Introduction to Data Assimilation
Lecture 4 Notes

Quanling Deng, YMSC, Tsingha University

1 Multi-Dimensional Kalman Filter

Model and Notation

Let $\mathbf{u}_m \in \mathbb{C}^N$ be the state and $\mathbf{v}_{m+1} \in \mathbb{C}^M$ the observation. We consider the linear model with affine forcing and Gaussian noises:

$$\mathbf{u}_{m+1} = \mathbf{A}\,\mathbf{u}_m + \mathbf{f}_m + \boldsymbol{\sigma}_m, \qquad \boldsymbol{\sigma}_m \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), \tag{1}$$

$$\mathbf{v}_{m+1} = \mathbf{G}\,\mathbf{u}_{m+1} + \boldsymbol{\sigma}^o_{m+1}, \qquad \boldsymbol{\sigma}^o_{m+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}^o), \tag{2}$$

with $\mathbf{A} \in \mathbb{C}^{N\times N}$, $\mathbf{G} \in \mathbb{C}^{M\times N}$, and covariance matrices $\mathbf{R} \succeq \mathbf{0}$, $\mathbf{R}^o \succ \mathbf{0}$ (positive semidefinite, PSD). Noises are mutually independent and independent of past states.

Forecast (Prediction) Step

Given the analysis ($\bar{\mathbf{u}}_{m|m}$, $\mathbf{R}_{m|m}$) at time m,

$$\bar{\mathbf{u}}_{m+1|m} = \mathbf{A}\,\bar{\mathbf{u}}_{m|m} + \mathbf{f}_m, \tag{3}$$

$$\mathbf{R}_{m+1|m} = \mathbf{A}\,\mathbf{R}_{m|m}\,\mathbf{A}^* + \mathbf{R}. \tag{4}$$

Analysis (Filtering) Step

Given $\mathbf{v}_{m+1}$, the Kalman gain, posterior mean, and posterior covariance are

$$\mathbf{K}_{m+1} = \mathbf{R}_{m+1|m}\,\mathbf{G}^*\big(\mathbf{G}\,\mathbf{R}_{m+1|m}\,\mathbf{G}^* + \mathbf{R}^o\big)^{-1}, \tag{5}$$

$$\bar{\mathbf{u}}_{m+1|m+1} = \bar{\mathbf{u}}_{m+1|m} + \mathbf{K}_{m+1}\big(\mathbf{v}_{m+1} - \mathbf{G}\,\bar{\mathbf{u}}_{m+1|m}\big), \tag{6}$$

$$\mathbf{R}_{m+1|m+1} = (\mathbf{I} - \mathbf{K}_{m+1}\mathbf{G})\,\mathbf{R}_{m+1|m}\,(\mathbf{I} - \mathbf{K}_{m+1}\mathbf{G})^* + \mathbf{K}_{m+1}\,\mathbf{R}^o\,\mathbf{K}^*_{m+1}. \tag{7}$$

In exact arithmetic with the optimal gain (5), the simplified covariance form

$$\mathbf{R}_{m+1|m+1} = (\mathbf{I} - \mathbf{K}_{m+1}\mathbf{G})\,\mathbf{R}_{m+1|m}$$

also holds; the Joseph form (7) is preferred for numerical robustness (losing symmetry).

Dimensions of Observation vs. State

The observation dimension M need not equal the state dimension N.

- **Partial observations** ($M \ll N$): each $\mathbf{v}_{m+1}$ senses only a few linear combinations of state components via $\mathbf{G}$; recovering unobserved components relies on dynamical coupling in $\mathbf{A}$ and repeated assimilation cycles.
- **Over-determined observations** ($M \gtrsim N$): e.g., many sensors measuring a lower-dimensional state; the same formulas apply with $\mathbf{G} \in \mathbb{C}^{M\times N}$ (Lagrangian DA).

1.1 Observability in the Multi-Dimensional Setting (Illustrative 2D Case)

Definition 1.1 (Observability). *The pair (F, g) is* observable *if the data channel is sensitive to the state.*

Definition 1.2 (Stochastic controllability). *The model is* stochastically controllable *if the process noise injects uncertainty into the state $\mathbf{R} \succ 0$ (positive definite).*

Example: Consider

$$\begin{aligned} u^{(1)}_{m+1} &= A_{11}\, u^{(1)}_m + A_{12}\, u^{(2)}_m + \sigma^{(1)}_m, \\ u^{(2)}_{m+1} &= A_{21}\, u^{(1)}_m + A_{22}\, u^{(2)}_m + \sigma^{(2)}_m, \end{aligned} \qquad v_{m+1} = g_1\, u^{(1)}_{m+1} + \sigma^o_{m+1},$$

i.e., only the first component is directly observed (scalar $g_1 \in \mathbb{C}$). Information can improve estimates of $u^{(2)}$ *only if* $u^{(2)}$ dynamically couples to $u^{(1)}$ through $A_{12} \neq 0$ or $A_{21} \neq 0$; otherwise $u^{(2)}$ remains effectively unobservable from v_{m+1} alone.

Remark 1.3 (Lighthouse tracking). A ship (u_m) at night is tracked by a lighthouse (g). No light $(g = 0)$ means no tracking; no waves $(r = 0)$ means perfectly predictable motion, but a slightly wrong heading can grow unless you get periodic flashes ($g \neq 0$, observability) and direction correction ($r = 0$, error correction, controllability) to update the course.

Takeaways for DA.

- *Observability over time:* Partial, indirect measurements can still render the system observable if the dynamics couple hidden to measured states.
- *Stochastic controllability:* Positive-definite Q (or, more precisely, $W_c(N) \succ 0$) is essential: directions not excited by the model noise cannot be corrected unless they become observable through coupling.
- *Design hint:* When building low-order DA models, ensure (F, H) is observable and that the noise model (F, R) excites all dynamically important directions.
- Measurement noise $R^o \succ 0$ does *not* affect observability/controllability themselves, only the statistical weight in the update.

2 Filtering Stability for the 1D Kalman Filter

2.1 Mean-square stability and the scalar Riccati map

Define the posterior variance $P_m := r_{m|m}$ and the prior $S_m := r_{m+1|m} = |F|^2 P_m + r$. The scalar Riccati recursion is

$$P_{m+1} = \Phi(S_m) := \frac{S_m\, r^o}{r^o + |g|^2 S_m}, \qquad S_m = |F|^2 P_m + r. \tag{8}$$

The map Φ is increasing and concave on $[0, \infty)$ with $\Phi(0) = 0$ and $\Phi(S) \nearrow r^o/|g|^2$ as $S \to \infty$.

Definition 2.1 (Mean-square stability). *The filter is* mean-square stable *if P_m remains bounded and $P_m \to P_\infty$ as $m \to \infty$ for some finite $P_\infty \geq 0$.*

Proposition 2.2 (Existence/uniqueness of P_∞). *If $g \neq 0$ and $r, r^o > 0$, the coupled map $P \mapsto \Phi(|F|^2 P + r)$ has a unique fixed point $P_\infty \in (0, r^o/|g|^2)$, and $P_m \to P_\infty$ from any $P_0 \geq 0$.*

2.2 Closed-form steady state and closed-loop factor

Let $a := F$. Eliminating $S = |a|^2 P + r$ in $P = \Phi(S)$ gives the quadratic

$$|g|^2|a|^2P^2 + \big(|g|^2 r + r^o(1 - |a|^2)\big)P - r\,r^o \;=\; 0. \tag{9}$$

The positive root is

$$P_\infty \;=\; \frac{-B + \sqrt{B^2 + 4\,|g|^2|a|^2 r\,r^o}}{2\,|g|^2|a|^2}, \qquad B := |g|^2 r + r^o(1 - |a|^2). \tag{10}$$

At steady state the prior is $S_\infty = |a|^2 P_\infty + r$ and the gain

$$K_\infty = \frac{S_\infty g^*}{r^o + |g|^2 S_\infty}, \qquad 1 - K_\infty g = \frac{r^o}{r^o + |g|^2 S_\infty} \in (0, 1).$$

The *closed-loop factor* for the mean ($u_{m+1|m+1} = (1 - Kg)u_{m+1|m} + Kv_{m+1}$) is

$$\alpha_\infty := a(1 - K_\infty g), \qquad |\alpha_\infty| = |a|\,\frac{r^o}{r^o + |g|^2 S_\infty} < 1,$$

so the mean recursion is stable.

Remark 2.3 (GPS vs. inertial guidance). Inertial sensors (model) drift; periodic GPS fixes (observations) pull the solution back. Even if the inertial prediction blows up, sufficiently frequent/accurate GPS updates yield a stable fused estimate.

3 Wiener and Markov Processes; Itô Integral and Itô's Formula

1. Wiener Process (Brownian Motion)

Definition 3.1 (Wiener Process)**.** *A real-valued stochastic process* $\{W(t)\}_{t\geq 0}$ *is called a* Wiener process *or* Brownian motion *if:*

1. $W(0) = 0$ *almost surely (a.s.),*
2. $W(t)$ *is continuous a.s.,*
3. $W(t)$ *has independent increments, with*

$$W(t) - W(s) \sim \mathcal{N}(0, t - s), \qquad 0 \leq s < t.$$

From property (3), increments satisfy:

$$\mathbb{E}[\Delta W_i] = 0, \qquad \mathbb{E}[(\Delta W_i)^2] = \Delta t_i,$$

where $\Delta W_i = W(t_i) - W(t_{i-1})$ and $\Delta t_i = t_i - t_{i-1}$. Hence sample paths of $W(t)$ are continuous but almost surely nowhere differentiable.

A *complex Wiener process* is defined by

$$W_c(t) = \frac{1}{\sqrt{2}}\big(W_R(t) + iW_I(t)\big),$$

where W_R and W_I are independent real Wiener processes. The scaling factor ensures the same variance as in the real case.

2. Markov Process

Definition 3.2 (Markov Property)**.** *A stochastic sequence* $\{x_n\}$ *has the* Markov property *if its conditional density satisfies*

$$p(x_n|x_{n-1}, x_{n-2}, \ldots, x_0) = p(x_n|x_{n-1}).$$

That is, the future state depends only on the present, not on the history. The Wiener process is a particular case of a Markov process. Depending on whether the state and time are continuous or discrete, common types include:

- **Markov chain:** discrete time, discrete state.
- **Markov jump process:** continuous time, discrete state.
- **SDE (stochastic differential equation):** continuous in both time and space.
- **Stochastic difference equation:** discrete time, continuous state.

3. Itô Stochastic Integral

Let $W(t)$ be a Wiener process and $G(x,t)$ an adapted (non-anticipating) process. The *Itô integral* is defined as

$$\int_{t_0}^{t} G(x(s), s)\, dW(s) := \text{m.s.} \lim_{n\to\infty} \sum_{j=1}^{n} G(x(t_{j-1}), t_{j-1})[W(t_j) - W(t_{j-1})], \tag{11}$$

where the limit is in the mean-square sense:

$$\lim_{n\to\infty} \mathbb{E}\Big[\big|X_n - X\big|^2\Big] = 0.$$

Adaptedness means $G(x,t)$ depends only on information available up to time t.

The function G must be evaluated at the *left endpoint* t_{j-1} of each subinterval. Evaluating instead at the midpoint defines the *Stratonovich integral*, which obeys different calculus rules but can be converted to Itô form.

Key Properties of Itô Calculus. For adapted processes G, H:

$$\begin{aligned}
\int_{t_0}^{t} G(s)\,[dW(s)]^2 &= \int_{t_0}^{t} G(s)\, ds, \quad \text{since } dW^2 = dt, \\
\mathbb{E}\left[\int_{t_0}^{t} G(s)\, dW(s)\right] &= 0, \\
\mathbb{E}\left[\left(\int_{t_0}^{t} G(s)\, dW(s)\right)\left(\int_{t_0}^{t} H(s)\, dW(s)\right)\right] &= \int_{t_0}^{t} \mathbb{E}[G(s)H(s)]\, ds.
\end{aligned}$$

For a smooth function $f(W(t), t)$,

$$df = \left(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial W^2}\right) dt + \frac{\partial f}{\partial W} dW.$$

A useful identity is

$$\int_{t_0}^{t} W(s)\, dW(s) = \tfrac{1}{2}\big[W^2(t) - W^2(t_0) - (t - t_0)\big].$$

4. Itô Stochastic Differential Equation (SDE)

Given $X(t_0) = X_0$, an *SDE* is written as

$$\frac{dX(t)}{dt} = A(X(t), t) + B(X(t), t)\, \dot{W}(t), \tag{12}$$

or equivalently in differential form,

$$dX(t) = A(X(t), t)\, dt + B(X(t), t)\, dW(t), \tag{13}$$

where $\dot{W}(t)$ denotes Gaussian white noise (the formal time derivative of $W(t)$).

The integral solution of (13) is

$$X(t) = X(t_0) + \int_{t_0}^{t} A(X(s), s)\, ds + \int_{t_0}^{t} B(X(s), s)\, dW(s), \tag{14}$$

where the second integral is the Itô integral (11).

5. Itô's Formula (Itô's Lemma)

Let x_t satisfy

$$dx_t = a_t\, dt + b_t\, dW_t, \quad \text{where } a_t = a(x_t, t),\ b_t = b(x_t, t).$$

For a smooth function $f(x, t)$, Itô's formula gives:

$$\begin{aligned} df(x_t, t) &= \frac{\partial f}{\partial x}\, dx_t + \tfrac{1}{2} \frac{\partial^2 f}{\partial x^2} (dx_t)^2 \\ &= \left(a_t \frac{\partial f}{\partial x} + \tfrac{1}{2} b_t^2 \frac{\partial^2 f}{\partial x^2} \right) dt + b_t \frac{\partial f}{\partial x}\, dW_t. \end{aligned} \tag{15}$$

This is the stochastic analogue of the chain rule. The extra term $\frac{1}{2} b_t^2 f_{xx}$ arises because $(dW_t)^2 = dt$ in mean-square sense.

6. Remarks and Discussion

- **Wiener process paths** are continuous but nowhere differentiable; thus, Itô integration replaces the usual Riemann calculus.
- **Markov property** ensures future evolution depends only on the current state—critical for deriving recursive filters like the Kalman filter.
- **Itô vs. Stratonovich:** Itô's integral uses left-endpoint evaluation and introduces correction terms like $\frac{1}{2} b^2 f_{xx}$; Stratonovich preserves ordinary chain rules but differs in stochastic interpretation.
- **Itô's lemma** is indispensable for deriving analytical solutions and moment equations for SDEs (e.g., Ornstein–Uhlenbeck process).
- **Application remark:** In data assimilation, SDEs driven by Wiener noise model uncertain dynamics, while Itô calculus enables proper estimation of variance propagation.