

Nonlinear inverse problems: least-squares estimation

Recall the context of linear inverse problems

$x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $m \geq n$, $A \in \mathbb{R}^{m \times n}$

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

→ minimization of a quadratic cost functional e.g.,

$$f(x) = \frac{1}{2} \sum_{i=1}^{m} (a_i^T \cdot x - b_i)^2 = \frac{1}{2} \sum_{i=1}^{m} r_i^2(x)$$

where $r_i : \mathbb{R}^n \to \mathbb{R}$, $r_i(x) = a_i^T \cdot x - b_i$

in general, we consider a problem where the observational operator is nonlinear,

$$y = h(x_t) + \varepsilon_o$$

where x_t is the true state, $x_t \in \mathbb{R}^n$

$h : \mathbb{R}^n \to \mathbb{R}^m$ is the observation operator

$y \in \mathbb{R}^m$ data vector, $\varepsilon_o \in \mathbb{R}^m$ observation error

The <u>nonlinear least-squares method</u> is

$$\boxed{\min_{x \in \mathbb{R}^n} f(x) \quad \text{where} \quad f(x) = \frac{1}{2} \sum_{i=1}^{m} r_i^2(x)}$$

$r_i(x) = h_i(x) - y_i \,, \qquad r_i : \mathbb{R}^n \to \mathbb{R}$

Denoting $r : \mathbb{R}^n \to \mathbb{R}^m \,, \quad r(x) = \begin{bmatrix} r_1(x) \\ \vdots \\ r_m(x) \end{bmatrix}$

we have $\quad f(x) = \frac{1}{2} \| r(x) \|^2 \longrightarrow$ Euclidean norm

The functions $r_i(x)$ are called <u>residuals</u>.

The <u>Jacobian matrix</u> of the residual function $r(x)$ is

$$J(x) = \frac{\partial r}{\partial x} \in \mathbb{R}^{m \times n} \,, \quad J_{ij}(x) = \frac{\partial r_i}{\partial x_j}(x)$$

$$i = 1:m \,, \quad j = 1:n$$

The gradient matrix of $r(x)$ is

$$\nabla r(x) = [\nabla r_1(x) \quad \dots \quad \nabla r_m(x)] = J^T(x) \in \mathbb{R}^{n \times m}$$

We assume that $r(x)$ is a smooth function (twice continuously differentiable)

The solution to the nonlinear optimization problem satisfies the first-order necessary conditions:

$$\nabla f(x^*) = 0 \qquad (*)$$

An approximate solution is obtained through an iterative process:

$x^{(0)}$ - initial guess

$$x^{(k+1)} = x^{(k)} + \alpha_k p_k$$

where $p_k \in \mathbb{R}^n$ is a specified direction (line search)

and α_k is a step length along p_k.

in particular, Newton's method applied to $(*)$ is

$$x^{(k+1)} = x^{(k)} - \left[\nabla^2 f(x^{(k)})\right]^{-1} \nabla f(x^{(k)})$$

or $x^{(k+1)} = x^{(k)} - \left[\nabla^2 f_k\right]^{-1} \nabla f_k$

where $\nabla f_k = \nabla f(x^{(k)})$ is the gradient of f evaluated at $x^{(k)}$ and

$\nabla^2 f_k = \nabla^2 f(x^{(k)}) \in \mathbb{R}^{n\times n}$ is the Hessian matrix of f at $x^{(k)}$.

Notice that the gradient $\nabla f(x) \in \mathbb{R}^n$ is

$$\nabla f(x) = \sum_{i=1}^{m} r_i(x) \nabla r_i(x) = [\nabla r(x)] r(x) = J^T(x) r(x)$$

$$\boxed{\nabla f(x) = J^T(x) r(x)}$$

and the Hessian matrix is expressed as

$$[\nabla^2 f]_{l,k} = \sum_{i=1}^{m} \frac{\partial}{\partial x_k}\left[r_i(x) \frac{\partial r_i(x)}{\partial x_l} \right] =$$

$$= \underbrace{\sum_{i=1}^{m} \frac{\partial r_i}{\partial x_k} \frac{\partial r_i}{\partial x_l}}_{(J^T J)_{l,k}} + \sum_{i=1}^{m} r_i \underbrace{\frac{\partial^2 r_i}{\partial x_k \partial x_l}}_{\nabla^2 r_i}$$

Thus
$$\boxed{\nabla^2 f(x) = J^T(x) J(x) + \sum_{i=1}^{m} r_i(x) \nabla^2 r_i(x)}$$

<u>Gauss-Newton method</u> (Matlab: lsqnonlin)

Use an approximation of the Hessian matrix

$$\boxed{\nabla^2 f_k \approx J_k^T J_k}$$ and line search iteration

$$x^{(k+1)} = x^{(k)} + \alpha_k p_k$$

where the direction p_k is obtained as

$$(J_k^T J_k)\, p_k = -\nabla f_k = -\underbrace{J_k^T r_k}_{J^T(x_k^{(k)})\, r(x^{(k)})}$$

<u>Property</u>: if $\nabla f_k \neq 0_n$ and the matrix $J_k = J(x^{(k)})$ has full rank then p_k is a <u>descent direction</u>

$$p_k^T \cdot \nabla f_k = -p_k^T (J_k^T J_k)\, p_k < 0$$

since $J_k^T J_k$ is positive definite.

Notice that p_k solves the <u>linear least-squares problem</u>

$$\min_{p \in \mathbb{R}^n} \frac{1}{2} \| J_k p + r_k \|^2 \Rightarrow p_k = -(J_k^T J_k)^{-1} J_k^T r_k$$

such that the descent direction is found by linear approximation:

$$r(x^{(k)} + p) \approx r_k + J_k p$$

linearization of r at $x^{(k)}$.

<u>A line search</u> is then performed to find the step length α_k:

$$\min_{\alpha > 0} \phi(\alpha) = f(x^{(k)} + \alpha p_k)$$

Notice $\phi'(\alpha) = \nabla f(x^{(k)} + \alpha p_k)^T \cdot p_k$

Nonlinear least-squares: general form

Assume: $y = h(x_t) + \varepsilon_o$, observational data

where ε_o is random vector of observational errors,

$$E[\varepsilon_o] = 0_m, \quad cov(\varepsilon_o) = R \in \mathbb{R}^{m \times m}$$

$$x_b = x_t + \varepsilon_b, \quad \text{prior estimate}$$

$$E[\varepsilon_b] = 0_n, \quad cov(\varepsilon_b) = B \in \mathbb{R}^{n \times n}$$

Least-squares estimation:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \| h(x) - y \|^2_{R^{-1}} + \frac{1}{2} \| x - x_b \|^2_{B^{-1}}$$

Linearization of $h(x)$ at x_b:

$$h(x) \approx h(x_b) + H_b (x - x_b)$$

where $H_b = \frac{\partial h}{\partial x}(x_b) \in \mathbb{R}^{m \times n}$

<u>Quadratic problem</u>: Let $\delta x = x - x_b$

$$\min_{\delta x \in \mathbb{R}^n} \frac{1}{2} \| H_b \, \delta x - (y - h(x_b)) \|^2_{R^{-1}} + \frac{1}{2} \| \delta x \|^2_{B^{-1}}$$

Solution (see equation of BLUE analysis)

$$\delta x_a = K [y - h(x_b)] \quad \text{(analysis increment)}$$

or

$$\boxed{x_a = x_b + K [y - h(x_b)]}$$

where

$$K = B H_b^T [H_b B H_b^T + R]^{-1} \in \mathbb{R}^{n \times m}$$

is the gain matrix

An iterative process may be formulated by setting $x_b = x_a$ (update the prior to current analysis estimate) and repeat the linearization.