Section 3.2.4: Batch (Way A) and Sequential (Way B) Derivations

1 Problem Setup

We measure the same unknown scalar x repeatedly. For $i = 1, \dots, k$,

$$y_i = x + \nu_i, \qquad \nu_i \sim \mathcal{N}(0, 1), \tag{1}$$

with $\{\nu_i\}$ independent. After k measurements, we will derive the estimate $\hat{x}_k$ and its variance $\mathrm{Var}(\hat{x}_k)$ in two ways:

- **Way A (Batch):** stack all measurements and solve once.
- **Way B (Sequential):** update $(\hat{x}_k, P_k)$ recursively when a new measurement arrives.

2 Way A: Batch (All-at-Once) Derivation

2.1 Step 1: Stack measurements into a single linear system

Define

$$y \triangleq \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix} \in \mathbb{R}^{k\times 1}, \qquad \nu \triangleq \begin{bmatrix} \nu_1 \\ \vdots \\ \nu_k \end{bmatrix} \in \mathbb{R}^{k\times 1}, \qquad C \triangleq \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^{k\times 1}. \tag{2}$$

Then all k scalar equations can be written compactly as

$$y = Cx + \nu. \tag{3}$$

Since each $\nu_i \sim \mathcal{N}(0, 1)$ and they are independent,

$$\nu \sim \mathcal{N}(0, I_k). \tag{4}$$

2.2 Step 2: Write the likelihood and the least-squares objective

Given x, the measurement vector y is Gaussian:

$$p(y \mid x) = \mathcal{N}(Cx, I_k). \tag{5}$$

Up to an additive constant independent of x, the negative log-likelihood is

$$-\log p(y \mid x) = \tfrac{1}{2}(y - Cx)^\top I_k^{-1}(y - Cx) = \tfrac{1}{2}\|y - Cx\|^2. \tag{6}$$

Thus maximizing likelihood is equivalent to minimizing the least-squares cost

$$J(x) \triangleq \|y - Cx\|^2. \tag{7}$$

2.3 Step 3: Solve the least-squares problem (normal equations)

Expand $J(x)$:

$$\begin{aligned} J(x) &= (y - Cx)^\top (y - Cx) && (8) \\ &= y^\top y - 2x\, C^\top y + x^2\, C^\top C. && (9) \end{aligned}$$

Differentiate with respect to the scalar x and set to zero:

$$\frac{dJ}{dx} = -2C^\top y + 2x\, C^\top C = 0 \quad \Longrightarrow \quad \hat{x}_k = (C^\top C)^{-1} C^\top y. \tag{10}$$

Now plug in the special $C = \mathbf{1}_k$:

$$C^\top C = \mathbf{1}_k^\top \mathbf{1}_k = k, \qquad C^\top y = \sum_{i=1}^{k} y_i. \tag{11}$$

Hence

$$\boxed{\hat{x}_k = \frac{1}{k} \sum_{i=1}^{k} y_i.} \tag{12}$$

2.4 Step 4: Derive the variance of the batch estimate

Start from

$$\begin{aligned} \hat{x}_k &= (C^\top C)^{-1} C^\top y && (13) \\ &= (C^\top C)^{-1} C^\top (Cx + \nu) && (14) \\ &= x + (C^\top C)^{-1} C^\top \nu. && (15) \end{aligned}$$

Therefore the estimation error is

$$\tilde{x}_k \triangleq \hat{x}_k - x = (C^\top C)^{-1} C^\top \nu. \tag{16}$$

Compute its variance:

$$\begin{aligned} \mathrm{Var}(\hat{x}_k) &= \mathrm{Var}(\tilde{x}_k) && (17) \\ &= (C^\top C)^{-1} C^\top\, \mathrm{Cov}(\nu)\, C\, (C^\top C)^{-1}. && (18) \end{aligned}$$

Since $\mathrm{Cov}(\nu) = I_k$,

$$\mathrm{Var}(\hat{x}_k) = (C^\top C)^{-1} C^\top I_k C (C^\top C)^{-1} = (C^\top C)^{-1}. \tag{19}$$

And because $C^\top C = k$,

$$\boxed{\mathrm{Var}(\hat{x}_k) = \frac{1}{k}.} \tag{20}$$

3 Way B: Sequential (Recursive) Derivation

3.1 Step 0: Treat the first k measurements as a Gaussian prior

From Way A, after k unit-variance measurements,

$$\hat{x}_k = \frac{1}{k}\sum_{i=1}^{k} y_i, \qquad P_k \triangleq \mathrm{Var}(\hat{x}_k) = \frac{1}{k}. \tag{21}$$

Interpret this as a prior distribution for x conditioned on the first k measurements:

$$x \mid y_{1:k} \sim \mathcal{N}(\hat{x}_k, P_k). \tag{22}$$

Now obtain one more measurement

$$y_{k+1} = x + \nu_{k+1}, \qquad \nu_{k+1} \sim \mathcal{N}(0, \sigma^2), \tag{23}$$

independent of everything else.

3.2 Step 1: Posterior is proportional to prior times likelihood

Prior:

$$p(x \mid y_{1:k}) \propto \exp\left(-\frac{1}{2P_k}(x-\hat{x}_k)^2\right). \tag{24}$$

Likelihood:

$$p(y_{k+1} \mid x) \propto \exp\left(-\frac{1}{2\sigma^2}(y_{k+1}-x)^2\right). \tag{25}$$

Multiply:

$$p(x \mid y_{1:k+1}) \propto \exp\left(-\frac{1}{2P_k}(x-\hat{x}_k)^2 - \frac{1}{2\sigma^2}(y_{k+1}-x)^2\right). \tag{26}$$

This exponent is quadratic in x, so the posterior is Gaussian.

3.3 Step 2: Expand and collect terms in x

Expand:

$$(x-\hat{x}_k)^2 = x^2 - 2\hat{x}_k x + \hat{x}_k^2, \qquad (y_{k+1}-x)^2 = x^2 - 2y_{k+1}x + y_{k+1}^2. \tag{27}$$

Ignoring constants independent of x, the exponent becomes

$$-\frac{1}{2}\left[\left(\frac{1}{P_k}+\frac{1}{\sigma^2}\right)x^2 - 2\left(\frac{\hat{x}_k}{P_k}+\frac{y_{k+1}}{\sigma^2}\right)x\right] + \text{const.} \tag{28}$$

Match this with the canonical Gaussian form

$$-\frac{1}{2}\left[\frac{1}{P_{k+1}}x^2 - 2\frac{\hat{x}_{k+1}}{P_{k+1}}x\right] + \text{const.} \tag{29}$$

3.4 Step 3: Read off the variance update (information form)

Matching x^2 coefficients yields

$$\boxed{\frac{1}{P_{k+1}} = \frac{1}{P_k} + \frac{1}{\sigma^2}.} \tag{30}$$

Since $P_k = 1/k$, we have $1/P_k = k$, so

$$\frac{1}{P_{k+1}} = k + \frac{1}{\sigma^2} \quad\Longrightarrow\quad \boxed{P_{k+1} = \frac{1}{k+\frac{1}{\sigma^2}} = \frac{\sigma^2}{\sigma^2 k + 1}.} \tag{31}$$

3.5 Step 4: Read off the mean update

Matching linear coefficients yields

$$\boxed{\frac{\hat{x}_{k+1}}{P_{k+1}} = \frac{\hat{x}_k}{P_k} + \frac{y_{k+1}}{\sigma^2}.} \tag{32}$$

Thus

$$\hat{x}_{k+1} = P_{k+1}\left(\frac{\hat{x}_k}{P_k} + \frac{y_{k+1}}{\sigma^2}\right). \tag{33}$$

Using $1/P_k = k$ and $P_{k+1} = \frac{1}{k+1/\sigma^2}$,

$$\hat{x}_{k+1} = \frac{1}{k + \frac{1}{\sigma^2}}\left(k\hat{x}_k + \frac{1}{\sigma^2}y_{k+1}\right). \tag{34}$$

This shows $\hat{x}_{k+1}$ is a weighted average of $\hat{x}_k$ and y_{k+1}, weighted by their precisions.

3.6 Step 5: Convert to innovation form

Subtract $\hat{x}_k$:

$$\hat{x}_{k+1} - \hat{x}_k = \frac{k\hat{x}_k + \frac{1}{\sigma^2}y_{k+1} - \hat{x}_k\left(k + \frac{1}{\sigma^2}\right)}{k + \frac{1}{\sigma^2}} \tag{35}$$

$$= \frac{\frac{1}{\sigma^2}(y_{k+1} - \hat{x}_k)}{k + \frac{1}{\sigma^2}} \tag{36}$$

$$= \frac{y_{k+1} - \hat{x}_k}{\sigma^2 k + 1}, \tag{37}$$

where the last step multiplies numerator and denominator by σ^2. Therefore

$$\boxed{\hat{x}_{k+1} = \hat{x}_k + \frac{y_{k+1} - \hat{x}_k}{\sigma^2 k + 1}.} \tag{38}$$