MTH 410/510 Inverse Problems & DA

Weighted least-squares estimation. & Best linear unbiased estimator (BLUE)

Preliminary concepts:

Definition A matrix $Q \in \mathbb{R}^{n\times n}$ is said to be positive definite if

1) Q is symmetric, $Q = Q^T$

and 2) for any $x \in \mathbb{R}^n \setminus \{0_n\}$, $x^T Q x > 0$.
$(x \neq 0_n)$

Notation Given a positive definite matrix we denote

$$\|x\|_Q = \sqrt{x^T Q x}, \quad \forall x \in \mathbb{R}^n$$

Then $\|\cdot\|_Q$ is a norm on $\mathbb{R}^n$.

We use the notation $Q > 0$ to specify a positive definite matrix Q

Weighted least-squares minimization

Let $y \in \mathbb{R}^m$ denote the data vector

$H \in \mathbb{R}^{m \times n}$ the observation operator

$x_b \in \mathbb{R}^n$ a prior estimate to $x_t \in \mathbb{R}^n$

Let $W \in \mathbb{R}^{n \times n}$, $W > 0$ and

$Q \in \mathbb{R}^{m \times m}$, $Q > 0$

weight matrices.

Consider the minimization problem

$$\min_{x \in \mathbb{R}^n} J(x) \qquad (*) \qquad (\text{WLS})$$

where $J(x) = \frac{1}{2} \| Hx - y \|_Q^2 + \frac{1}{2} \| x - x_b \|_W^2$

<u>Property</u> There is a unique solution to the minimization problem $(*)$. The solution denoted x_a^* is obtained by solving the linear system

$$(H^T Q H + W) x_a^* = H^T Q y + W x_b$$

<u>Property</u> The solution x_a^* to the weighted least-squares (WLS) problem (*) is expressed as

$$x_a^* = x_b + K(y - Hx_b) = Ky + (I - KH)x_b$$

where $K \in \mathbb{R}^{n \times m}$ is defined as

$$K = (H^T Q H + W)^{-1} H^T Q \quad \text{gain matrix}$$

<u>Remark</u> : Let $\varepsilon_o = y - Hx_t$ denote the observation error and assume that

$E[\varepsilon_o] = 0_{\mathbb{R}^m}$ (unbiased).

Let $\varepsilon_b = x_b - x_t$ denote the error in the prior estimate and assume that

$E[\varepsilon_b] = 0_{\mathbb{R}^n}$ (unbiased).

Then for any specification of the weight matrices $Q > 0$ and $W > 0$, the (WLS) solution x_a^* is an unbiased estimate to x_t,

$$E[x^*] = x_t$$

Denoting $\varepsilon_a^* = x_a^* - x_t$ we have :

$$\varepsilon_a^* = x_b + K(y - Hx_b) - x_t$$

$$= K(y - Hx_t) + (I - KH)(x_b - x_t)$$

$$= K\varepsilon_o + (I - KH)\varepsilon_b$$

Therefore $E[\varepsilon_a^*] = K E[\varepsilon^o] + [I - KH] E[\varepsilon_b]$

$$= 0 .$$

Furthermore, assuming that the observation errors and the errors in the prior estimate are uncorrelated, i.e. $E[\varepsilon_o \varepsilon_b^T] = 0_{m \times n}$

we may express the covariance $\mathrm{cov}(\varepsilon_a^*)$ as follows :

$$\mathrm{cov}(\varepsilon_a^*) = KRK^T + (I - KH)B(I-KH)^T$$

where $R = E[\varepsilon_o \varepsilon_o^T]$

is the covariance matrix of the observation errors and

$$B = E[\varepsilon_b \varepsilon_b^T]$$

is the covariance matrix of the errors in the prior estimate.

The best linear unbiased estimation problem is formulated as:

$$\min_{K \in \mathbb{R}^{m \times m}} \mathrm{Tr}\, E[\varepsilon_a^* \varepsilon_a^{*T}] \stackrel{\text{def}}{=} \sum_{i=1}^{n} \mathrm{var}(\varepsilon_{a,i}^*)$$

that is, we search for the optimal gain matrix K to minimize the total error variance given by the trace of the error covariance matrix.

A simple scalar example:

$$y = x_t + \varepsilon_o \quad , \quad E(\varepsilon_o) = 0$$

$$x_b = x_t + \varepsilon_b \quad , \quad E(\varepsilon_b) = 0$$

Assume that ε_o & ε_b are uncorrelated

$$E[\varepsilon_o \varepsilon_b] = 0 .$$

Then $\quad x_a = K y + (1-K)\, x_b$

$$\varepsilon_a = x_a - x_t = K \varepsilon_o + (1-K)\varepsilon_b$$

$$\sigma_a^2 = E[\varepsilon_a^2] = K^2 \sigma_o^2 + (1-K)^2 \sigma_b^2$$

Optimal analysis: Solve for K

$$\min_K \sigma_a^2 (K)$$

Solution $\quad K\sigma_o^2 + (1-K)\,\sigma_b^2 = 0$

$$K = \frac{\sigma_b^2}{\sigma_o^2 + \sigma_b^2}$$

Thus $\quad x_a^* = \dfrac{\sigma_b^2}{\sigma_o^2 + \sigma_b^2}\, y + \dfrac{\sigma_o^2}{\sigma_o^2 + \sigma_b^2}\, x_b$

and $\sigma_a^2 = \dfrac{\sigma_o^2 \sigma_b^2}{\sigma_o^2 + \sigma_b^2} \leq \min\{\sigma_o^2, \sigma_b^2\}$

Notice that $\dfrac{1}{\sigma_a^2} = \dfrac{1}{\sigma_o^2} + \dfrac{1}{\sigma_b^2}$

Thus the optimal (BLUE) analysis error variance is smaller than each of σ_o^2, σ_b^2.

in addition x_a^* is the solution to the optimization problem:

$$\min_x \left\{ \frac{1}{2} \frac{(x - x_b)^2}{\sigma_b^2} + \frac{1}{2} \frac{(x - y)^2}{\sigma_o^2} \right\}$$

Since $J'(x) = \dfrac{x - x_b}{\sigma_b^2} + \dfrac{x - y}{\sigma_o^2} = 0$

implies

$$x = \frac{\sigma_o^2}{\sigma_o^2 + \sigma_b^2} x_b + \frac{\sigma_b^2}{\sigma_o^2 + \sigma_b^2} y$$

thus $x = x_a^*$

<u>Theorem</u> The optimal least-squares estimator, or BLUE analysis, is expressed as

$$x_a = x_b + K(y - Hx_b)$$

where

$$K = BH^T(HBH^T + R)^{-1}$$

is the optimal gain (weight) matrix of the analysis.

The error covariance of the BLUE analysis is

$$A = (I - KH)B$$

The BLUE analysis is the solution to the optimization problem

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

$$J(x) = \frac{1}{2}\|x - x_b\|^2_{B^{-1}} + \frac{1}{2}\|y - Hx\|^2_{R^{-1}}$$

(thus $Q = R^{-1}$, $W = B^{-1}$ in (WLS)).

Proof Recall that for any specification of the gain matrix K, the analysis error covariance is express as

$$A = KRK^T + (I - KH)B(I - KH)^T$$

thus BLUE is obtained by solving the matrix optimization problem

$$\min_{K \in \mathbb{R}^{n \times m}} Tr[A]$$

Using calculus of variations, we express

$$\delta Tr[A] = \delta Tr[KRK^T] + \delta Tr[KHBH^T K^T]$$
$$- \delta Tr[KHB] - \delta Tr[BH^T K^T]$$

$$\delta Tr[A] = 2\, Tr[KR(\delta K)^T] + 2\, Tr[KHBH^T(\delta K)^T]$$
$$- 2\, Tr[BH^T(\delta K)^T]$$
$$= 2\, Tr\{[K(R + HBH^T) - BH^T](\delta K)^T\}$$
$$= 2\langle K(R + HBH^T) - BH^T, \delta K\rangle_{n \times m}$$

Therefore, the first order optimality condition is

$$K(R + HBH^T) - BH^T = 0$$

thus the optimal gain matrix is

$$K = BH^T(HBH^T + R)^{-1}$$

To obtain the expression for the optimal (BLUE) analysis error covariance, notice that

$$K(HBH^T + R) = BH^T$$

implies

$$KRK^T + KHBH^TK^T = BH^TK^T$$

thus

$$A = KRK^T + (I - KH)B(I - KH)^T$$

simplifies to

$$A = (I - KH)B$$

Remark: Notice the following matrix identity:

$$BH^T(HBH^T+R)^{-1} = (B^{-1}+H^TR^{-1}H)^{-1}H^TR^{-1}$$

Then we may express the optimal gain

$$K = (B^{-1}+H^TR^{-1}H)^{-1}H^TR^{-1}$$

and the BLUE analysis error covariance is

$$A = B - (B^{-1}+H^TR^{-1}H)^{-1}H^TR^{-1}HB$$

$$A = (B^{-1}+H^TR^{-1}H)^{-1}$$

Notice that the analysis error covariance for BLUE is the inverse of the Hessian matrix

$$A = [\nabla^2 J]^{-1}$$

and

$$A^{-1} = B^{-1}+H^TR^{-1}H = \nabla^2 J.$$

Notice that the K operator (gain matrix) is determined by BH^TR^{-1} :

$$K = BH^T(HBH^T + R)^{-1} =$$

$$= BH^T\left[(HBH^TR^{-1} + I)R\right]^{-1}$$

$$= BH^TR^{-1}\left[H(BH^TR^{-1}) + I\right]^{-1}$$

Thus for any specification of the weight matrices $(\hat{B}, \hat{R})$ such that

$$\hat{B}H^T\hat{R}^{-1} = BH^TR^{-1}$$

we have $\hat{x}_a = x_a$

Conversely, we have

$\hat{K} = K$ if and only if

$$\hat{B}H^T\hat{R}^{-1} = BH^TR^{-1}$$

Additional properties of BLUE

The analysis error covariance satisfies :

i) $A - B \leq 0$ (negative semi-definite)

ii) $HAH^T - R < 0$ (negative definite)

such that $\sigma_{a,i}^2 \leq \sigma_{b,i}^2$, $i = 1:n$

$$\mathrm{Var}\left[(H\varepsilon_a)_i\right] < \sigma_{o,i}^2 , \quad i = 1:m$$

Proof : since $A = (I - KH)B$

we have

$$A - B = -KHB = -BH^T(HBH^T+R)^{-1}HB \leq 0$$

$$HAH^T - R = HBH^T - HKHBH^T - R =$$

$$= -R(HBH^T+R)^{-1}R < 0$$

Tikhonov regularization solution as BLUE

The solution $x_{\lambda,L}$ in the Tikhonov regularization is obtained by minimizing

$$J_{\lambda,L}(x) = \|Ax-b\|^2 + \lambda^2 \|L(x-x_0)\|^2$$

Assuming that the noise in data $\varepsilon_0 = b - b_t$ has the covariance

$$\mathrm{cov}(\varepsilon_0) = \sigma_0^2 I$$

(thus uncorrelated errors in data and same variance σ_0^2 for all components)

then setting $x_0 = x_b$ and $LL^T = B^{-1}$

the specification of the regularization parameter $\lambda^2 = \sigma_0^2$

provides $x_{\lambda,L} = x_a$ (BLUE analysis)