Sequential Models in Data Science

Recursive Bayesian Filtering and Markov Models

Ariel Fuxman & Gilad Zusman

May 2023

1 Kalman Filtering With Biased Noise

Before deriving the equations, we explicitly give the model's assumptions. Lemmas 1 and 2 are the same from the second assignement. We think of y_k as the measurements at time k, and x_k as the state at time k.

Assumption 1. *The model satisfies:*

$$\mathbb{P}\left(x_k|x_{1:k-1}, y_{1:k-1}\right) = \mathbb{P}\left(x_k|x_{k-1}\right)$$

Assumption 2. *The model satisfies:*

$$\mathbb{P}\left(y_k|x_{1:k}, y_{1:k-1}\right) = \mathbb{P}\left(y_k|x_k\right)$$

Assumption 3. *The initial distribution of x_0 is given by:*

$$x_0 \sim \mathcal{N}\left(m_0, P_0\right),$$

where $m_0 \in \mathbb{R}^N, P_0 \in M_{N\times N}\left(\mathbb{R}\right)$ are known.

Theorem. *The **Kalman filter** with biased noise scheme is:*

1. The initial distribution of x_0 is given by:

$$x_0 \sim \mathcal{N}\left(m_0, P_0\right),$$

where $m_0 \in \mathbb{R}^N, P_0 \in M_{N\times N}(\mathbb{R})$ are known.

2. The prediction step is given by: $\mathbb{P}(x_k|y_{1:k-1}) = \mathcal{N}\left(m_k^-, P_k^-\right)$,where

$$\begin{aligned}
m_k^- &= A_{k-1}m_{k-1} + m_q, \\
P_k^- &= A_{k-1}P_{k-1}A_{k-1}^T + Q_{k-1}.
\end{aligned}$$

3. The update step is given by: $\mathbb{P}(x_k|y_{1:k}) = \mathcal{N}(m_k, P_k)$,where

$$\begin{aligned}
v_k &= y_k - H_k m_k^- - m_r \\
S_k &= H_k P_k^- H_k^T + R_k \\
K_k &= P_k^- H_k^T S_k^{-1} \\
m_k &= m_k^- + K_k v_k \\
P_k &= P_k^- - K_k S_k K_k^T
\end{aligned}$$

Proof. We begin by deriving the prediction:

Using **Assumption 1**,

$$\begin{aligned}
x_k|x_{k-1}, y_{1:k-1} = x_k|x_{k-1} &\sim \mathcal{N}(A_{k-1}x_{k-1} + m_q, Q), \\
x_{k-1}|y_{1:k-1} &\sim \mathcal{N}(m_{k-1}, P_{k-1}),
\end{aligned}$$

so by **Lemma 1**,

$$x_k|y_{1:k-1} \sim \mathcal{N}\left(m_k^-, P_k^-\right),$$

where

$$\begin{aligned}
m_k^- &= A_{k-1}m_{k-1} + m_q, \\
P_k^- &= A_{k-1}P_{k-1}A_{k-1}^T + Q_{k-1}.
\end{aligned}$$

Now, regarding the update:

Using **Assumption 2**:

$$
\begin{aligned}
y_k | x_k, y_{1:k-1} = y_k | x_k &\sim \mathcal{N}\left(H_k x_k + m_r, R_k\right), \\
x_k | y_{1:k-1} &\sim \mathcal{N}\left(m_k^-, P_k^-\right),
\end{aligned}
$$

so using **Lemma 1**,

$$
\begin{pmatrix} x_k \\ y_k \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} m_k^- \\ H_k m_k^- + m_r \end{pmatrix}, \begin{pmatrix} P_k^-, & P_k^- H_k^T \\ H_k P_k^- & H_k P_k^- H_k^T + R_k \end{pmatrix}\right).
$$

Using **Lemma 2** (allowed since P_k^- is symmetric),

$$
x_k | y_{1:k} \sim \mathcal{N}\left(m_k, P_k\right),
$$

where

$$
\begin{aligned}
m_k &= m_k^- + P_k^- H_k^T \left(H_k P_k^- H_k^T + R_k\right)^{-1} \left(y_k - H_k m_k^- - m_r\right), \\
P_k &= P_k^- - P_k^- H_k^T \left(H_k P_k^- H_k^T + R_k\right)^{-1} H_k P_k^-.
\end{aligned}
$$

□

2 Extended Kalman Filtering

1. Define:

$$
\begin{aligned}
f(x) &= x - 0.01 \sin(x) \\
h(x) &= 0.5 \sin(2x) \\
f'(x) &= 1 - 0.01 \cos(x) \\
h'(x) &= \cos(2x)
\end{aligned}
$$

so

1. The prediction step is given by: $\mathbb{P}\left(x_k|y_{1:k-1}\right) = \mathcal{N}\left(m_k^-, P_k^-\right)$,where

$$\begin{aligned}
m_k^- &= m_{k-1} - 0.01\sin\left(m_{k-1}\right), \\
P_k^- &= \left(1 - 0.01\cos\left(m_{k-1}\right)\right)^2 P_{k-1} + Q_{k-1}.
\end{aligned}$$

2. The update step is given by: $\mathbb{P}\left(x_k|y_{1:k}\right) = \mathcal{N}\left(m_k, P_k\right)$,where

$$\begin{aligned}
v_k &= y_k - 0.5\sin\left(2m_k^-\right) \\
S_k &= \left(\cos\left(2m_k^-\right)\right)^2 P_k^- + R_k \\
K_k &= \cos\left(2m_k^-\right) P_k^- S_k^{-1} \\
m_k &= m_k^- + K_k v_k \\
P_k &= P_k^- - K_k^2 S_k
\end{aligned}$$

2. To obtain the statistical linearization, we first have to calculate some integrals to find the expectations. We use definitions and lemmas in Fourier analysis. The ones seen in the course "Fourier Series and Integral Transforms" will be given without proof.

Definition 1. For $f \in L^1\left(\mathbb{R}\right)$, define the Fourier transform of f:

$$\widehat{f}\left(\omega\right) = \mathscr{F}\left[f\right]\left(\omega\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)e^{-i\omega x}dx.$$

Lemma 1. *For $f \in L^1\left(\mathbb{R}\right)$ and a frequency $\omega \in \mathbb{R}$,*

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)\cos\left(\omega x\right)dx &= 2\pi\Re\left(\widehat{f}\left(\omega\right)\right), \\
\int_{-\infty}^{\infty} f(x)\sin\left(\omega x\right)dx &= -2\pi\Im\left(\widehat{f}\left(\omega\right)\right),
\end{aligned}$$

where $\Re$ and $\Im$ are the real and imaginary parts respectively.

Lemma 2. *If $f \in L^1\left(\mathbb{R}\right)$ and $xf(x) \in L^1\left(\mathbb{R}\right)$, then $\widehat{f}\left(\omega\right) \in C^1\left(\mathbb{R}\right)$ and*

$$\mathscr{F}\left[xf(x)\right]\left(\omega\right) = i\frac{d}{d\omega}\widehat{f}\left(\omega\right).$$

Lemma 3. *The Fourier transform of a guassian is:*

$$\mathscr{F}\left[e^{-\frac{1}{2}x^2}\right](\omega) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\omega^2}.$$

Lemma 4. *We have that:*

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\cos(\omega x)\,dx = \sqrt{2\pi}e^{-\frac{1}{2}\omega^2},$$
$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\sin(\omega x)dx = 0.$$

Proof. This is an immediate consequence of **Lemmas 3** and **1**. □

Now we are going to have to compute some expectations with respect to a general gaussian variable $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$.

Lemma 5. *We have that:*

$$\mathbb{E}_{X\sim\mathcal{N}(\mu,\sigma^2)}\left[\sin(aX)\right] = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\sin(ax)\,dx = \sin(a\mu)\,e^{-\frac{1}{2}(a\sigma)^2}.$$

Proof. Substituting $t = x - \mu$, we get the integral

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}t^2}\sin(at + a\mu)\,dt = \sin(a\mu)\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}t^2}\cos(at)\,dt + \cos(a\mu)\underbrace{\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}t^2}\sin(at)\,dt}_{0} =$$
$$= \sin(a\mu)\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}t^2}\cos(at)\,dt.$$

(Here we used using **Lemma 4**). Substituting $\tau = \frac{t}{\sigma}$ and using the same **Lemma** again, we get:

$$\sigma\sin(\mu)\int_{-\infty}^{\infty} e^{-\frac{1}{2}\tau^2}\cos(a\sigma\tau)\,d\tau = \sigma\sin(a\mu)\sqrt{2\pi}e^{-\frac{1}{2}(a\sigma)^2}.$$

□

Lemma 6. *We have that:*

$$\mathscr{F}\left[xe^{-\frac{1}{2}x^2}\right](\omega) = -\frac{1}{\sqrt{2\pi}}i\omega e^{-\frac{1}{2}\omega^2}.$$

Proof. This is an immediate consequence of **Lemmas 3** and **2**. □

Lemma 7. *We have that:*

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} x \cos(\omega x)\, dx = 0,$$
$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} x \sin(\omega x) dx = \sqrt{2\pi}\omega e^{-\frac{1}{2}\omega^2}.$$

Proof. This is an immediate consequence of **Lemmas 6** and **1**. □

Lemma 8. *We have that:*

$$\mathbb{E}_{X\sim\mathcal{N}(\mu,\sigma^2)}\left[X\sin(aX)\right] = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} x\sin(ax)\, dx = e^{-\frac{1}{2}(a\sigma)^2}\left(a\sigma^2\cos(a\mu) + \mu\sin(a\mu)\right).$$

Proof. Substituting $t = x - \mu$ and using **Lemma 5**, we get the integral:

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}t^2}(t+\mu)\sin(at + a\mu)\, dt = \underbrace{\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}t^2} t\sin(at+a\mu)\, dt}_{I} + \underbrace{\mu\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}t^2}\sin(at+a\mu)\, dt}_{\mu\sigma\sin(a\mu)\sqrt{2\pi}e^{-\frac{1}{2}(a\sigma)^2}}.$$

Regarding I, we use **Lemma 7**:

$$I = \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}t^2} t\sin(at+a\mu)\, dt = \cos(a\mu)\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}t^2} t\sin(at)\, dt + \sin(a\mu)\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}t^2} t\cos(at)\, dt =$$
$$= \cos(a\mu)\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}t^2} t\sin(at)\, dt.$$

Substituting $\tau = \frac{t}{\sigma}$, we get the integral:

$$\sigma^2\cos(a\mu)\int_{-\infty}^{\infty} e^{-\frac{1}{2}\tau^2}\tau\sin(a\sigma\tau)\, d\tau = a\sigma^3\cos(a\mu)\sqrt{2\pi}e^{-\frac{1}{2}(a\sigma)^2}.$$

Returning to the original integral,

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} x\sin(x)\, dx = a\sigma^3\cos(a\mu)\sqrt{2\pi}e^{-\frac{1}{2}(a\sigma)^2} + \mu\sigma\sin(a\mu)\sqrt{2\pi}e^{-\frac{1}{2}(a\sigma)^2} =$$
$$= \sigma\sqrt{2\pi}e^{-\frac{1}{2}(a\sigma)^2}\left(a\sigma^2\cos(a\mu) + \mu\sin(a\mu)\right).$$

□

Lemma 9. *In the Kalman filter, the expectations are:*

$$\mathbb{E}\left[f(x_{k-1})\right] = m_{k-1} - 0.01\sin\left(m_{k-1}\right)e^{-\frac{1}{2}P_{k-1}}$$
$$\mathbb{E}\left[f(x_{k-1})\delta x_{x-1}\right] = P_{k-1}\left(1 - 0.01e^{-\frac{1}{2}P_{k-1}}\cos\left(m_{k-1}\right)\right)$$
$$\mathbb{E}\left[h(x_k)\right] = 0.5\sin\left(2m_k^-\right)e^{-2P_k^-}$$
$$\mathbb{E}\left[h(x_k)\delta x_x\right] = e^{-2P_k^-}P_k^-\cos\left(2m_k^-\right)$$

Proof. This is due to **Lemmas 5** and **8,** the linearity of the expectation operator and the fact that for any random variable X,

$$\mathrm{Var}\left[X\right] = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}\left[X\right]\right)^2.$$

□

We are ready obtain the statistical linearization:

1. The prediction step is given by: $\mathbb{P}\left(x_k|y_{1:k-1}\right) = \mathcal{N}\left(m_k^-, P_k^-\right)$,where

$$m_k^- = m_{k-1} - 0.01\sin\left(m_{k-1}\right)e^{-\frac{1}{2}P_{k-1}},$$
$$P_k^- = P_{k-1}\left(1 - 0.01e^{-\frac{1}{2}P_{k-1}}\cos\left(m_{k-1}\right)\right)^2 + Q_{k-1}.$$

2. The update step is given by: $\mathbb{P}\left(x_k|y_{1:k}\right) = \mathcal{N}\left(m_k, P_k\right)$,where

$$v_k = y_k - 0.5\sin\left(2m_k^-\right)e^{-2P_k^-}$$
$$S_k = e^{-4P_k^-}\cos^2\left(2m_k^-\right)P_k^- + R_k$$
$$K_k = e^{-2P_k^-}\cos\left(2m_k^-\right)P_k^-\left(S_k\right)^{-1}$$
$$m_k = m_k^- + K_k v_k$$
$$P_k = P_k^- - K_k^2 S_k$$

3 Grid Based Filter

1. Note that:

$$
\begin{aligned}
w^i_{k-1|k-1} &:= \mathbb{P}\left(x_{k-1} = x^i | y_{1:k-1}\right), \\
w^i_{k|k-1} &:= \mathbb{P}\left(x_k = x^i | y_{1:k-1}\right), \\
w^i_{k|k} &:= \mathbb{P}\left(x_k = x^i | y_{1:k}\right).
\end{aligned}
$$

Now, using marginalization,

$$
w^i_{k|k-1} = \mathbb{P}\left(x_k = x^i | y_{1:k-1}\right) = \sum_{j=1}^{N} \mathbb{P}\left(x_k = x^i | y_{1:k-1}, x_{k-1} = x^j\right) \mathbb{P}\left(x_{k-1} = x^i | y_{1:k-1}\right) = \sum_{j=1}^{N} \mathbb{P}\left(x^i | x^j\right) w^i_{k-1|k-1},
$$

since if we already know $x_{k-1} = j$, then to predict x_k, we don't care anymore about the measurements $y_{1:k-1}$.

2. Using Bayes' formula and marginalization,

$$
\begin{aligned}
w^i_{k|k} &= \mathbb{P}\left(x_k = x^i | y_{1:k}\right) = \mathbb{P}\left(x_k = x^i | y_{1:k-1}, y_k\right) = \frac{\mathbb{P}\left(x_k = x^i | y_{1:k-1}\right) \mathbb{P}\left(y_k | y_{1:k-1}, x_k = x^i\right)}{\mathbb{P}\left(y_k | y_{1:k-1}\right)} = \\
&= \frac{\mathbb{P}\left(x_k = x^i | y_{1:k-1}\right) \mathbb{P}\left(y_k | y_{1:k-1}, x_k = x^i\right)}{\sum_{j=1}^{N} \mathbb{P}\left(y_k | x_k = x^j, y_{1:k-1}\right) \mathbb{P}\left(x_k = x^j | y_{1:k-1}\right)} = \frac{w^i_{k|k-1} \mathbb{P}\left(y_k | x_k = x^i\right)}{\sum_{j=1}^{N} w^j_{k|k-1} \mathbb{P}\left(y_k | x_k = x^j\right)}.
\end{aligned}
$$

3. In our specific case, our state space is:

$$
x^i = x_0 - NL + iL, \quad i = 0, \ldots, 2N
$$

and the prediction is:

$$
\mathbb{P}\left(x^i | x^j\right) = \begin{cases} \frac{1}{2}, & i = j+1 \text{ or } i = j-1 \\ 0, & \text{otherwise} \end{cases}
$$

$$
w^i_{k|k-1} = \frac{1}{2}\left(w^{i-1}_{k-1|k-1} + w^{i+1}_{k-1|k-1}\right).
$$

(Here we assumed that if $i \notin \{0, \dots, 2N\}$ then $w^i_{k-1|k-1} = 0$).

To obtain the update, note that $r_k \sim \mathcal{U}\left(0, \frac{1}{2}\right)$, so $r_k + \frac{1}{2} \sim \mathcal{U}\left(\frac{1}{2}, 1\right)$. Then,

$$\mathbb{P}\left(y_k | x_k\right) = \mathbb{P}\left(\left(r_k + \frac{1}{2}\right) x_k \middle| x_k\right) = \mathcal{U}\left(\min\left(\frac{x_k}{2}, x_k\right), \max\left(\frac{x_k}{2}, x_k\right)\right),$$

since x_k might be negative. This allows us to compute $w^i_{k|k}$ (see the code).

4 Rejection Sampling and Monte Carlo Integration

For the sake of clarity, we denote our random variables with upper case letters: X, U.

1. By the definition of conditional probability,

$$\mathbb{P}\left(X \le t \middle| U \le \frac{f(X)}{cg(X)}\right) = \frac{\mathbb{P}\left(X \le t \middle| U \le \frac{f(X)}{cg(X)}\right)}{\mathbb{P}\left(U \le \frac{f(X)}{cg(X)}\right)}.$$

2. If $U \sim \mathcal{U}(0,1)$, then $\mathbb{P}(U \le t) = t, \ \forall t \in [0,1]$. Using marginalization,

$$\begin{aligned}\mathbb{P}\left(U \le \frac{f(X)}{cg(X)}\right) &= \int_{-\infty}^{\infty} \mathbb{P}\left(U \le \frac{f(X)}{cg(X)} \middle| X = x\right) g(x) dx = \\ &= \int_{-\infty}^{\infty} \frac{f(x)}{cg(x)} g(x) dx = \frac{1}{c} \int_{-\infty}^{\infty} f(x) dx = \frac{1}{c}.\end{aligned}$$

3. Using marginalization again,

$$\begin{aligned}\mathbb{P}\left(X \le t, U \le \frac{f(X)}{cg(X)}\right) &= \int_{-\infty}^{\infty} \mathbb{P}\left(X \le t, U \le \frac{f(X)}{cg(X)} \middle| X = x\right) g(x) dx = \\ &= \int_{-\infty}^{t} \mathbb{P}\left(U \le \frac{f(x)}{cg(x)}\right) g(x) dx = \\ &= \frac{1}{c} \int_{-\infty}^{t} f(x) dx = \frac{1}{c} F(t).\end{aligned}$$

which means that:

$$\mathbb{P}\left(X \le t \middle| U \le \frac{f(X)}{cg(X)}\right) = \frac{\mathbb{P}\left(X \le t, U \le \frac{f(X)}{cg(X)}\right)}{\mathbb{P}\left(U \le \frac{f(X)}{cg(X)}\right)} = \frac{\frac{1}{c} F(t)}{\frac{1}{c}} = F(t).$$

To compute the integral, we notice that $f(x)$ is a distribution (as a mixture of gaussians). So

$$\int_{-\infty}^{\infty} \left(x^2 - 3x + 5\right) \cos\left(x\right) f(x) dx = \mathbb{E}_{X \sim f(x)} \left[\left(X^2 - 3X + 5\right) \cos\left(X\right)\right].$$

This, combined with the central limit theorem, leads us to the following algorithm:

1. Draw N samples $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x)$ using rejection sampling.
2. Compute $Y_i = \left(X_i^2 - 3X_i + 5\right) \cos\left(X_i\right), \ \forall i = 1, \ldots, N$.
3. Return $\int_{-\infty}^{\infty} \left(x^2 - 3x + 5\right) \cos\left(x\right) f(x) dx \approx \frac{1}{N} \sum_{i=1}^{N} Y_i$.

Using a graphing software, it is evident that we can take $c = 24.75$. It is important to try to choose c as small as possible so the algorithm runs faster (since then the probability that our sample gets rejected is lower).