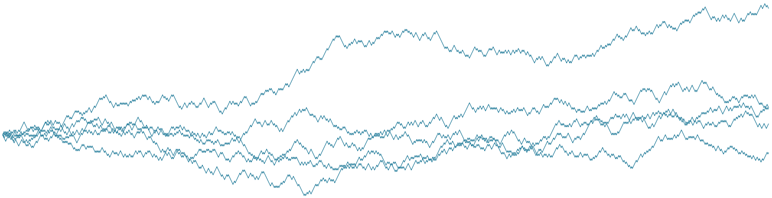
The Continuous-time Kalman Filter

UC Berkeley STAT 248, Spring 2022: Final presentation

Andrej Leban



OVERVIEW

A stochastic process

All variables under consideration are, in principle, vectors:

- **State random variable**: $X = X_t$.
- **Parameter (non-random) variable**: t, usually denoting time.

A STOCHASTIC PROCESS

All variables under consideration are, in principle, vectors:

- **State random variable**: $X = X_t$.
- **Parameter (non-random) variable**: t, usually denoting time.

Classification:

State space	***Discrete***	Discrete parameter chain	Continuous parameter chain
	Continuous	Random sequence (topic of this class)	Stochastic process
		Discrete	***Continuous***
		Parameter Set	

Stochastic processes

- Defined by the Probability law: Full joint distribution function $F_{X_{t_0},\ldots:\ \forall t \geq t_0}$ / full joint density function $f_{X_{t_0},\ldots:\ \forall t \geq t_0}$ / full joint characteristic function $\varphi_{X_{t_0},\ldots:\ \forall t \geq t_0}$
- Difficult to express in general: for *Markov* and *Gaussian* process the first-order: $f_{X_{t}:\ t \geq t_0}$ and the second-order: $f_{X_{t,\tau}:\ t,\tau \geq t_0}$ *marginals* are enough to determine the probability law.

Stochastic processes

- Defined by the PROBABILITY LAW: Full joint distribution function $F_{X_{t_0},\dots:\ \forall t \geq t_0}$ / full joint density function $f_{X_{t_0},\dots:\ \forall t \geq t_0}$ / full joint characteristic function $\varphi_{X_{t_0},\dots:\ \forall t \geq t_0}$
- Difficult to express in general: for *Markov* and *Gaussian* process the first-order: $f_{X_{t:\ t \geq t_0}}$ and the second-order: $f_{X_{t,\tau:\ t,\tau \geq t_0}}$ *marginals* are enough to determine the probability law.

Some important *statistics* functions:

- The mean value function: $m_X(t) = \mathbb{E}[X_t]\ (t)$
- The (auto) correlation function: $\gamma_X(t,\tau) = \mathbb{E}[X_t\ X_\tau]\ (t)$
- The (auto) covariance function: $c_X(t,\tau) = \mathbb{E}[(X_t - m_X(t))\ (X_\tau - m_X(\tau))]\ (t)$

STOCHASTIC PROCESSES

- Defined by the PROBABILITY LAW: Full joint distribution function $F_{X_{t_0},\dots:\ \forall t\geq t_0}$ / full joint density function $f_{X_{t_0},\dots:\ \forall t\geq t_0}$ / full joint characteristic function $\varphi_{X_{t_0},\dots:\ \forall t\geq t_0}$
- Difficult to express in general: for *Markov* and *Gaussian* process the first-order: $f_{X_{t:\ t\geq t_0}}$ and the second-order: $f_{X_{t,\tau:\ t,\tau\geq t_0}}$ *marginals* are enough to determine the probability law.

Some important *statistics* functions:

- The mean value function: $m_X(t) = \mathbb{E}[X_t]\ (t)$
- The (auto) correlation function: $\gamma_X(t,\tau) = \mathbb{E}[X_t\ X_\tau]\ (t)$
- The (auto) covariance function: $c_X(t,\tau) = \mathbb{E}[(X_t - m_X(t))\ (X_\tau - m_X(\tau))]\ (t)$

Stationarity:

- **Weak**: $m_X(t) = const.\qquad c_X(t, t+\tau) = const.;\ \forall\tau$
- **Strong**: $f_{X_{t_0,\dots,t_n}} = f_{X_{t_0+\tau,\dots,t_n+\tau}};\ \forall\tau\qquad$ (*n - order*)

Mean-square calculus

Limit in mean-square:

$$\text{l.i.m. } x_n = x \Leftrightarrow \lim_{n \to \infty} \mathbb{E}[|x - x_n|^2] = 0$$

Mean-square calculus

Limit in mean-square:

$$\text{l.i.m. } x_n = x \Leftrightarrow \lim_{n\to\infty} \mathbb{E}[|x - x_n|^2] = 0$$

Re-defining all the usual operations of calculus in the mean-square sense, we get very useful properties:

- $m_{\dot{X}}(t) = \dot{m}_X(t) \qquad \mathbb{E}[\int_a^b X_t dt] = \int_a^b m_X(t) dt$
- $\gamma_{\dot{X}\dot{X}}(t,\tau) = \frac{\partial \gamma(t,\tau)}{\partial t \partial \tau^T} \qquad \mathbb{E}[\int_a^b X_t dt \int_c^d X_\tau d\tau] = \int_a^b \int_c^d \gamma_{X,X}(t,\tau)\, dt\, d\tau$
- $c_{\dot{X}\dot{X}}(t,\tau) = \frac{\partial c(t,\tau)}{\partial t \partial \tau^T} \qquad \text{Cov}(\int_a^b X_t dt, \int_c^d X_\tau d\tau) = \int_a^b \int_c^d c_{X,X}(t,\tau)\, dt\, d\tau$

and the fundamental theorem of (mean-square) calculus:

$$X_t - X_a = \int_a^t X_\tau \, d\tau$$

The Brownian motion and White noise processes

The Brownian Motion (Wiener - Levy) process - β_t:

- $X_t \sim N(0, C(t));\ \forall t$
- $\{X_t\}$ has stationary and independent **increments**:
  $X_t - X_\tau \overset{D}{=} X_{t+h} - X_{\tau+h}; \qquad \forall h, \forall t > \tau$
- $X_t - X_\tau \sim N(0,\ \sigma^2(t-\tau)); \qquad \forall \tau, \forall t > \tau$

THE BROWNIAN MOTION AND WHITE NOISE PROCESSES

The Brownian Motion (Wiener - Levy) process - β_t:

- $X_t \sim N(0, C(t));\ \forall t$
- $\{X_t\}$ has stationary and independent **increments**:
  $X_t - X_\tau \stackrel{D}{=} X_{t+h} - X_{\tau+h}; \qquad \forall h, \forall t > \tau$
- $X_t - X_\tau \sim N(0,\ \sigma^2(t-\tau)); \qquad \forall \tau, \forall t > \tau$

Is a *Gaussian* and a *Markov* stochastic process.

The Brownian motion and White noise processes

The Brownian Motion (Wiener - Levy) process - β_t:

- $X_t \sim N(0, C(t));\ \forall t$
- $\{X_t\}$ has stationary and independent **increments**:
  $X_t - X_\tau \overset{D}{=} X_{t+h} - X_{\tau+h}; \qquad \forall h, \forall t > \tau$
- $X_t - X_\tau \sim N(0,\ \sigma^2(t-\tau)); \qquad \forall \tau, \forall t > \tau$

Is a *Gaussian* and a *Markov* stochastic process.

The White noise process:

- $\{X_t\}$ is mutually independent with all other states: $X_t \perp\!\!\!\perp X_\tau; \qquad \forall t, \tau$
- The power spectral density of the correlation function is constant, hence the name.
- For a white *Gaussian* process: $C_{X,X}(t,\ t+\tau) = Q(t)\ \delta(t-\tau)$

THE BROWNIAN MOTION AND WHITE NOISE PROCESSES

The Brownian Motion (Wiener - Levy) process - β_t:

- $X_t \sim N(0, C(t));\ \forall t$
- $\{X_t\}$ has stationary and independent **increments**:
  $X_t - X_\tau \stackrel{D}{=} X_{t+h} - X_{\tau+h}; \qquad \forall h, \forall t > \tau$
- $X_t - X_\tau \sim N(0,\ \sigma^2(t-\tau)); \qquad \forall \tau, \forall t > \tau$

Is a *Gaussian* and a *Markov* stochastic process.

The White noise process:

- $\{X_t\}$ is mutually independent with all other states: $X_t \perp\!\!\!\perp X_\tau; \qquad \forall t, \tau$
- The power spectral density of the correlation function is constant, hence the name.
- For a white *Gaussian* process: $C_{X,X}(t,\ t+\tau) = Q(t)\ \delta(t-\tau)$

The latter, together with the rules of mean-square calculus, gives:

$$w_t = d\beta_t$$

THE SDE

In general, for a *random*, not necessarily linear function f

$$\dot{X}_t = f(x_t, w_t, t) \;\Leftrightarrow\; X_t - X_{t_0} = \int_{t_0}^{t} f(x_\tau, w_\tau, \tau)\, d\tau, \tag{1}$$

where w_t is itself a random function - the "forcing", "input" term.

THE SDE

In general, for a *random*, not necessarily linear function f

$$\dot{X}_t = f(x_t, w_t, t) \;\Leftrightarrow\; X_t - X_{t_0} = \int_{t_0}^{t} f(x_\tau, w_\tau, \tau)\, d\tau, \qquad (1)$$

where w_t is itself a random function - the "forcing", "input" term.

We will restrict ourselves to the separable form - the *Langevin equation*:

$$\dot{X}_t = f(x_t, t) + G(w_t, t)\, w_t \;\Leftrightarrow\; f(x_t, t) + G(w_t, t)\, d\beta_t, \qquad (2)$$

where w_t is a Gaussian white noise. $\mathbb{E}[d\beta_t\, d\beta_t^T] = Q(t)$.

The SDE

In general, for a *random*, not necessarily linear function f

$$\dot{X}_t = f(x_t, w_t, t) \;\Leftrightarrow\; X_t - X_{t_0} = \int_{t_0}^{t} f(x_\tau, w_\tau, \tau)\, d\tau, \tag{1}$$

where w_t is itself a random function - the "forcing", "input" term.

We will restrict ourselves to the separable form - the *Langevin equation*:

$$\dot{X}_t = f(x_t, t) + G(w_t, t)\, w_t \;\Leftrightarrow\; f(x_t, t) + G(w_t, t)\, d\beta_t, \tag{2}$$

where w_t is a Gaussian white noise. $\mathbb{E}[d\beta_t \, d\beta_t^T] = Q(t)$.

The SDE

In general, for a *random*, not necessarily linear function f

$$\dot{X}_t = f(x_t, w_t, t) \;\Leftrightarrow\; X_t - X_{t_0} = \int_{t_0}^{t} f(x_\tau, w_\tau, \tau)\, d\tau, \tag{1}$$

where w_t is itself a random function - the "forcing", "input" term.

We will restrict ourselves to the separable form - the *Langevin equation*:

$$\dot{X}_t = f(x_t, t) + G(w_t, t)\, w_t \;\Leftrightarrow\; f(x_t, t) + G(w_t, t)\, d\beta_t, \tag{2}$$

where w_t is a Gaussian white noise. $\mathbb{E}[d\beta_t \, d\beta_t^T] = Q(t)$.

What is $d\beta_t$? We can side-step this question with the fundamental theorem:

$$X_t - X_{t_0} = \underbrace{\int_{t_0}^{t} f(x_\tau, \tau)\, d\tau}_{\text{Mean-square Riemann integral}} + \underbrace{\int_{t_0}^{t} G(x_\tau, \tau)\, d\beta_\tau}_{\text{Ito integral}} \tag{3}$$

The Ito integral: intuition and definition

The increments are Markov by the property of the Brownian motion:

$$(X_{t+\delta t} - X_t)|X_t \propto (\beta_{t+\delta t} - \beta_t)$$

The Ito integral: intuition and definition

The increments are Markov by the property of the Brownian motion:

$$(X_{t+\delta t} - X_t)|X_t \propto (\beta_{t+\delta t} - \beta_t)$$

For illustration: say $X_t \equiv e^{\beta_t}$:

$$\delta X_t \approx e^{\beta_t + \delta\beta_t} - e^{\beta_t} \approx X_t(\delta\beta_t + \frac{1}{2}\delta\beta_t^2 + \dots)$$

$$\mathbb{E}[\delta X_t - X_t \delta\beta_t] = \mathcal{O}(\delta t) \qquad \text{(if using only 1st order!)}$$

$$\implies dX_t = X_t d\beta_t + \frac{1}{2} X_t d\beta_t^2$$

$$\implies X_t - 1 = \int_0^t X_t \, d\beta_\tau + \frac{1}{2}\int_0^t X_t \, d\beta_\tau^2$$

THE ITO INTEGRAL: INTUITION AND DEFINITION

The increments are Markov by the property of the Brownian motion:

$$(X_{t+\delta t} - X_t)|X_t \propto (\beta_{t+\delta t} - \beta_t)$$

For illustration: say $X_t \equiv e^{\beta_t}$:

$$\delta X_t \approx e^{\beta_t + \delta\beta_t} - e^{\beta_t} \approx X_t(\delta\beta_t + \frac{1}{2}\delta\beta_t^2 + \dots)$$

$$\mathbb{E}[\delta X_t - X_t \delta\beta_t] = \mathcal{O}(\delta t) \qquad \text{(if using only 1st order!)}$$

$$\implies dX_t = X_t d\beta_t + \frac{1}{2} X_t d\beta_t^2$$

$$\implies X_t - 1 = \int_0^t X_t \, d\beta_\tau + \frac{1}{2}\int_0^t X_t \, d\beta_\tau^2$$

FIRST- and SECOND order Ito stochastic integrals for the Brownian motion:
For a random function: $g_t(\omega) \perp\!\!\!\perp (\beta_t - \beta_\tau)$, $\int_T \mathbb{E}[|g_t(\omega)|^2]dt < \infty$:

$$\int_0^t g_t(\omega)\, d\beta_\tau = \text{l.i.m.}_{\rho\to 0} \sum_i g_t(\omega)(\beta_{t_{i+1}} - \beta_{t_i}) \tag{4}$$

$$\int_0^t g_t(\omega)\, d\beta_\tau^2 = \text{l.i.m.}_{\rho\to 0} \sum_i g_t(\omega)(\beta_{t_{i+1}} - \beta_{t_i})^2 = \sigma^2 \int_0^t g_t(\omega)\, dt, \tag{5}$$

where ρ is the maximum sequential distance on the partition.

THE ITO STOCHASTIC DIFFERENTIAL

For an arbitrary (nice enough) function of X_t φ, its *stochastic differential* is:

$$d\varphi = \frac{\partial\varphi}{\partial x}dt + \frac{\partial\varphi}{\partial x^T}d_{Xt} + \frac{1}{2}tr\left(G(t)Q(t)G(t)^T\frac{\partial^2\varphi}{\partial x\partial x^T}\right)dt \tag{6}$$

A solution for the Ito integral of a given function $\psi = \frac{\partial\varphi}{\partial x}$ can thus be obtained from the fundamental theorem:

$$\int_a^b \psi(\beta_t)d\beta_t = \varphi(\beta_b) - \varphi(\beta_a) - \frac{\sigma^2}{2}\int_a^b \frac{\partial^2\varphi}{\partial x\partial x^T}dt$$

KOLMOGOROV'S EQUATION

Recap: for the Brownian motion, the *marginal* and the *transition probability* are the *probability law*.

Kolmogorov's equation

Recap: for the Brownian motion, the *marginal* and the *transition probability* are the *probability law*.

For the (Langevin) Ito SDE:

$$dX_t = f(x_t, t)\, dt + g(x_t, t)\, d\beta_t$$

with non-random functions f, g, A. Kolmogorov has derived a PDE for the evolution of the *transition probability*:

$$\frac{\partial p(X_t|X_\tau)}{\partial t} = \frac{\partial (p(X_t|X_\tau)\, f(x,t))}{\partial x} + \frac{1}{2}\frac{\partial^2 (p(X_t|X_\tau)\, g^2(x,t))}{\partial x^2}, \qquad \forall t > \tau \tag{7}$$

In the context of Physics this diffusion equation is called the Fokker-Planck equation. It can be formerly encapsulated by a *diffusion operator* $\mathcal{L}(p)$.

Initial condition: $\lim_{t\to\tau} p_{X_t|X_\tau}(x_t|x_\tau) = \delta(x_t - x_\tau)$.
Boundary conditions: $p_{X_t|X_\tau}(\pm\infty \mid x_\tau) = 0$.

THE GENERAL FILTERING PROBLEM

We can now bring observations into the picture as another *Langevin* equation:

$$y_t = h(x_t, t) + v_t \;\Leftrightarrow dz_t = h(x_t, t) + d\eta_t, \tag{8}$$

where v_t is another (independent) white-noise process, and $d\eta_t$ a Brownian motion: $\mathbb{E}[d\eta_t d\eta_t^T] = R(t)$.

The general filtering problem

We can now bring observations into the picture as another *Langevin* equation:

$$y_t = h(x_t, t) + v_t \;\Leftrightarrow dz_t = h(x_t, t) + d\eta_t, \tag{8}$$

where v_t is another (independent) white-noise process, and $d\eta_t$ a Brownian motion: $\mathbb{E}[d\eta_t d\eta_t^T] = R(t)$.

THE GENERAL FILTERING PROBLEM

We can now bring observations into the picture as another *Langevin* equation:

$$y_t = h(x_t, t) + v_t \;\Leftrightarrow dz_t = h(x_t, t) + d\eta_t, \tag{8}$$

where v_t is another (independent) white-noise process, and $d\eta_t$ a Brownian motion: $\mathbb{E}[d\eta_t d\eta_t^T] = R(t)$.

Given a suitable prior p_{X_0}, the *conditional density* $p_{X_t|Y_{0:t}}$ is the complete solution of the filtering problem.

The general filtering problem

We can now bring observations into the picture as another *Langevin* equation:

$$y_t = h(x_t, t) + v_t \;\Leftrightarrow dz_t = h(x_t, t) + d\eta_t, \tag{8}$$

where v_t is another (independent) white-noise process, and $d\eta_t$ a Brownian motion: $\mathbb{E}[d\eta_t d\eta_t^T] = R(t)$.

Given a suitable prior p_{X_0}, the *conditional density* $p_{X_t|Y_{0:t}}$ is the complete solution of the filtering problem.

Goal:

- Estimate the conditional (posterior) mean: $\hat{x}_t = \mathbb{E}[X_t|Y_{0:t}]$
- This is the *optimal* (minimum variance) estimate for $\mathbb{E}[L(x_t - \hat{x}_t)|Y_{0:t}]$ for a class of *loss* functions $L(x - \hat{x})$

The effect of discrete observations - between observations

Between observations, the density evolution must obey Kolmogorov's equation $\mathcal{L}(p_{X_t|Y_{0:t}})$.

The effect of discrete observations - between observations

Between observations, the density evolution must obey Kolmogorov's equation $\mathcal{L}(p_{X_t|Y_{0:t}})$.

We can use the properties of the stochastic differential of a (nice enough) function φ (6) to adapt it to the *conditional mean* $\hat{x}_t$ and the *conditional covariance*:

$$\hat{P}_t^\tau = Cov(X_t,\ X_t|Y_{0:\tau}),\ (\tau = t \text{ for filtering problems})$$

In case the filtering problem is **linear** with the Brownian process, these two **uniquely determine** the state.

THE EFFECT OF DISCRETE OBSERVATIONS - BETWEEN OBSERVATIONS

Between observations, the density evolution must obey Kolmogorov's equation $\mathcal{L}(p_{X_t|Y_{0:t}})$.

We can use the properties of the stochastic differential of a (nice enough) function φ (6) to adapt it to the *conditional mean* $\hat{x_t}$ and the *conditional covariance*:

$$\hat{P}_t^\tau = Cov(X_t,\ X_t|Y_{0:\tau}),\ (\tau = t \text{ for filtering problems})$$

In case the filtering problem is **linear** with the Brownian process, these two **uniquely determine** the state.

Define the *conditional expectation operator* for a function φ:

$$\hat{\varphi}^\tau(X_t) = \mathbb{E}_\tau[\varphi(X_t)|Y_{0:\tau}] = \int \varphi(x_t)\, p_{X_t|Y_{0:\tau}}(x_t)\, dx_t \tag{9}$$

THE EFFECT OF DISCRETE OBSERVATIONS - BETWEEN OBSERVATIONS

Between observations, the density evolution must obey Kolmogorov's equation $\mathcal{L}(p_{X_t|Y_{0:t}})$.

We can use the properties of the stochastic differential of a (nice enough) function φ (6) to adapt it to the *conditional mean* $\hat{x_t}$ and the *conditional covariance*:

$$\hat{P}_t^\tau = Cov(X_t,\ X_t|Y_{0:\tau}),\ (\tau = t \text{ for filtering problems})$$

In case the filtering problem is **linear** with the Brownian process, these two **uniquely determine** the state.

Define the *conditional expectation operator* for a function φ:

$$\hat{\varphi}^\tau(X_t) = \mathbb{E}_\tau[\varphi(X_t)|Y_{0:\tau}] = \int \varphi(x_t)\, p_{X_t|Y_{0:\tau}}(x_t)\, dx_t \tag{9}$$

The Kolmogorov equation for these two becomes:

$$\dot{\hat{x}}_t^t = \widehat{f(x_t,t)}^t$$
$$\dot{\hat{P}}_t^t = \left[\widehat{X_t f(x_t,t)^T}^t - \widehat{X_t}^t \widehat{f(x_t,t)^T}^t\right] + \left[\widehat{f(x_t,t)X_t^T}^t - \widehat{f(x_t,t)}^t (\widehat{X_t}^t)^T\right] + \widehat{G(t)Q(t)G(t)^T}^t$$

THE EFFECT OF DISCRETE OBSERVATIONS - AT THE OBSERVATIONS

We wish to determine the relation between $p_{X_t|Y_{0:t}^{-}} = p_{X_t|Y_{0:t_{k-1}}}$ and $p_{X_t|Y_{0:t}}$, i.e. what happens at the observation at the time t_k.

THE EFFECT OF DISCRETE OBSERVATIONS - AT THE OBSERVATIONS

We wish to determine the relation between $p_{X_t|Y_{0:t}^{-}} = p_{X_t|Y_{0:t_{k-1}}}$ and $p_{X_t|Y_{0:t}}$, i.e. what happens at the observation at the time t_k.

The latter must satisfy Bayes' rule:

$$p_{X_{t_k}|Y_{0:t_k}} = \frac{p_{Y_{t_k}|X_{t_k},Y_{0:t_{k-1}}}\, p_{X_{t_k}|Y_{0:t_{k-1}}}}{p_{Y_{t_k}|Y_{0:t_{k-1}}}}$$

THE EFFECT OF DISCRETE OBSERVATIONS - AT THE OBSERVATIONS

We wish to determine the relation between $p_{X_t|Y^-_{0:t}} = p_{X_t|Y_{0:t_{k-1}}}$ and $p_{X_t|Y_{0:t}}$, i.e. what happens at the observation at the time t_k.

The latter must satisfy Bayes' rule:

$$p_{X_{t_k}|Y_{0:t_k}} = \frac{p_{Y_{t_k}|X_{t_k},Y_{0:t_{k-1}}}\; p_{X_{t_k}|Y_{0:t_{k-1}}}}{p_{Y_{t_k}|Y_{0:t_{k-1}}}}$$

For a white noise process, we recover the familiar **general filtering update rule**:

$$p_{X_{t_k}|Y_{0:t_k}} = \frac{p_{Y_{t_k}|X_{t_k}}\; p_{X_{t_k}|Y^-_{0:t_k}}}{\int p_{Y_{t_k}|\xi_{t_k}}\; p_{\xi_{t_k}|Y^-_{0:t_k}}\; d\xi} \tag{10}$$

For continuous observations $y(t)$, $Y^-_{0:t}$ signifies the *left limit*.

CONTINUOUS OBSERVATIONS: THE KUSHNER EQUATION

Problem: there is no "time between observations", so Kolmogorov's equation needs to be modified to account for the observations.

CONTINUOUS OBSERVATIONS: THE KUSHNER EQUATION

Problem: there is no "time between observations", so Kolmogorov's equation needs to be modified to account for the observations.

The differential equation for $p_{X_t|Y_{0:t}}$ with continuous observations obeys the *Kushner equation*:

$$dp = \mathcal{L}(p) + (h_t - \hat{h}_t^t)^T R^{-1}(t)\,(dz_t - \hat{h}_t^t\,dt), \tag{11}$$

where $h_t = h(X_t, t)$ is the "forcing" term from the Langevin equation for the **observations**, and dz are the continuous observations.

The general evolution of moments

As with the Kolmogorov equation, the Kushner equation can also be adapted for the conditional expectation of (nice enough) functions of X_t. Thus we obtain for the first two moments:

$$d\hat{x}_t = \widehat{f_t}^t dt + \left[(\widehat{X_t h_t}^t)^T - \widehat{X_t}^t \widehat{h}_t^t \right] R^{-1}(t) \left[dz_t - \widehat{h}_t^t \, dt \right]$$

$$d(\hat{P}_t)_{ij} = \left[\widehat{(X_t)_i (f_t)_j}^t - \widehat{(X_t)_i}^t \widehat{(f_t)_j}^t \right] + \left[\widehat{(X_t)_j (f_t)_i}^t - \widehat{(X_t)_j}^t \widehat{(f_t)_i}^t \right] + \widehat{(G(t)Q(t)G(t)^T)_{ij}}^t$$

$$- \left[\widehat{(X_t)_i (h_t)}^t - \widehat{(X_t)_i}^t \widehat{(h)_j}^t \right] R^{-1}(t) \left[\widehat{(X_t)_j (h_t)}^t - \widehat{(X_t)_j}^t \widehat{(h)_j}^t \right]$$

$$+ \left[\widehat{(X_t)_i (X_t)_j (h_t)}^t - \widehat{(X_t)_i (X_t)_j}^t \widehat{(h)_j}^t - \widehat{(X_t)_i}^t \widehat{(X_t)_j^t (h)_j}^t - \widehat{(X_t)_j}^t \widehat{(X_t)_i^t (h)_j}^t + 2 \widehat{(X_t)_i}^t \widehat{(X_t)_j}^t \widehat{(h)_j}^t \right]$$

$$R^{-1}(t) \left[dz_t - \widehat{h}_t^t \, dt \right]$$

LINEAR GAUSSIAN FILTERING PROBLEM

Given the complexity of the above, we now limit ourselves to **linear, Gaussian white noise** filtering problems:

$$
\begin{aligned}
dX_t &= F(t)\, X_t\, dt + G(t)\, d\beta_t \\
dY_t &= M(t)\, X_t\, dt + d\eta_t
\end{aligned}
$$

LINEAR GAUSSIAN FILTERING PROBLEM

Given the complexity of the above, we now limit ourselves to **linear, Gaussian white noise** filtering problems:

$$dX_t = F(t)\, X_t\, dt + G(t)\, d\beta_t$$
$$dY_t = M(t)\, X_t\, dt + d\eta_t$$

- The state and time dynamics are separated.

Linear Gaussian filtering problem

Given the complexity of the above, we now limit ourselves to **linear, Gaussian white noise** filtering problems:

$$dX_t = F(t)\, X_t\, dt + G(t)\, d\beta_t$$
$$dY_t = M(t)\, X_t\, dt + d\eta_t$$

- The state and time dynamics are separated.
- Thus, the conditional covariance P is no longer a function of the state: $\hat{P}_t^t = P_t^t$.

Recap: The discrete-discrete Kalman filter

$$X_{t+1} = F(t)\, X_t + G(t)\, w_{t+1}$$
$$y_{t+1} = M(t)\, X_t + v_t$$

Recap: The discrete-discrete Kalman filter

$$X_{t+1} = F(t)\, X_t + G(t)\, w_{t+1}$$
$$y_{t+1} = M(t)\, X_t + v_t$$

Since the parameter space t is discrete, Kolmogorov's equation (7) simply decomposes into difference equations. For the moments, this gives:

$$\hat{x}^t_{t+1} = F(t)\, \hat{x}^t_t$$
$$P^t_{t+1} = F(t) P^t_t F(t)^T + G(t) Q(t+1) G(t)^T,$$

which are the familiar *one-step ahead prediction* relations.

RECAP: THE DISCRETE-DISCRETE KALMAN FILTER

$$X_{t+1} = F(t)\, X_t + G(t)\, w_{t+1}$$
$$y_{t+1} = M(t)\, X_t + v_t$$

Since the parameter space t is discrete, Kolmogorov's equation (7) simply decomposes into difference equations. For the moments, this gives:

$$\hat{x}^t_{t+1} = F(t)\, \hat{x}^t_t$$
$$P^t_{t+1} = F(t)P^t_t F(t)^T + G(t)Q(t+1)G(t)^T,$$

which are the familiar *one-step ahead prediction* relations.
At the observations, we proceed using the conjugacy of the Gaussians to simplify (10).
If we define the *Kalman Gain* as:

$$K(t) = P^{t-}_t M^T(t) \left[M^T(t) P^{t-}_t M^T(t) + R(t) \right]^{-1}$$

We recover the familiar *filtering update* relations:

$$\hat{x}^t_t = \hat{x}^{t-1}_t + K(t) \left[y_t - M(t)\hat{x}^{t-1}_t \right]$$
$$P^t_t = P^{t-1}_t - K(t)M(t)P^{t-1}_t$$

The continuous-discrete Kalman filter

$$\dot{X}_t = F(t)\, X_t\, dt + G(t)\, d\beta_t$$
$$y_{t_k} = M(t_k)\, X_{t_k} + v_k$$

Note: we have the marginal $p_{X_t|Y_{0:t}}$ in closed form; one could simply evolve the state using Kolmogorov's equation.

The continuous-discrete Kalman filter

$$\dot{X}_t = F(t)\, X_t\, dt + G(t)\, d\beta_t$$
$$y_{t_k} = M(t_k)\, X_{t_k} + v_k$$

Note: we have the marginal $p_{X_t|Y_{0:t}}$ in closed form; one could simply evolve the state using Kolmogorov's equation.
The Kolmogorov equation for the evolution of moments between observations simplifies to:

$$\dot{\hat{x}}_t^t = F(t)\, \hat{x}_t^t$$
$$\dot{P}_t^t = F(t)P_t^t + P_t^t F(t)^T + G(t)Q(t)G(t)^T$$

THE CONTINUOUS-DISCRETE KALMAN FILTER

$$\dot{X}_t = F(t)\, X_t\, dt + G(t)\, d\beta_t$$
$$y_{t_k} = M(t_k)\, X_{t_k} + v_k$$

Note: we have the marginal $p_{X_t|Y_{0:t}}$ in closed form; one could simply evolve the state using Kolmogorov's equation.
The Kolmogorov equation for the evolution of moments between observations simplifies to:

$$\dot{\hat{x}}_t^t = F(t)\, \hat{x}_t^t$$
$$\dot{P}_t^t = F(t)P_t^t + P_t^t F(t)^T + G(t)Q(t)G(t)^T$$

Since the observations are still discrete, we simply replace the previous values of the states with the left limits in the preceding version:

$$\hat{x}_{t_k}^{t_k^+} = \hat{x}_{t_k}^{t_k^-} + K(t_k)\left[y_{t_k} - M(t_k)\hat{x}_{t_k}^{t_k^-}\right]$$
$$P_t^{t_k^+} = P_{t_k}^{t_k^-} - K(t_k)M(t_k)P_t^{t_k^-}$$

This filter can be reformulated as a discrete filter by integrating over all intervals $[t_k, t_{k+1}]$ and using the *state transition matrix* $\Phi(t_{k+1}, t_k)$.

THE CONTINUOUS-CONTINUOUS KALMAN FILTER: THE KALMAN-BUCY FILTER

$$\dot{X}_t = F(t)\, X_t\, dt + G(t)\, d\beta_t$$
$$\dot{z}_t = M(t)\, X_t\, dt + d\eta_t$$

The continuous-continuous Kalman filter: the Kalman-Bucy filter

$$\dot{X}_t = F(t)\, X_t\, dt + G(t)\, d\beta_t$$
$$\dot{z}_t = M(t)\, X_t\, dt + d\eta_t$$

Here, we have to adapt the *Kushner equation* for the moments. Fortunately, the separability $f(X_t, t) = F(t)\, X_t \ldots$ significantly simplifies the terms of the type:

$$(\widehat{X_t f^T}^t - \widehat{X_t}^t \widehat{f^T}^t) = (\widehat{X_t X_t^T}^t - \widehat{X_t}^t \widehat{X_t^T}^t)\, F(t)^T \text{ etc.}$$

THE CONTINUOUS-CONTINUOUS KALMAN FILTER: THE KALMAN-BUCY FILTER

$$\dot{X}_t = F(t)\, X_t\, dt + G(t)\, d\beta_t$$
$$\dot{z}_t = M(t)\, X_t\, dt + d\eta_t$$

Here, we have to adapt the *Kushner equation* for the moments. Fortunately, the separability $f(X_t, t) = F(t)\, X_t \ldots$ significantly simplifies the terms of the type:

$$(\widehat{X_t f^T}^t - \widehat{X_t}^t \widehat{f^T}^t) = (\widehat{X_t X_t^T}^t - \widehat{X_t}^t \widehat{X_t^T}^t)\, F(t)^T \text{ etc.}$$

By defining the *Kalman gain* in this instance as:

$$K(t) = P_t^t M(t)^T R(t)^{-1},$$

we get:

$$d\hat{x}_t^t = F(t)\, \hat{x}_t^t\, dt + K(t)\left[dz_t - M(t)\, \hat{x}_t^t\, dt\right]$$
$$\dot{P}_t^t = F(t)\, P_t^t + P_t^t\, F(t)^T + G(t)\, Q(t)\, G(t)^T - K(t)\, M(t)\, P_t^t,$$

where dz_t is the (continuously) observed value.

The continuous-continuous Kalman filter: the Kalman-Bucy filter

$$\dot{X}_t = F(t)\, X_t\, dt + G(t)\, d\beta_t$$
$$\dot{z}_t = M(t)\, X_t\, dt + d\eta_t$$

Here, we have to adapt the *Kushner equation* for the moments. Fortunately, the separability $f(X_t, t) = F(t)\, X_t \ldots$ significantly simplifies the terms of the type:

$$(\widehat{X_t f^T}^t - \widehat{X_t}^t \widehat{f^T}^t) = (\widehat{X_t X_t^T}^t - \widehat{X_t}^t \widehat{X_t^T}^t)\, F(t)^T \text{ etc.}$$

By defining the *Kalman gain* in this instance as:

$$K(t) = P_t^t M(t)^T R(t)^{-1},$$

we get:

$$d\hat{x}_t^t = F(t)\, \hat{x}_t^t\, dt + K(t)\, [dz_t - M(t)\, \hat{x}_t^t\, dt]$$
$$\dot{P}_t^t = F(t)\, P_t^t + P_t^t\, F(t)^T + G(t)\, Q(t)\, G(t)^T - K(t)\, M(t)\, P_t^t,$$

where dz_t is the (continuously) observed value.

THE KALMAN-BUCY FILTER - SOME PROPERTIES

- There is no more separation between the evolution between- and the jumps at - the observations: we get one equation.

THE KALMAN-BUCY FILTER - SOME PROPERTIES

- There is no more separation between the evolution between- and the jumps at - the observations: we get one equation.
- The Kalman gain leverages the influence of the residual on the state.

THE KALMAN-BUCY FILTER - SOME PROPERTIES

- There is no more separation between the evolution between- and the jumps at - the observations: we get one equation.
- The Kalman gain leverages the influence of the residual on the state.
- The equation for the evolution of the conditional variance is known as the *Riccati equation*.

CONCLUSION

- Dynamic programming / Reinforcement learning

Conclusion

- Dynamic programming / Reinforcement learning
- Statistical mechanics

CONCLUSION

- Dynamic programming / Reinforcement learning
- Statistical mechanics
- Dynamical systems

References I

[1] Alazard, D. Introduction to kalman filtering. https://personnel.isae-supaero.fr/IMG/pdf/introkalman_e_08072019.pdf.

[2] Durbin, J., and Koopman, S. J. *Time series analysis by State Space Methods*. Oxford University Press, 2012.

[3] GELB, A. *Applied Optimal estimation. written by technical staff, the Analytic Sciences Corporation. edited by Arthur Gelb, etc*. M.I.T. Press, 1974.

[4] Jazwinski, A. H. *Stochastic processes and filtering theory*. Academic Press, 1970.

[5] Joshi, M. S. *The concepts and practice of Mathematical Finance*. Cambridge University Press, 2011.

[6] Kitagawa, G., and Gersh, W. *Smoothness priors analysis of time series*. Springer, 1996.

[7] Lewis, F. L., Popa, D., and Xie, L. *Optimal and robust estimation: With an introduction to stochastic control theory*. CRC Press, 2008.

[8] Triantafyllopoulos, K. *Bayesian inference of State Space Models: Kalman filtering and beyond*. Springer, 2021.