cdot b) u_i \right\|^2$$

$$= \left\| \sum_{i=1}^{n} (f_i - 1)(u_i^T \cdot b) u_i - \sum_{i=n+1}^{m} (u_i^T \cdot b) u_i \right\|^2$$

$$= \sum_{i=1}^{n} \left((f_i - 1) u_i^T \cdot b \right)^2 + \sum_{i=n+1}^{m} (u_i^T \cdot b)^2$$

we have that

$$\xi' = \frac{d\xi}{d\lambda} = -\frac{4}{\lambda} \sum_{i=1}^{n} (1 - f_i(\lambda))\, f_i^2(\lambda) \left(\frac{u_i^T \cdot b}{\sigma_i}\right)^2$$

$$\rho' = \frac{d\rho}{d\lambda} = -\lambda^2 \xi'$$

since $0 \leq f_i(\lambda) < 1$ for any $\lambda > 0$

we have that

$\frac{d\xi}{d\lambda} < 0$ thus $\|x_\lambda\|$ decreases with λ

$\frac{d\rho}{d\lambda} > 0$ thus $\|A\lambda_K - b\|$ increases with λ

in addition, since

$$\frac{d\xi}{d\rho} = -\frac{1}{\lambda^2}$$

we have that $\|x_\lambda\|^2$ is a decreasing function of $\|Ax_\lambda - b\|^2$