Dynamic inverse problems

Inverse problems with *static target*:

Assume an additive noise model,

$$Y = f(X) + E, \quad E \sim \mathcal{N}(0, \sigma^2 I).$$

Repeated independent observations: measure Y N times, assuming that the target remains the same during the process.

$$D = \{y_1, y_2, \ldots, y_N\} = \text{data},$$

$$y_j = f(x) + e_j.$$

Likelihood:

$$\pi(y_1, y_2, \ldots, y_N \mid x) \propto \prod_{j=1}^{N} \pi(y_j \mid x).$$

In particular, with Gaussian noise,

$$\pi(y_j \mid x) \propto \exp\left(-\frac{1}{2\sigma^2}\|f(x) - y_j\|^2\right).$$

Then,

$$\pi(y_1, y_2, \ldots, y_N \mid x) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^{N}\|f(x) - y_j\|^2\right).$$

$$
\begin{aligned}
\sum_{j=1}^{N} \|f(x) - y_j\|^2 &= N\|f(x)\|^2 - 2f(x)^{\mathrm{T}} \underbrace{\sum_{j=1}^{N} y_j}_{N\overline{y}} + \sum_{j=1}^{N} y_j^{\mathrm{T}} y_j \\
&= N\left(\|f(x)\|^2 - 2f(t)^{\mathrm{T}}\overline{y} + \overline{y}^{\mathrm{T}}\overline{y}\right) + N \underbrace{\left(\frac{1}{N}\sum_{j=1}^{N} y_j^{\mathrm{T}} y_j - \overline{y}^{\mathrm{T}}\overline{y}\right)}_{=C} \\
&= N\|f(x) - \overline{y}\|^2 + \underbrace{NC}_{=\text{constant}} .
\end{aligned}
$$

Therefore

$$\begin{aligned}\pi(y_1, y_2, \ldots, y_N \mid x) \quad &\propto \quad \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^{N}\|f(x) - y_j\|^2\right) \\ &\propto \quad \exp\left(-\frac{1}{2(\sigma^2/N)}\|f(x) - \overline{y}\|^2\right).\end{aligned}$$

Hence: Repeating the measurement *independently* N times is equivalent to replacing the model with

$$\overline{Y} = f(X) + E, \quad E \sim \mathcal{N}\left(0, \frac{\sigma^2}{N}\right).$$

Variance reduction of the noise!

It is quintessential that the target does not change during the measurement process.

Examples where the condition may *not* be valid:

- MEG, EEG: repeated measurement of evoked potential (tiring, learning)
- Process tomography: integration of the signal
- Target tracking
- Monitoring of a chemical system
- Weather forecasting

More general observation model:

$$Y_j = f(X_j) + E_j, \quad j = 1, 2, \ldots$$

Clearly, the observations cannot be integrated unless we have a dynamic prior model.

One of the simplest dynamic prior models is a 1–Markov evolution model,

$$X_{j+1} = g_j(X_j) + V_{j+1},$$

where $g_j : \mathbb{R}^n \to \mathbb{R}^n$ is presumably known, and V_{j+1} is an *innovation process.*

- Static measurment: $g_j(x) = x$, $V_{j+1} = 0$.
- Random walk model: often used in lack of anything more sophisticated:

$$X_{j+1} = X_j + V_{j+1}, \quad V_{j+1} \sim \mathcal{N}(0, \gamma^2 I).$$

  Despite of its simplicity, its powerfulness should not be underestimated.

- First order differential equation: assume that the unknown is a time dependent vector $x(t) \in \mathbb{R}^n$, satisfying *ideally* the differential equation

$$x'(t) = f(x(t), t).$$

Time discretization: let

$$t_j = jh, j = 0, 1, \ldots$$

and write

$$x_j = x(t_j).$$

Finite difference equation, using forward Euler (for simplicity; I apologize),

$$x_{j+1} = x_j + hf(x_j, t_j) + v_{j+1},$$

where v_{j+1} accounts for discretization errors as well as possible deviations from the ideal world.

Bayes filtering, basic form

Evolution–observation model:

$$
\begin{aligned}
X_{j+1} &= g_j(X_j) + V_{j+1}, \quad j = 0, 1, 2, \ldots \\
Y_j &= f_j(X_j) + E_j, \quad j = 1, 2, \ldots .
\end{aligned}
$$

Observations, or data:

$$Y_j = y_j, \quad j = 1, 2, \ldots$$

We assume further that the prior probability density of X_0 is given.

Adaptive algorithm

The goal is to design an algorithm along the following lines:

- Given the density of X_0, *predict* the density of X_1 using the prior evolution model,
- Using the predicted density of X_1 as prior, calculate the posterior density $\pi(x_1 \mid y_1)$,
- used the posterior density $\pi(x_1 \mid y_1)$, predict the density of X_2,
- Using the predicted density of X_2 as prior, calculate the posterior density $\pi(x_2 \mid y_2)$,
- Continue similarly.

Hence, what we need is

- Prediction step: Given the density of X_j, calculate the density of X_{j+1} from

$$X_{j+1} = g_j(X_j) + V_{j+1}.$$

  (Propagation problem)

- Correction step: Given the prior density of X_j, calculate the posterior density $\pi(x_j \mid y_j)$ using the observation model

$$Y_j = f(X_j) + E_j.$$

  (Inverse problem)

Particular approaches

- Linear model, Gaussian innovation and error: classical Kalman filtering.
- Linearization, approximation by Gaussian densities: Extended Kalman filtering.
- Nonlinear and/or non-Gaussian models: MCMC approach, known as Particle filtering.

Evolution–observation model:

$$
\begin{aligned}
X_{j+1} &= AX_j + V_{j+1}, \quad j = 0, 1, 2, \ldots \\
Y_j &= BX_j + E_j, \quad j = 1, 2, \ldots.
\end{aligned}
$$

Assumptions of the noise processes and the initial process:

1. Normality:
$$V_j \sim \mathcal{N}(0, \Gamma_j), \quad E_j \sim \mathcal{N}(0, \Sigma_j).$$
2. Independency: Variables V_j, E_j, all mutually independent.
3. Initial density:
$$X_0 \sim \mathcal{N}(x_0, D_0),$$
and X_0 is independent of the noise processes.

Propagation

Observation: To completely specify a Gaussian density, it is enough to know the mean and the variance.

Assume that

$$X_j \sim \mathcal{N}(x_j, D_j).$$

Mean: We have

$$X_{j+1} = AX_j + V_{j+1},$$

implying that the mean is

$$\begin{aligned} x_{j+1} &= \mathrm{E}\{X_{j+1}\} = A\mathrm{E}\{X_j\} + \mathrm{E}\{V_{j+1}\} \\ &= Ax_j. \end{aligned}$$

Hence: Propagate the mean with A.

Covariance

$$X_{j+1} - x_{j+1} = A(X_j - x_j) + V_{j+1},$$

and by independency,

$$\begin{aligned} &= \mathrm{E}\big\{(X_{j+1} - x_{j+1})(X_{j+1} - x_{j+1})^{\mathrm{T}}\big\} \\ &= \mathrm{E}\big\{(A(X_j - x_j) + V_{j+1})(A(X_j - x_j) + V_{j+1})^{\mathrm{T}}\big\} \\ &= \mathrm{E}\big\{(A(X_j - x_j)(X_j - x_j)^{\mathrm{T}} A^{\mathrm{T}}\big\} + \mathrm{E}\big\{V_{j+1} V_{j+1}^{\mathrm{T}}\big\} \\ &= A D_j A^{\mathrm{T}} + \Gamma_{j+1}. \end{aligned}$$

Hence, after propagation,

$$X_{j+1} \sim \mathcal{N}(A x_j, A D_j A^{\mathrm{T}} + \Gamma_{j+1}).$$

Correction

To solve the correction step, consider a linear inverse problem,

$$Y = BX + E,$$

where

$$X \sim \mathcal{N}(\overline{x}, D), \quad E \sim \mathcal{N}(0, \Sigma).$$

We need to solve the posterior density $\pi(x \mid y)$.

There are two equivalent approaches, both being useful to know.

Approach 1: Use Bayes' formula

Bayes' formula says that

$$\pi(x \mid y) \propto \pi_{\text{prior}}(x)\pi(y \mid x).$$

In this case,

$$\pi_{\text{prior}}(x) \propto \exp\left(-\frac{1}{2}(x-\overline{x})^{\mathrm{T}} D^{-1}(x-\overline{x})\right),$$

$$\pi(y \mid x) \propto \exp\left(-\frac{1}{2}(y-Bx)^{\mathrm{T}} \Sigma^{-1}(y-Bx)\right).$$

Therefore

$$\pi(x \mid y) \propto \exp\left(-\frac{1}{2}(x-\overline{x})^{\mathrm{T}} D^{-1}(x-\overline{x})-\frac{1}{2}(y-Bx)^{\mathrm{T}} \Sigma^{-1}(y-Bx)\right).$$

Organize terms according to their order:

Denote $x_\mathrm{c} = x - \overline{x}$, $y_\mathrm{c} = y - B\overline{x}$.

$$(x - \overline{x})^\mathrm{T} D^{-1} (x - \overline{x}) + (y - Bx)^\mathrm{T} \Sigma^{-1} (y - Bx)$$

$$= x_\mathrm{c}^\mathrm{T} D^{-1} x_\mathrm{c} + (y_\mathrm{c} - Bx_\mathrm{c})^\mathrm{T} \Sigma^{-1} (y_\mathrm{c} - Bx_\mathrm{c})$$

$$\underbrace{x_\mathrm{c}^\mathrm{T} (D^{-1} + B^\mathrm{T} \Sigma^{-1} B) x_\mathrm{c}}_{\text{quadratic}} - \underbrace{2x_\mathrm{c}^\mathrm{T} B^\mathrm{T} \Sigma^{-1} y_\mathrm{c}}_{\text{linear}} + \underbrace{y_\mathrm{c}^\mathrm{T} \Sigma^{-1} y_\mathrm{c}}_{\text{constant}} .$$

Denote

$$C = D^{-1} + B^{\mathrm{T}}\Sigma^{-1}B, \quad q = B^{\mathrm{T}}\Sigma^{-1}y_{\mathrm{c}}.$$

The above expression reads

$$x_{\mathrm{c}}^{\mathrm{T}} C x_{\mathrm{c}}^{\mathrm{T}} - 2x_{\mathrm{c}}^{\mathrm{T}} q + \text{constant}$$

$$= (x_{\mathrm{c}} - C^{-1}q)^{\mathrm{T}} C (x_{\mathrm{c}} - C^{-1}q) + \text{another constant}.$$

Hence, the posterior density is

$$\pi(x \mid y) \propto \exp\left(-\frac{1}{2}(x_{\mathrm{c}} - C^{-1}q)^{\mathrm{T}} C (x_{\mathrm{c}} - C^{-1}q)\right),$$

Explicitly:

$$\pi(x \mid y) \propto \exp\left(-\frac{1}{2}(x-\widehat{x})^{\mathrm{T}}\Gamma^{-1}\mathrm{post}(x-\widehat{x})\right),$$

where

$$\widehat{x} = \overline{x} + (D^{-1} + B^{\mathrm{T}}\Sigma^{-1}B)^{-1}(B^{\mathrm{T}}\Sigma^{-1}(y - B\overline{x}))$$

$$\Gamma_{\mathrm{post}} = (D^{-1} + B^{\mathrm{T}}\Sigma^{-1}B)^{-1}$$

Notice, if

$$D = \gamma^2 I, \quad \Sigma = \sigma^2 I,$$

the midpoint is simply

$$\begin{aligned} \widehat{x} &= \overline{x} + \left(\frac{1}{\gamma^2} I + \frac{1}{\sigma^2} B^{\mathrm{T}} B \right)^{-1} \left(\frac{1}{\sigma^2} B^{\mathrm{T}} (y - B\overline{x}) \right) \\ &= \overline{x} + \left(\delta I + B^{\mathrm{T}} B \right)^{-1} \left(B^{\mathrm{T}} (y - B\overline{x}) \right), \quad \delta = \frac{\sigma^2}{\gamma^2}. \end{aligned}$$

This is the Tikhonov regularized solution with regularization parameter δ.

Approach 2: Use conditioning

Consider the random variable

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} X \\ BX + E \end{bmatrix}.$$

Calculate the mean and covariance:

$$\mathrm{E}\{Z\} = \begin{bmatrix} \overline{x} \\ B\overline{x} \end{bmatrix}.$$

For calculating the covariance, assume for simplicity that the means have been subtracted.

$$ZZ^{\mathrm{T}} = \begin{bmatrix} XX^{\mathrm{T}} & X(BX+E)^{\mathrm{T}} \\ (BX+E)X^{\mathrm{T}} & (BX+E)(BX+E)^{\mathrm{T}} \end{bmatrix}.$$

Expectation, remembering that $\mathrm{E}\{XE^{\mathrm{T}}\} = 0$, gives

$$\mathrm{Cov}(Z) = \begin{bmatrix} D & DB^{\mathrm{T}} \\ BD & BDB^{\mathrm{T}} + \Sigma \end{bmatrix} = M.$$

Hence, the joint probability density of X and Y is

$$\pi(x, y) \propto \exp\left(-\frac{1}{2} \begin{bmatrix} x - \overline{x} \\ y - B\overline{x} \end{bmatrix}^{\mathrm{T}} M^{-1} \begin{bmatrix} x - \overline{x} \\ y - B\overline{x} \end{bmatrix} \right).$$

Unfolding:

We know that

$$\pi(x \mid y) \propto \pi(x, y)$$

Writing a partitioning of M^{-1} as

$$M^{-1} = S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix},$$

we arrange terms of different degree of $x_\mathrm{c} = x - \overline{x}$:

$$\begin{bmatrix} x - \overline{x} \\ y - B\overline{x} \end{bmatrix}^\mathrm{T} \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} x - \overline{x} \\ y - B\overline{x} \end{bmatrix} = \begin{bmatrix} x_\mathrm{c} \\ y_\mathrm{c} \end{bmatrix}^\mathrm{T} \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} x_\mathrm{c} \\ y_\mathrm{c} \end{bmatrix}$$

$$= x_\mathrm{c}^\mathrm{T} S_{11} x_\mathrm{c} + 2x_\mathrm{c}^\mathrm{T} S_{12} y_\mathrm{c} + \text{constant} \quad (S_{21} = S_{12}^\mathrm{T})$$

$$= (x_\mathrm{c} + S_{11}^{-1} S_{12} y_\mathrm{c})^\mathrm{T} S_{11} (x_\mathrm{c} + S_{11}^{-1} S_{12} y_\mathrm{c}) + \text{another constant}.$$

Hence, we have

$$\pi(x \mid y) \propto \exp\left(-\frac{1}{2}(x-\widehat{x})^{\mathrm{T}}\Gamma^{-1}\mathrm{post}(x-\widehat{x})\right),$$

where

$$\widehat{x} = \overline{x} - S_{11}^{-1}S_{12}(y - B\overline{x}),$$

$$\Gamma_{\mathrm{post}} = S_{11}^{-1}.$$

We need to resolve S_{11} and S_{12}.

We have

$$S = M^{-1}.$$

Partitionings

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \quad M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}.$$

Consider the equation

$$Ma = b, \text{or equivalently, } a = Sb.$$

Let

$$a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

$$Ma = \left[\begin{array}{c} M_{11}a_1 + M_{12}a_2 \\ M_{21}a_1 + M_{22}a_2 \end{array} \right] = \left[\begin{array}{c} b_1 \\ b_2 \end{array} \right].$$

From the second equation, we eliminate a_2:

$$a_2 = M_{22}^{-1}(b_2 - M_{21}a_1).$$

Substitute into the first equation and solve for a_1:

$$(M_{11} - M_{12}M_{22}^{-1}M_{21})a_1 = b_1 - M_{12}M_{22}^{-1}b_2,$$

or

$$a_1 = \underbrace{(M_{11} - M_{12}M_{22}^{-1}M_{21})^{-1}}_{=S_{11}} b_1 + \underbrace{-(M_{11} - M_{12}M_{22}^{-1}M_{21})^{-1}M_{12}M_{22}^{-1}}_{=S_{12}} b_2.$$

The matrix

$$\widetilde{M}_{22} = M_{11} - M_{12} M_{22}^{-1} M_{21}$$

is called the *Schur complement* of M_{22}.

In terms of Schur complements,

$$\begin{aligned} \widehat{x} &= \overline{x} - S_{11}^{-1} S_{12} (y - B\overline{x}) \\ &= \overline{x} + M_{12} M_{22}^{-1} (y - B\overline{x}), \end{aligned}$$

$$\Gamma_{\text{post}} = \widetilde{M}_{22},$$

where

$$\begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} = \begin{bmatrix} D & DB^{\mathrm{T}} \\ BD & BDB^{\mathrm{T}} + \Sigma \end{bmatrix}.$$

Finally, we have

$$\widehat{x} = \overline{x} + DB^{\mathrm{T}}(BDB^{\mathrm{T}} + \Sigma)^{-1}(y - B\overline{x}),$$

and

$$\Gamma_{\mathrm{post}} = D - DB^{\mathrm{T}}(BDB^{\mathrm{T}} + \Sigma)^{-1}BD.$$

Consider again the special case,

$$D = \gamma^2 I, \quad \Sigma = \sigma^2 I.$$

We have

$$\begin{aligned} \widehat{x} &= \overline{x} + \gamma^2 B^{\mathrm{T}} (\gamma^2 B B^{\mathrm{T}} + \sigma^2 I)^{-1} (y - B\overline{x}) \\ &= \overline{x} + B^{\mathrm{T}} (B B^{\mathrm{T}} + \delta I)^{-1} (y - B\overline{x}), \quad \delta = \frac{\sigma^2}{\gamma^2} \end{aligned}$$

This is known as the *Wiener filter* solution to the ill-posed problem $y = Bx$.

Wiener versus Tikhonov

We have found to *equivalent* formulas for the midpoint and the variance:

$$\begin{aligned}\widehat{x} &= \overline{x} + B^{\mathrm{T}}(BB^{\mathrm{T}} + \delta I)^{-1}(y - B\overline{x}) \\ &= \overline{x} + \left(\delta I + B^{\mathrm{T}}B\right)^{-1}\left(B^{\mathrm{T}}(y - B\overline{x})\right),\end{aligned}$$

and

$$\begin{aligned}\Gamma_{\mathrm{post}} &= \gamma^2(I - B^{\mathrm{T}}(BB^{\mathrm{T}} + \delta I)^{-1}B) \\ &= \gamma^2\left(I + \frac{1}{\delta}B^{\mathrm{T}}B\right)^{-1}\end{aligned}$$

Which formula to use?

Engineering rule of thumb:

Let $B \in \mathbb{R}^{m \times n}$. Then

$$B^{\mathrm{T}} B \in \mathbb{R}^{n \times n}, \quad B B^{\mathrm{T}} \in \mathbb{R}^{m \times m}.$$

It is tempting to say:

Wiener filter for underdetermined problems $(m < n)$

Tikhonov for overdetermined problems $(n < m)$

In practice, all depends on solvers!

1. Initialize: $j = 0$, x_0 and D_0 given.

2. *Prediction step:* Calculate

$$\begin{aligned}\overline{x}_{j+1} &= Ax_j, \\ \overline{D}_{j+1} &= AD_jA^{\mathrm{T}} + \Gamma_{j+1}.\end{aligned}$$

3. *Updating step:* Calculate

$$\begin{aligned}x_{j+1} &= \overline{x}_{j+1} + \overline{D}_{j+1}B^{\mathrm{T}}(B\overline{D}_{j+1}B^{\mathrm{T}} + \Sigma_{j+1})^{-1}(y_{j+1} - B\overline{x}_{j+1}), \\ D_{j+1} &= \overline{D}_{j+1} - \overline{D}_{j+1}B^{\mathrm{T}}(B\overline{D}_{j+1}B^{\mathrm{T}} + \Sigma_{j+1})^{-1}B\overline{D}_{j+1}.\end{aligned}$$

4. Increase j by one and repeat from 2.

Comments

The above version is based on the Wiener filter form of the updating. Alternatively, one may use the regularized normal equation (Tikhonov) form.

The expensive step in Kalman filtering is the computation of the covariance in Step 3.

$$K_{j+1} = \overline{D}_{j+1} B^{\mathrm{T}} (B \overline{D}_{j+1} B^{\mathrm{T}} + \Sigma_{j+1})^{-1}$$

is often called the *Kalman gain* matrix.