Observation sensitivity. Influence matrix

Consider the solution $x_{\lambda,L}$ to the general form of the Tikhonov regularization problem,

$$(A^T A + \lambda^2 L^T L)\, x_{\lambda,L} = A^T b + \lambda^2 L^T L x_0$$

We are interested to analyze how variations in data, δb and prior estimate δx_0 will impact the solution $x_{\lambda,L}$ or, in certain practical situations, a "quantity of interest" given by a function

$$\mathcal{Q}(x_{\lambda,L}).$$

Notice that $x_{\lambda,L}$ is a linear function of (b, x_0),

$$x_{\lambda,L} = (A^T A + \lambda^2 L^T L)^{-1} A^T b + \lambda^2 (A^T A + \lambda^2 L^T L)^{-1} L^T L x_0$$

The sensitivity of the solution $x_{\lambda,L}$ to the data vector b (observations) is given by the Jacobian matrix

$$\frac{\partial x_{\lambda,L}}{\partial b} = \left(A^T A + \lambda^2 L^T L\right)^{-1} A^T \in \mathbb{R}^{n \times m}$$

where $\left[\frac{\partial x_{\lambda,L}}{\partial b}\right]_{i,j} = \frac{\partial x_{\lambda,L}(i)}{\partial b(j)}$

indicates the impact of a variation in the data component $b(j) = b_j$ on the solution component $x_{\lambda,L}(i)$,

$$\delta x_{\lambda,L}(i) = \frac{\partial x_{\lambda,L}(i)}{\partial b(j)} \, \delta b(j)$$

Large sensitivity values identify those data components $b(j)$ where variations (such as noise or noise reduction) $\delta b(j)$ have a large impact on the solution component $x_{\lambda,L}(i)$.

The <u>influence matrix</u> of the observations is defined as

$$P = \frac{\partial [A x_{\lambda,L}]}{\partial b} = A(A^T A + \lambda^2 L^T L)^{-1} A^T \in \mathbb{R}^{m \times m}$$

Each entry $P_{i,j} = \dfrac{\partial [A x_{\lambda,L}]_i}{\partial b(j)}$

measures (describes) the influence (impact) of the data component $b(j)$ on the model prediction (fitted value) $[A x_{\lambda,L}](i)$

Notice that P is symmetric and positive semi-definite matrix.

if A has full row rank then P is positive definite.

in the particular case when $A \in \mathbb{R}^{m \times n}$ has full column rank and $\lambda = 0$ we have

$$P = A(A^T A)^{-1} A^T \quad \text{such that } P^2 = P$$

(P is idempotent).

The sensitivity of $x_{\lambda,L}$ to the prior estimate x_0 is

$$\frac{\partial x_{\lambda,L}}{\partial x_0} = \lambda^2 (A^T A + \lambda^2 L^T L)^{-1} L^T L \in \mathbb{R}^{n \times n}$$

Notice that the sensitivity to observations and the sensitivity to the prior satisfy the identity

$$\frac{\partial x_{\lambda,L}}{\partial b} A + \frac{\partial x_{\lambda,L}}{\partial x_0} = I_{n \times n}$$

The sensitivity of the model prediction $A x_{\lambda,L}$ to x_0 is expressed as

$$\frac{\partial [A x_{\lambda,L}]}{\partial x_0} = \lambda^2 A (A^T A + \lambda^2 L^T L)^{-1} L^T L \in \mathbb{R}^{m \times n}$$

such that we have the identity

$$\frac{\partial [A x_{\lambda,L}]}{\partial b} A + \frac{\partial [A x_{\lambda,L}]}{\partial x_0} = A$$

Often in practice we are interested to evaluate the sensitivity of a specified quantity of interest (QoI) $g(x_{\lambda,L})$ with respect to data and prior estimate, where $g : \mathbb{R}^n \to \mathbb{R}$.

For example, $g(x) = \|x\|^2$

or a linear aspect

$$g(x) = (w, x) = \sum_{i=1}^{n} w_i x_i.$$

We view $x_{\lambda,L}$ as a function of b and x_0, $x_{\lambda,L}(b, x_0)$

Accordingly, we view the QoI as

$$\tilde{g} : \mathbb{R}^{m+n} \to \mathbb{R}$$

$$\tilde{g}(b, x_0) = g(x_{\lambda,L}(b, x_0))$$

The sensitivity to the prior estimate x_0 is the gradient vector

$$\nabla_{x_0} \tilde{g} = \begin{bmatrix} \frac{\partial \tilde{g}}{\partial x_0(1)} \\ \vdots \\ \frac{\partial \tilde{g}}{\partial x_0(n)} \end{bmatrix} \in \mathbb{R}^n$$ expressed as

$$\nabla_{x_0} \tilde{g}(b, x_0) = \nabla_{x_0} x_{\lambda,L} \nabla_x g(x_{\lambda,L})$$

where $\nabla_{x_0} x_{\lambda,L} = \left(\frac{\partial x_{\lambda,L}}{\partial x_0} \right)^T$

such that

$$\nabla_{x_0} \tilde{g}(b, x_0) = \lambda^2 L^T L (A^T A + \lambda^2 L^T L)^{-1} \nabla_x g(x_{\lambda,L})$$

in particular, if $g(x) = \|x\|^2$ then

$$\nabla_{x_0} \tilde{g}(b, x_0) = 2\lambda^2 L^T L (A^T A + \lambda^2 L^T L)^{-1} x_{\lambda,L}$$

if $g(x) = (w, x) = w^T \cdot x$ then

$$\nabla_{x_0} \tilde{g}(b, x_0) = \lambda^2 L^T L (A^T A + \lambda^2 L^T L)^{-1} w$$

The sensitivity to observations of the QoI

is the gradient vector

$$\nabla_b \tilde{g} = \begin{bmatrix} \frac{\partial \tilde{g}}{\partial b_1} \\ \vdots \\ \frac{\partial \tilde{g}}{\partial b_m} \end{bmatrix} \in \mathbb{R}^m$$

and may be expressed using chain rule differentiation

as

$$\nabla_b \tilde{g}(b, x_0) = [\nabla_b x_{\lambda,L}] \cdot [\nabla_x g(x_{\lambda,L})]$$

$$\nabla_b \tilde{g}(b, x_0) = A(A^T A + \lambda^2 L^T L)^{-1} \nabla_x g(x_{\lambda,L})$$

in particular, if $g(x_{\lambda,L}) = \|x_{\lambda,L}\|^2$

then $\nabla_x g(x_{\lambda,L}) = 2x_{\lambda,L}$

and the observation sensitivity is

$$\nabla_b \|x_{\lambda,L}\|^2 = 2A(A^T A + \lambda^2 L^T L)^{-1} x_{\lambda,L}$$

if the QoI is $g(x_{\lambda,L}) = (w, x_{\lambda,L}) = w^T \cdot x_{\lambda,L}$

then $\nabla_x g(x_{\lambda,L}) = w$

and the observation sensitivity is

$$\nabla_b [w^T \cdot x_{\lambda,L}] = A(A^T A + \lambda^2 L^T L)^{-1} w$$