Stochastic Differential Equations: The Kalman-Bucy Filter

Nora Røhnebæk Aasen

November 22, 2021

Abstract

In many real life applications of stochastic differential equations, we will not have direct observations of our process. The Kalman-Bucy filter is an estimator for when we have continuous observations of a linear system. This paper mainly focuses on the theoretical justification for why this filter works. It also includes one example for how this filter can be used for parameter estimation.

1 Introduction

Suppose we have a linear stochastic differential equation given by

$$dX_t = A(t)X_t dt + \alpha(t)dB_t, \tag{1}$$

where $A, \alpha : \mathbb{R} \to \mathbb{R}$ and B_t is a 1-dimensional Brownian motion[1]. If we assume that there exists a solution to this system, then, using Itô's Formula, we can show that the solution will be on the form

$$X_t = \exp\left\{\int_0^t A(s)ds\right\} X_0 + \int_0^t \exp\left\{\int_s^t A(u)du\right\} \alpha(s)dB_s. \tag{2}$$

See appendix A for a detailed calculation of this solution.

Assume now that there exists a linear system as described above, but we are not able to measure it directly. Instead, we are only able to obtain inaccurate observations of our system at time t, denoted Z_t. We then get two stochastic differential equations,

$$dX_t = A(t)X_t dt + \alpha(t)dB_t \tag{3}$$

$$dZ_t = G(t)X_t dt + \gamma(t)d\tilde{B}_t. \tag{4}$$

Hence, the filtering problem can be formulated as follows:

What is the best estimate of the state of our system X_t,
based only on the observations Z_t?

First of all, how do we rephrase this question in mathematical terms? Note that we have a underlying probability space $(\Omega, \mathcal{F}, P)$ which corresponds to the Brownian motion $[B_t, \tilde{B}_t]$, and expectation E is defined with respect to the probability measure P. We want to find the best estimate, $\hat{X}_t$, of a state of the system X_t, based on the observations of Z_t. The set of observations $\{Z_s, s \leq t\}$ will generate a sub σ-algebra, $\mathcal{G}_t \subset \mathcal{F}$, so when we say we want the best estimate based on Z_t, what we want is to find $\hat{X}_t$ such that

- $\hat{X}_t$ is $\mathcal{G}_t$-measurable,
- $E[(X_t - \hat{X}_t)^2] \leq E[(X_t - Y)^2]$, where Y is any $\mathcal{G}_t$-measurable random variable in $L^2(\Omega, P)$.

It turns out that $\hat{X}_t = E[X_t|\mathcal{G}_t]$, but why this holds and how to find the solution is what will be discussed in the upcoming sections. This will eventually result in *the Kalman-Bucy filter*. This paper is mainly based off of chapter six in *Stochastic Differential Equations* by Bernt Øksendal [1]. Most claims will be stated without proof, as they can be retraced here.

2 The Kalman-Bucy Filter

Let Z_t, the observations of the system X_t, be such that the equations (3) and (4) are satisfied. Additionally, we assume that A, α, G and γ are bounded on bounded intervals, in order to ensure existence and uniqueness of the solution. Furthermore, we assume that $Z_0 = 0$, that the range of

[1] It is not necessary to restrict ourselves to one dimension, but we avoid the technical details that arise in the several-dimensional case since the concepts remain the same.

γ is sufficiently far from 0 on bounded intervals, and that X_0 has a Gaussian distribution which is independent from dB_t and $d\tilde{B}_t$.

In the introduction we said that if Z_t generates a σ-algebra $\mathcal{G}_t$, then $\hat{X}_t = E[X_t|\mathcal{G}_t]$. This is a general result for all random variables in Hilbert spaces.

Theorem 1. *The best estimate in mean square error for the system X_t is given by,*

$$\hat{X}_t = E[X_t|\mathcal{G}_t]. \tag{5}$$

We'll take the time to write out the proof, as it is not included in [1].

Proof. Let $Y : \Omega \to \mathbb{R}$ be any $\mathcal{G}$-measurable random variable in $L^2(P)$. Then

$$\begin{aligned} E[(X-Y)^2] &= E[(X - E[X|\mathcal{G}] + E[X|\mathcal{G}] - Y)^2] \\ &= E[(X - E[X|\mathcal{G}])^2] - 2E[(X - E[X|\mathcal{G}])(Y - E[X|\mathcal{G}])] + E[(Y - E[X|\mathcal{G}])^2]. \end{aligned}$$

Next, note that

$$\begin{aligned} 2E[(X - E[X|\mathcal{G}])(Y - E[X|\mathcal{G}])] &= 2E[E[(X - E[X|\mathcal{G}])(Y - E[X|\mathcal{G}])|\mathcal{G}]] \\ &= 2E[(Y - E[X|\mathcal{G}])E[(X - E[X|\mathcal{G}])|\mathcal{G}]] = 0, \end{aligned}$$

since $E[(X - E[X|\mathcal{G}])|\mathcal{G}] = E[X|\mathcal{G}] - E[X|\mathcal{G}]$. Here we used the law of total expectation and theorem B.2.e) on page 320 in [1]. Hence,

$$E[(X-Y)^2] = E[(X - E[X|\mathcal{G}])^2] + E[(Y - E[X|\mathcal{G}])^2] \geq E[(X - E[X|\mathcal{G}])^2],$$

with equality if and only if $Y = E[X|\mathcal{G}]$. □

Next we consider this nice connection between the conditional expectation and orthogonal projections in Hilbert spaces.

Lemma 1. *Let $\mathcal{H} \subset \mathcal{F}$ be a σ-algebra and let $X \in L^2(P)$ be $\mathcal{F}$-measurable. Then, for any subspace $\mathcal{N}$ of the Hilbert space $L^2(P)$, there exist an orthogonal projection onto $\mathcal{N}$, $\mathcal{P}_{\mathcal{N}}$, such that*

$$\mathcal{P}_{\mathcal{N}}(X) = E[X|\mathcal{H}]. \tag{6}$$

It seems our job now is to find $E[X|\mathcal{G}]$, using the fact that it is a projection onto a subspace of L^2. This will be a much easier task due to the fact that the vector (X_t, Z_t) is a Gaussian process (page 92 in [1]). When working with Gaussian random variables, everything is, roughly speaking, easier to work with, as it obeys linearity. This is exactly what the Kalman-Bucy filter takes advantage of. Let us first define $\mathcal{L} = \mathcal{L}(Z,t)$ as the closure in L^2 of the set of functions $\{c_0 + c_1 Z_{s_1} + \ldots + c_k Z_{s_k}\}$, with $s_j \leq t, c_j \in \mathbb{R}$. We now get this useful result.

Lemma 2. *Let $X, Z_1, \ldots, Z_k$ be random variables such that the vector $(X, Z_{s_1}, \ldots, Z_{s_k})$ has a Gaussian distribution. If we let $\mathcal{L}$ be as before, then*

$$E[X|\mathcal{G}] = \mathcal{P}_{\mathcal{L}}(X),$$

where $\mathcal{P}_{\mathcal{L}}(X)$ denotes the projection of the random variable X down to the subspace $\mathcal{L}$.

In other words, instead of finding $E[X_t|\mathcal{G}_t]$, we can calculate $\mathcal{P}_{\mathcal{L}}(X_t)$.

Lemma 3. *Assume that $\mathcal{L}(Z,t)$ is defined as before. Then*

$$\mathcal{L}(Z,t) = \{c_0 + \int_0^T f(t)dZ_t | f \in L^2[0,T], c_0 \in \mathbb{R}\}.$$

The proof can be found on page 93 in [1], and consists of three steps:

- Show that the right side is contained in $\mathcal{L}(Z,t)$.
- Show that the right side contains all functions on the form $c_0 + c_1 Z_{s_1} + \ldots + c_k Z_{s_k}$.
- Show that the right side is closed in L^2.

Since $\mathcal{L}(Z,t)$ is a closure in L^2, meaning $\mathcal{L}(Z,t)$ is the smallest, closed subset that contains all linear combinations of Z_{s_i}'s, this implies that the left side is contained in the right side and the conclusion then follows.

This means our solution $\hat{X}_t$ will be on the form $c_0 + \int_0^T f(t)dZ_t$. However, how can we interpret the integral with respect to Z_t? Clearly Z_t is not a Brownian motion, as it is not stationary with independent increments. However, if we could show that $\mathcal{L}(Z,t) = \mathcal{L}(R,t)$, where R_t is a Brownian motion, the integral can be interpreted as an Itô integral.

In order to achieve this we will first use the innovation process, and then tweak what we get into a Brownian motion. Sounds like a plan!

Definition 1. The *innovation process*, N_t, is defined as

$$N_t = Z_t - \int_0^t G(s)\hat{X}_s ds,$$

or, equivalently,

$$dN_t = dZ_t - G(t)\hat{X}_t dt = G(t)(X_t - \hat{X}_t)dt + \gamma(t)d\tilde{B}t. \tag{7}$$

Let's state some useful properties of N_t:

1. N_t have independent increments.
2. $\mathcal{L}(N,t) = \mathcal{L}(Z,t)$ for all $t \geq 0$.
3. N_t is Gaussian.
4. $E[N_t^2] = \int_0^t \gamma^2(s)ds$.

See page 94 in [1] for justification. As we can see, N_t is pretty close to being a Brownian motion. In fact, the only problem is property 4., since for a Brownian motion B_t we have that $E[B_t^2] = t$. Lucky for us, only a minor correction is needed. Define now

$$dR_t = \frac{1}{\gamma(t)} dN_t. \tag{8}$$

This is always defined, since we, at the beginning of section 2, assumed that $\gamma(s)$ was bounded away from zero. R_t is the Brownian motion such that $\mathcal{L}(Z,t) = \mathcal{L}(R,t)$! Furthermore, the projection down to $\mathcal{L}(R,t)$ has the following nice expression,

$$\hat{X}_t = \mathcal{P}_{\mathcal{L}(R,t)}(X_t) = E[X_t] + \int_0^t \frac{\partial}{\partial s} E[X_t R_s] dR_s. \tag{9}$$

3 The Kalman-Bucy filter

Let us summarize what we have so far, and how we can use this to find an expression for $\hat{X}$. We started with a set of differential equations

$$\begin{aligned} dX_t &= A(t)X_t dt + \alpha(t)dB_t \\ dZ_t &= G(t)X_t dt + \gamma(t)d\tilde{B}_t, \end{aligned}$$

where Z_t is the observation of an unobservable linear system X_t at time $t \in [0,T]$. Since this system is linear, the vector $[X_t, Z_t]$ is a Gaussian process, and we can therefore define the best estimator as an orthogonal projection onto a subspace of linear functions in L^2 as follows,

$$\hat{X}_t = E[X_t|\mathcal{G}_t] = \mathcal{P}_{\mathcal{L}}(X_t).$$

We know that every function in $\mathcal{L}(Z,t)$ can be written on the form $c_0 + \int_0^t f(t)dZ_t$, but we can't interpret the integral with respect to Z_t. Therefore, we find the Brownian motion, R_t, which is such that $\mathcal{L}(Z,t) = \mathcal{L}(R,t)$. The projection of $\hat{X}_t$ into $\mathcal{L}(R,t)$ is one that we are able to calculate.

We now have the expression

$$\hat{X}_t = E[X_t] + \int_0^t \frac{\partial}{\partial s} E[X_t R_s] dR_s.$$

Let us first try to find what $E[X_t R_s]$ is. From the definition of dR_t (8) and dN_t (7) it is clear that

$$R_s = \int_0^s \frac{1}{\gamma(r)} dN_r = \int_0^s \frac{G(r)}{\gamma(r)}(X_r - \hat{X}_r)dr + d\tilde{B}_s. \tag{10}$$

Set $\tilde{X}_t = X_t - \hat{X}_t$, and note that $\hat{X}_t \perp \tilde{X}_t$. By replacing R_s with what we found in equation (10) we get

$$E[X_t R_s] = E[X_t \int_0^s \frac{G(r)}{\gamma(r)} \tilde{X} dr] + E[X_t \tilde{B}_s] = \int_0^s \frac{G(r)}{\gamma(r)} E[X_t \tilde{X}_r] dr.$$

We said at the beginning that the solution of the linear system X_t is

$$X_t = \exp\left\{\int_0^t A(s)ds\right\} X_0 + \int_0^t \exp\left\{\int_s^t A(u)du\right\} \alpha(s)dB_s. \tag{11}$$

Note in particular that $E[X_t] = \exp\left\{\int_0^t A(s)ds\right\} E[X_0]$. We can rewrite $E[X_t\tilde{X}_r]$ as

$$E[X_t\tilde{X}_r] = \exp\left\{\int_r^t A(u)du\right\} E[X_r\tilde{X}_r] = \exp\left\{\int_r^t A(u)du\right\} E[(\tilde{X}_r)^2],$$

where we used the fact that $E[X_r\tilde{X}_r] = E[(\hat{X}_r - \tilde{X}_r)\tilde{X}_r] = E[\hat{X}_r\tilde{X}_r] + E[(\tilde{X}_r)^2]$. Now define $S(r) = E[(\tilde{X}_r)^2]$, then

$$E[X_t R_s] = \int_0^s \frac{G(r)}{\gamma(r)} \exp\left\{\int_r^t A(u)du\right\} S(r)dr$$

and

$$\frac{\partial}{\partial s} E[X_t R_s] = \frac{G(s)}{\gamma(s)} \exp\left\{\int_s^t A(u)du\right\} S(s). \tag{12}$$

Denote $\frac{\partial}{\partial s}E[X_tR_s]$ by $f(s,t)$. Before proceeding forwards, let's take another look at $S(t)$. This is the mean square error of our estimate at any point in time $r \geq 0$. It turns out that $S(t)$ satisfies *the deterministic Riccati equation*,

$$\frac{dS}{dt} = 2A(t)S(t) - \frac{G^2(t)}{\gamma^2(t)}S^2(t) + \alpha^2(t).$$

This is justified on page 99-100 in [1], so we skip those details here.

Remark. The Riccati equation is a differential equation that can be written on the form

$$y'(x) = f_0(x) + f_1(x)y(x) + f_2(x)y^2(x).$$

Riccati equations can be solved by quadratures, which means that the solution is expressed in terms of integrals [2].

Let us return to our search for an expression for $\hat{X}_t$. We have

$$\int_0^u \left(\int_0^t \frac{\partial}{\partial t}f(s,t)\right)dt = \hat{X}_u - c_0(u) - \int_0^u f(s,s)dR_s.$$

Rearranging the terms and taking the derivative on both sides results in

$$d\hat{X}_t = c_0'(t)dt + f(t,t)dR_t + \left(\int_0^t \frac{\partial}{\partial t}f(s,t)dR_s\right)dt.$$

This we can rewrite as

$$d\hat{X}_t = F(t)\hat{X}_tdt + \frac{G(t)S(t)}{\gamma(t)}dR_t. \tag{13}$$

by first substituting f with what we had in (12), and the rearranging the terms such that c_0 is cancelled out.

Finally, since $dN_t = dZ_t - G(t)\hat{X}_tdt$,

$$dR_t = \frac{1}{\gamma(t)}(dZ_t - G(t)\hat{X}_tdt),$$

and substituting this into equation (13) yields

$$d\hat{X}_t = (F(t) - \frac{G^2(t)S(t)}{\gamma^2(t)})\hat{X}_tdt + \frac{G(t)S(t)}{\gamma^2(t)}dZ_t.$$

Theorem 2 (The Kalman-Bucy filter in 1 dimension)**.** *Let Z_t and X_t be two stochastic processes given by*

$$\begin{aligned} dX_t &= A(t)X_tdt + \alpha(t)dB_t \\ dZ_t &= G(t)X_tdt + \gamma(t)d\tilde{B}_t, \end{aligned}$$

where Z_t is the linear observation of the linear system X_t, and Z_t, X_t satisfy all conditions listed at the beginning of section 2. Then the best mean-square estimator for X_t is $\hat{X}_t = E[X_t|\mathcal{G}]$, where $\mathcal{G}_t$ is the σ-algebra generated by Z_t, and $\hat{X}_t$ will satisfy

$$d\hat{X}_t = (F(t) - \frac{G^2(t)S(t)}{\gamma^2(t)})\hat{X}_tdt + \frac{G(t)S(t)}{\gamma^2(t)}dZ_t. \tag{14}$$

$S(t)$ is the mean-square error of the estimate at time t, and satisfy the deterministic Ricatti equation

$$\frac{dS}{dt} = 2A(t)S(t) - \frac{G^2(t)}{\gamma^2(t)}S^2(t) + \alpha^2(t). \tag{15}$$

Remark. The Kalman-Bucy filter is the best mean square estimator when the condition of normality holds. If this criteria is not fulfilled, however, other estimators can be better mean square estimators than the Kalman-Bucy filter, but it remains the best linear estimator even then.

4 Example

We have mainly motivated our search by considering a stochastic differential equation which cannot be observed our measured directly. For this example we shall consider instead how to use the Kalman-Bucy filter to estimate a parameter. This is example 6.2.11 in [1].

Suppose that we have observations of a stochastic process given by

$$dZ_t = \theta M(t)dt + N(t)dB_t.$$

Here, M and N are known functions, whereas θ is an unknown drift parameter, which does not depend on t. Therefore $d\theta = 0$. How can we use the Kalman-Bucy filter in this situation? Start by noting that

$$\begin{aligned} \frac{dS}{dt} &= -\frac{M^2(t)}{N^2(t)} S^2(t), \\ \implies \frac{1}{S^2(t)} dS &= -\frac{M^2(t)}{N^2(t)} ds. \end{aligned}$$

This is a separable ordinary differential equation, so we get

$$S(t) = \frac{1}{S_0^{-1} + \int_0^t \frac{M^2(s)}{N^2(s)} dt}.$$

Next, from equation (14) we get

$$d\hat{\theta}_t = \frac{1}{S_0^{-1} + \int_0^t \frac{M^2(s)}{N^2(s)} ds} \left(\frac{M(t)}{N^2(t)} dZ_t - \frac{M^2(t)}{N^2(t)} \hat{\theta} dt \right).$$

Separate this so we get terms depending on θ by itself as follows,

$$\left(S_0^{-1} + \int_0^t \frac{M^2(s)}{N^2(s)} ds \right) d\hat{\theta}_t + \frac{M^2(t)}{N^2(t)} \hat{\theta} dt = \frac{M(t)}{N^2(t)} dZ_t.$$

The left hand side is the derivative of $\left(S_0^{-1} + \int_0^t \frac{M^2(s)}{N^2(s)} ds \right) \hat{\theta}_t$ (this is simply the product rule), and son integrating both sides yield

$$\left(S_0^{-1} + \int_0^t \frac{M^2(s)}{N^2(s)} ds \right) \hat{\theta}_t - S_0^{-1} \hat{\theta}_0 = \int_0^t \frac{M(t)}{N^2(t)} dZ_t,$$

which finally gives us,

$$\hat{\theta}_t = \frac{S_0^{-1} \hat{\theta}_0 + \int_0^t \frac{M(t)}{N^2(t)} dZ_t}{S_0^{-1} + \int_0^t \frac{M^2(s)}{N^2(s)} ds}. \tag{16}$$

Remark. When $S_0^{-1} = 0$, equation (16) gives the same estimate as the maximum likelihood estimate for θ.

Appendices

A Calculation of solution

Here I show the steps for how to solve a linear stochastic differential equation on the form

$$dX_t = A(t)X_t dt + \alpha(t)dB_t.$$

Start by setting the terms depending on X_t by itself, such that

$$dX_t - A(t)X_t dt = \alpha(t)dB_t.$$

First we multiply both sides of this expression with a not yet specified Y_t, and get

$$Y_t dX_t - Y_t A(t)X_t dt = Y_t\alpha(t)dB_t. \tag{17}$$

Then, we can find Y_t such that,

$$d(Y_t X_t) = Y_t dX_t - Y_t A(t)X_t dt, \tag{18}$$

that is, the left side of equation (17). Assume Y_t is an Itô process, and that

$$dY_t = f_1(t, Y_t)dt + f_2(t, Y_t)dB_t,$$

with $dY_0 = 0$. Then

$$\begin{aligned} d(X_tY_t) &= Y_t dX_t + X_t dY_t + dX_t dY_t \\ &= Y_t dX_t + X_t(f_1(t, Y_t)dt + f_2(t, Y_t)dB_t) + \alpha(t)f_2(t, Y_t)dt \end{aligned} \tag{19}$$

since $dX_t dY_t = \alpha(t)f_2(t, Y_t)dt$.[2]
Comparison of the right side of equation (18) with the right side of equation (19) yields

$$\begin{aligned} -Y_t A(t)X_t dt &= (f_1(t, Y_t)X_t dt + \alpha(t)f_2(t, Y_t)dt \\ 0 &= X_t f_2(t, Y_t)dB_t. \end{aligned}$$

Hence,

$$f_2(t, Y_t) = 0 \qquad f_1(t, Y_t) = -Y_t A(t).$$

From this we can find an expression of Y_t as follows:

$$\begin{aligned} dY_t &= -Y_t A(t)dt \\ \implies \frac{1}{Y_t}dY_t &= -A(t)dt \\ \implies \ln(Y_t) &= -\int_0^t A(s)ds \\ \implies Y_t &= \exp\left\{-\int_0^t A(s)ds\right\}. \end{aligned}$$

[2]Remember that $(dt)^2 = dB_t dt = dt dB_t = 0$.

Looking at equation (17) again, we get that

$$\begin{aligned} d(Y_t X_t) &= Y_t \alpha(t) dB_t \\ Y_t X_t &= X_0 + \int_0^t Y_s \alpha(s) dB_s \\ X_t &= Y_t^{-1} \left(X_0 + \int_0^t Y_s \alpha(s) dB_s \right). \end{aligned}$$

This finally yields,

$$X_t = X_0 \exp \left\{ \int_0^t A(s) ds \right\} + \int_0^t \exp \left\{ \int_s^t A(u) du \right\} \alpha(s) dB_s.$$

Bibliography

[1] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Universitext. Springer Berlin Heidelberg, 2013.

[2] Encyclopedia of Mathematics. Riccati equation. `https://encyclopediaofmath.org/index.php?title=Riccati_equation`, (accessed: 20.11.2021).