Applied Math Reading Seminar

Data Assimilation III: Extended and Unscented Kalman Filters

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

April 3, 2023

Data Assimilation Setup

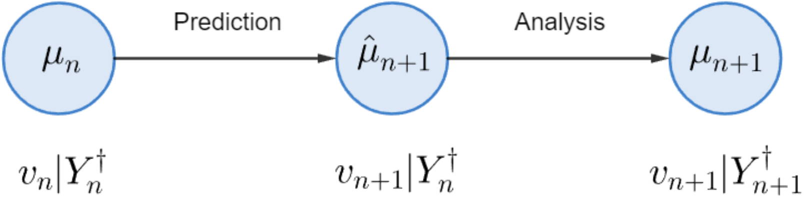
State Space Model

$$\begin{aligned}
\text{Dynamics Model:}\quad & v_{n+1} = \Psi\left(v_n\right) + \xi_n,\ \ n \geq 0 \\
\text{Data Model:}\quad & y_{n+1} = h(v_{n+1}) + \eta_{n+1},\ \ n \geq 0 \\
\text{Probabilistic Structure:}\quad & v_0 \sim \mathcal{N}\left(m_0, C_0\right),\ \xi_n \sim \mathcal{N}(0, \Sigma),\ \eta_n \sim \mathcal{N}(0, \Gamma) \text{ i.i.d.} \\
\text{Probabilistic Structure:}\quad & v_0 \perp \{\xi_n\} \perp \{\eta_n\}
\end{aligned}$$

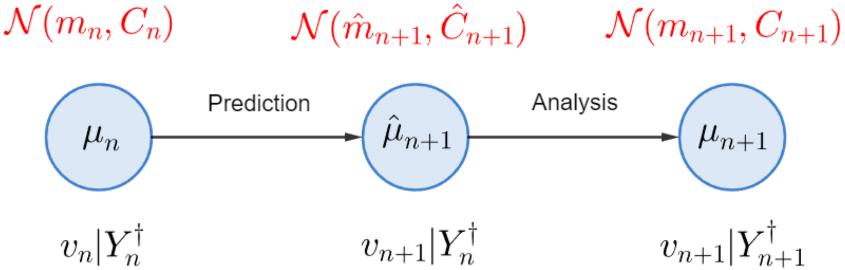
Assume C_0, Σ, Γ are positive definite, $\Psi \in C(\mathbb{R}^d, \mathbb{R}^d)$, and $h \in C(\mathbb{R}^d, \mathbb{R}^k)$. In general, Ψ and h will be nonlinear functions. Let $y_i^\dagger$ denote the observed value of y_i, and $Y_n^\dagger := \{y_1^\dagger, \dots, y_n^\dagger\}$.

Filtering Problem: Find the distribution μ_n of $v_n | Y_n^\dagger,\ n \geq 0$.

Prediction and Analysis



Linear case


$\mathcal{N}(m_n, C_n)$
$\mathcal{N}(\hat{m}_{n+1}, \hat{C}_{n+1})$
$\mathcal{N}(m_{n+1}, C_{n+1})$
μ_n
Prediction
$\hat{\mu}_{n+1}$
Analysis
μ_{n+1}
$v_n | Y_n^\dagger$
$v_{n+1} | Y_n^\dagger$
$v_{n+1} | Y_{n+1}^\dagger$

Kalman filter algorithm for linear case

$\Psi(\cdot) = M$ and $h(\cdot) = H$,

$$\begin{aligned} v_{n+1} &= M v_n + \xi_n, && \xi_n \sim \mathcal{N}(0, \Sigma) \\ y_{n+1} &= H v_{n+1} + \eta_{n+1}, && \eta_n \sim \mathcal{N}(0, \Gamma) \end{aligned}$$

Kalman filter algorithm for linear case

$\Psi(\cdot) = M$ and $h(\cdot) = H$,

$$\begin{aligned} v_{n+1} &= Mv_n + \xi_n, && \xi_n \sim \mathcal{N}(0, \Sigma) \\ y_{n+1} &= Hv_{n+1} + \eta_{n+1}, && \eta_n \sim \mathcal{N}(0, \Gamma) \end{aligned}$$

- **Prediction Step.**

$$\begin{aligned} \hat{m}_{n+1} &= Mm_n, \\ \hat{C}_{n+1} &= MC_nM^\top + \Sigma \end{aligned}$$

Kalman filter algorithm for linear case

$\Psi(\cdot) = M$ and $h(\cdot) = H$,

$$\begin{aligned} v_{n+1} &= M v_n + \xi_n, & \xi_n &\sim \mathcal{N}(0, \Sigma) \\ y_{n+1} &= H v_{n+1} + \eta_{n+1}, & \eta_n &\sim \mathcal{N}(0, \Gamma) \end{aligned}$$

- **Prediction Step.**

$$\begin{aligned} \hat{m}_{n+1} &= M m_n, \\ \hat{C}_{n+1} &= M C_n M^\top + \Sigma \end{aligned}$$

- **Analysis Step.**

$$\begin{aligned} m_{n+1} &= \hat{m}_{n+1} + \hat{C}_{n+1} H^\top (H \hat{C}_{n+1} H^\top + \Gamma)^{-1} (y^\dagger_{n+1} - H \hat{m}_{n+1}), \\ C_{n+1} &= \hat{C}_{n+1} - \hat{C}_{n+1} H^\top (H \hat{C}_{n+1} H^\top + \Gamma)^{-1} H \hat{C}_{n+1} \end{aligned}$$

Kalman filter algorithm for linear case

$\Psi(\cdot) = M$ and $h(\cdot) = H$,

$$
\begin{aligned}
v_{n+1} &= Mv_n + \xi_n, & \xi_n &\sim \mathcal{N}(0, \Sigma) \\
y_{n+1} &= Hv_{n+1} + \eta_{n+1}, & \eta_n &\sim \mathcal{N}(0, \Gamma)
\end{aligned}
$$

- **Prediction Step.**

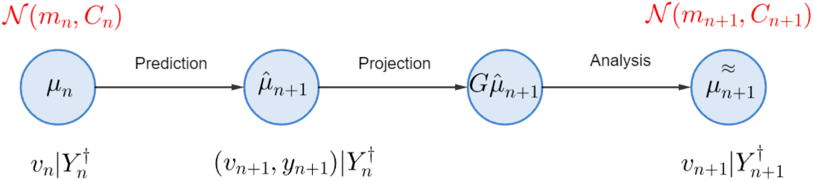
$$
\begin{aligned}
\hat{m}_{n+1} &= Mm_n, \\
\hat{C}_{n+1} &= MC_nM^\top + \Sigma
\end{aligned}
$$

- **Analysis Step.**

$$
\begin{aligned}
m_{n+1} &= \hat{m}_{n+1} + \hat{C}_{n+1}H^\top(H\hat{C}_{n+1}H^\top + \Gamma)^{-1}(y^\dagger_{n+1} - H\hat{m}_{n+1}), \\
C_{n+1} &= \hat{C}_{n+1} - \hat{C}_{n+1}H^\top(H\hat{C}_{n+1}H^\top + \Gamma)^{-1}H\hat{C}_{n+1}
\end{aligned}
$$

- What about nonlinear case?

Gaussian Projected Filtering



Gaussian Projection

Definition

Let $\mathfrak{G}(\mathbb{R}^d)$ denote the set of Gaussian probability measures on $\mathbb{R}^d$. For any probability measure μ on $\mathbb{R}^d$, define its *Gaussian projection* as the Gaussian distribution with the same mean and covariance matrix as μ, i.e. for $u \sim \mu$,

$$G\mu = \mathcal{N}(m^\mu, C^\mu), \quad m^\mu = \mathbb{E}(u), \quad C^\mu = \mathbb{E}((u - \mathbb{E}(u))(u - \mathbb{E}(u))^\top).$$

Gaussian Projection

Definition

Let $\mathfrak{G}(\mathbb{R}^d)$ denote the set of Gaussian probability measures on $\mathbb{R}^d$. For any probability measure μ on $\mathbb{R}^d$, define its *Gaussian projection* as the Gaussian distribution with the same mean and covariance matrix as μ, i.e. for $u \sim \mu$,

$$G\mu = \mathcal{N}(m^\mu, C^\mu), \quad m^\mu = \mathbb{E}(u), \quad C^\mu = \mathbb{E}((u - \mathbb{E}(u))(u - \mathbb{E}(u))^\top).$$

It is possible to show that $G\mu$ is the closest Gaussian distribution to μ, if we measure distance between probability distributions by the Kullback-Leibler divergence:

$$G\mu = \mathrm{argmin}_{\pi \in \mathfrak{G}} d_{\mathsf{KL}}(\mu \| \pi).$$

Formulas for Gaussian Projected Filtering

- **Prediction and Gaussian Projection Step.** $v_n \sim \mu_n = \mathcal{N}(m_n, C_n)$,

$$\hat{v}_{n+1} = \Psi(v_n) + \xi_n$$
$$\hat{y}_{n+1} = h(\hat{v}_{n+1}) + \eta_{n+1},$$

Formulas for Gaussian Projected Filtering

- **Prediction and Gaussian Projection Step.** $v_n \sim \mu_n = \mathcal{N}(m_n, C_n)$,

$$\begin{aligned} \hat{v}_{n+1} &= \Psi(v_n) + \xi_n \\ \hat{y}_{n+1} &= h(\hat{v}_{n+1}) + \eta_{n+1}, \end{aligned}$$

$$\hat{m}_{n+1} := \mathbb{E}(\hat{v}_{n+1}), \qquad \hat{C}_{n+1} := \mathrm{Cov}(\hat{v}_{n+1}), \qquad \hat{o}_{n+1} := \mathbb{E}(\hat{y}_{n+1}),$$

$$\hat{C}^{vy}_{n+1} := \mathrm{Cov}(\hat{v}_{n+1}, \hat{y}_{n+1}), \quad \hat{C}^{yy}_{n+1} := \mathrm{Cov}(\hat{y}_{n+1}, \hat{y}_{n+1})$$

$$(v_{n+1}, y_{n+1}) | Y_n^\dagger \sim \mathcal{N}\left(\begin{bmatrix} \hat{m}_{n+1} \\ \hat{o}_{n+1} \end{bmatrix}, \begin{bmatrix} \hat{C}_{n+1} & \hat{C}^{vy}_{n+1} \\ \hat{C}^{vy}_{n+1} & \hat{C}^{yy}_{n+1} \end{bmatrix} \right), \quad \text{approximately.}$$

Formulas for Gaussian Projected Filtering

- **Prediction and Gaussian Projection Step.** $v_n \sim \mu_n = \mathcal{N}(m_n, C_n)$,

$$\begin{aligned} \hat{v}_{n+1} &= \Psi(v_n) + \xi_n \\ \hat{y}_{n+1} &= h(\hat{v}_{n+1}) + \eta_{n+1}, \end{aligned}$$

$$\begin{aligned} &\hat{m}_{n+1} := \mathbb{E}(\hat{v}_{n+1}), & &\hat{C}_{n+1} := \mathrm{Cov}(\hat{v}_{n+1}), & &\hat{o}_{n+1} := \mathbb{E}(\hat{y}_{n+1}), \\ &\hat{C}^{vy}_{n+1} := \mathrm{Cov}(\hat{v}_{n+1}, \hat{y}_{n+1}), & &\hat{C}^{yy}_{n+1} := \mathrm{Cov}(\hat{y}_{n+1}, \hat{y}_{n+1}) \end{aligned}$$

$$(v_{n+1}, y_{n+1}) | Y_n^\dagger \sim \mathcal{N}\left(\begin{bmatrix} \hat{m}_{n+1} \\ \hat{o}_{n+1} \end{bmatrix}, \begin{bmatrix} \hat{C}_{n+1} & \hat{C}^{vy}_{n+1} \\ \hat{C}^{vy}_{n+1} & \hat{C}^{yy}_{n+1} \end{bmatrix} \right), \quad \text{approximately.}$$

- **Analysis Step.** It turns out that after conditioning on $y^\dagger_{n+1}$ we obtain

$$\begin{aligned} m_{n+1} &= \hat{m}_{n+1} + \hat{C}^{vy}_{n+1}(\hat{C}^{yy}_{n+1})^{-1}(y^\dagger_{n+1} - \hat{o}_{n+1}), \\ C_{n+1} &= \hat{C}_{n+1} - \hat{C}^{vy}_{n+1}(\hat{C}^{yy}_{n+1})^{-1}(\hat{C}^{vy}_{n+1})^\top \end{aligned}$$

- How to actually compute $\hat{m}_{n+1}, \dots, \hat{C}^{yy}_{n+1}$?

Pushforwards of measures

Question

Let $\theta \in \mathbb{R}^{N_\theta}$ be a random vector with mean m and covariance matrix C, and f_1 and f_2 be two (nonlinear) functions. How can one approximate the means and covariances $\mathbb{E}[f_i(\theta)]$ and $\mathrm{Cov}[f_1(\theta), f_2(\theta)]$ of the transformed random vectors $f_1(\theta)$ and $f_2(\theta)$?

Pushforwards of measures

Question

Let $\theta \in \mathbb{R}^{N_\theta}$ be a random vector with mean m and covariance matrix C, and f_1 and f_2 be two (nonlinear) functions. How can one approximate the means and covariances $\mathbb{E}[f_i(\theta)]$ and $\mathrm{Cov}[f_1(\theta), f_2(\theta)]$ of the transformed random vectors $f_1(\theta)$ and $f_2(\theta)$?

- Linearize f_i → **Extended** Kalman Filter
- Unscented transform (quadrature rule) → **Unscented** Kalman Filter

Linearization

Question

Let $\theta \in \mathbb{R}^{N_\theta}$ be a random vector with mean m and covariance matrix C, and f_1 and f_2 be two (nonlinear) functions. How can one approximate the means and covariances $\mathbb{E}[f_i(\theta)]$ and $\mathrm{Cov}[f_1(\theta), f_2(\theta)]$ of the transformed random vectors $f_1(\theta)$ and $f_2(\theta)$?

Assuming C is small,

$$f_i(\theta) \approx f_i(m) + Df_i(m)(\theta - m)$$

$$\begin{aligned} \mathbb{E}[f_i(\theta)] &\approx f_i(m) \\ \mathrm{Cov}[f_1(\theta), f_2(\theta)] &\approx Df_1(m) \cdot C \cdot Df_2(m)^\top \end{aligned}$$

$\rightarrow$ Extended Kalman Filter

(Modified) Unscented Transform

Question

Let $\theta \in \mathbb{R}^{N_\theta}$ be a random vector with mean m and covariance matrix C, and f_1 and f_2 be two (nonlinear) functions. How can one approximate the means and covariances $\mathbb{E}[f_i(\theta)]$ and $\mathrm{Cov}[f_1(\theta), f_2(\theta)]$ of the transformed random vectors $f_1(\theta)$ and $f_2(\theta)$?

- Choose $2N_\theta + 1$ deterministic **sigma points** $\theta^0, \ldots, \theta^{2N_\theta} \in \mathbb{R}^{N_\theta}$ depending on m and C (to be precisely defined later).

(Modified) Unscented Transform

Question

Let $\theta \in \mathbb{R}^{N_\theta}$ be a random vector with mean m and covariance matrix C, and f_1 and f_2 be two (nonlinear) functions. How can one approximate the means and covariances $\mathbb{E}[f_i(\theta)]$ and $\mathrm{Cov}[f_1(\theta), f_2(\theta)]$ of the transformed random vectors $f_1(\theta)$ and $f_2(\theta)$?

- Choose $2N_\theta + 1$ deterministic **sigma points** $\theta^0, \ldots, \theta^{2N_\theta} \in \mathbb{R}^{N_\theta}$ depending on m and C (to be precisely defined later).
- Map the sigma points under f_i. Then the desired means and covariances can be approximated by a quadrature rule of the form

$$\mathbb{E}[f_i(\theta)] \approx f_i(\theta^0),$$

$$\mathrm{Cov}[f_1(\theta), f_2(\theta)] \approx \sum_{j=1}^{2N_\theta} W_j (f_1(\theta^j) - \mathbb{E}f_1(\theta))(f_2(\theta^j) - \mathbb{E}f_2(\theta))^\top.$$

- → Unscented Kalman Filter

(Modified) Unscented Transform details

The sigma points $\theta^0, \dots, \theta^{2N_\theta} \in \mathbb{R}^{N_\theta}$ are defined by

$$\theta^0 = m, \quad \theta^j = m + c_j[\sqrt{C}]_j, \quad \theta^{j+N_\theta} = m - c_j[\sqrt{C}]_j, \quad (1 \le j \le N_\theta),$$

where $[\sqrt{C}]_j$ is the jth column of the Cholesky factor of C.

$$c_1 = c_2 = \cdots = c_{N_\theta} = \sqrt{N_\theta + \lambda},$$

$$W_1 = W_2 = \cdots = W_{2N_\theta} = \frac{1}{2(N_\theta + \lambda)}$$

$$\lambda = a^2 N_\theta - N_\theta, \quad a = \min\left\{\sqrt{\frac{4}{N_\theta}}, 1\right\}$$

Other variations are possible.

References

[1] E. Calvello, S. Reich, and A. M. Stuart, *Ensemble Kalman methods: a mean field perspective*, arXiv preprint arXiv:2209.11371, (2022).

[2] D. Z. Huang, T. Schneider, and A. M. Stuart, *Unscented Kalman inversion*, arXiv preprint arXiv:2102.01580, 1 (2021).

[3] D. Sanz-Alonso, A. M. Stuart, and A. Taeb, *Inverse problems and data assimilation*, arXiv preprint arXiv:1810.06191, (2018).