Application of Girsanov's Theorem for Brownian Motion and Poisson Processes to Common Filtering Methods in Engineering

SooJean Han*

April 21, 2021

Abstract

Filtering and state estimation problems are common, but theoretically, little of it is understood, especially the conditional expectation projection mechanism. In this article, we seek to provide a better understanding of the projection component by making theoretical connections between innovation processes and the change of measure formulas from the theory of stochastic processes to two well-known methods of single-target, single-sensor filtering: 1) the Kalman filter, and 2) the particle filter. A notable distinction between the particle filtering case and the Kalman filtering case is the transformation of the quantity in the conditional expectation $\mathbb{E}[\mathbf{x}|\mathscr{A}]$: while Girsanov's theorem transforms $\mathbf{x}$ using the Radon-Nikodym derivative, the method of innovation processes transforms the quantity $\mathscr{A}$.

1 Introduction

The problem of filtering and state estimation is undoubtedly important in the fields of control theory, signal processing, and statistics. Some notable practical applications of filtering are vision-based localization and mapping [1] and dynamic target-tracking [2]. In literature, filters are often designed to be robust against additive Gaussian white noise due to their prevalence and the advantageous properties which makes filter synthesis easier to perform than non-Gaussian noise. However, some non-Gaussian noise phenomena shows up in the real-world just as commonly as Gaussian white noise does; one notable class are the sudden impulsive perturbations of Poisson-distributed shot noise, which arise in signal-processing neuronal spikes arising from brain activity in neuroscience [3], and large fluctuations in stock prices in financial economics [4]. In the field of robotics, they can arise as massive proprioceptive measurement errors or large disturbances due to obstacle collisions.

Consequently, there has been rich development occurring in the theory of filters since the seminal work of the discrete-time Kalman filter [5] and the continuous-time Kalman-Bucy filter. One immediate limitation to these traditional Kalman filters is that they only deal with linear dynamics; the well-known EKF [6] and UKF [7] were developed to address nonlinear dynamics. A second limitation is that the Kalman filter only considers white noise. To this end, the Kalman-Bucy filter has been extended to account for a specific class of Lévy noise [8].

The problem of filtering is an important one. There are many problems in real life where it is important to obtain the true state value from noisy measurements. Examples include vision-based localization and

*SooJean Han is with the Department of Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125, USA. Email: `soojean@caltech.edu`

mapping [1], dynamic sensor networks, consensus networks [9, 10] in robotics, signal-processing neuronal spikes arising from brain activity in neuroscience [3], and stock price modeling in financial economics [4].

Due to its prevalence in practical real-world scenarios, there has also been a lot of theory devoted to it. Since the seminal work of Kalman filtering [5] and its extensions [6, 7]... Particle filtering [11, 12], with feedback [13]. Nonlinear filtering approaches [14].

In this article, we address two questions related to the theoretical aspect of Bayesian filtering approaches. First, how do the change of measure formulas and the innovation process approach relate to the conditional expectation projection mechanism that is seen in Bayesian filtering? More specifically, how are the innovation process approach and Girsanov's theorem related to each other? Second, we provide extensions to systems perturbed by a class of noise processes broader than Gaussian white noise for both particle filtering and Kalman filtering.

2 The Standard Bayesian Approach

Consider the following general nonlinear system

$$d\mathbf{x}(t) = F_c(t, \mathbf{x}(t), \mathbf{w}(t))dt \tag{1a}$$

$$\mathbf{y}(t) = h(\mathbf{x}(t), \mathbf{v}(t)) \tag{1b}$$

where $\mathbf{x}(t) \in \mathcal{X} := \mathbb{R}^{n_x}, \mathbf{y}(t) \in \mathcal{Y} := \mathbb{R}^{n_y}, n_y \leq n_x$, $F_c : \mathbb{R}^+ \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_w} \to \mathbb{R}^{n_x}, h : \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \to \mathbb{R}^{n_y}$ are possibly nonlinear, time-varying functions, stochastic disturbances $\mathbf{w}(t)$ are i.i.d, $\mathbf{v}(t)$ are i.i.d and belong on probability space $(\Omega, \mathcal{F}, P)$, where Ω denotes the sample space of all possible outcomes, $\mathcal{F}$ is the set of events, and P is the probability function.

We assume that the system is *Markovian*: 1) given $\mathbf{x}(s)$ for some $s \leq t$, $\mathbf{x}(t)$ is independent of $\{\mathbf{x}(r) : r < s\}$ and all past measurements $\{\mathbf{y}(r) : r < s\}$, 2) given $\mathbf{x}(t)$, $\mathbf{y}(t)$ is independent of all past states and measurements.

For state estimation problems, it is often convenient to discretize the system with some (possibly fixed) sample time $\Delta t > 0$.

$$\mathbf{x}_{k+1} = F_d(k, \mathbf{x}_k, \mathbf{w}_k) \tag{2a}$$

$$\mathbf{y}_k = h(\mathbf{x}_k, \mathbf{v}_k) \tag{2b}$$

This discretized system is also assumed to be Markovian in a similar way as its continuous-time analog. An alternative way of describing (2) is through the use of probability distributions. Denote $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$ to be the probability of transitioning from $\mathbf{x}_k$ to $\mathbf{x}_{k+1}$ for any pair of states $\mathbf{x}_k, \mathbf{x}_{k+1} \in \mathcal{X}$. Further denote $\ell(\mathbf{y}_k|\mathbf{x}_k)$ to be the likelihood probability of observing $\mathbf{y}_k \in \mathcal{Y}$ given $\mathbf{x}_k \in \mathcal{X}$.

The objective of the *filtering problem* is to estimate the full state $\mathbf{x}_k$ given data observations $\mathscr{A}_k := \sigma(\mathbf{y}_1, \cdots, \mathbf{y}_k)$. Taking a Bayesian approach, we construct the pdf $f(\mathbf{x}_k|\mathscr{A}_k)$ given $f(\mathbf{x}_{k-1}|\mathscr{A}_{k-1})$ at each timestep k through a two-step process: prediction and measurement update.

1. **Prediction**: given $f(\mathbf{x}_{k-1}|\mathscr{A}_{k-1})$, use the Chapman-Kolmogorov equation to predict $\mathbf{x}_k$

$$f(\mathbf{x}_k|\mathscr{A}_{k-1}) = \int f(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathscr{A}_{k-1}) f(\mathbf{x}_{k-1}|\mathscr{A}_{k-1}) d\mathbf{x}_{k-1} = \int f(\mathbf{x}_k|\mathbf{x}_{k-1}) f(\mathbf{x}_{k-1}|\mathscr{A}_{k-1}) d\mathbf{x}_{k-1} \tag{3}$$

   where the second equality follows from Markovian assumptions. In this form, $f(\mathbf{x}_k|\mathbf{x}_{k-1})$ can be directly determined from the system dynamics.

2. **Measurement Update**: given $f(\mathbf{x}_k|\mathscr{A}_{k-1})$, use Bayes' Rule to incorporate new data point $\mathbf{y}_k$ into our prediction:

$$f(\mathbf{x}_k|\mathscr{A}_k) = \frac{f(\mathbf{x}_k, \mathbf{y}_k|\mathscr{A}_{k-1})}{f(\mathbf{y}_k|\mathscr{A}_{k-1})} = \frac{\ell(\mathbf{y}_k|\mathbf{x}_k) f(\mathbf{x}_k|\mathscr{A}_{k-1})}{f(\mathbf{y}_k|\mathscr{A}_{k-1})} \tag{4}$$

where the second equality follows from Markovian assumptions, $f(\mathbf{y}_k|\mathbf{x}_k)$ can be determined from the observation equation, and

$$f(\mathbf{y}_k|\mathscr{A}_{k-1}) = \int \ell(\mathbf{y}_k|\mathbf{x}_k) f(\mathbf{x}_k|\mathscr{A}_{k-1}) d\mathbf{x}_k$$

Remark 1. The two steps are alternatively referred to as the **prior estimate** and the **posterior update**. □

Remark 2. The *prediction problem*, which predicts $\mathbf{x}_{k+1}$ from past data observations $\mathscr{A}_k := \sigma(\mathbf{y}_1, \cdots, \mathbf{y}_k)$, is very similar to the filtering problem: simply reverse the above two steps to construct the pdf $f(\mathbf{x}_{k+1}|\mathscr{A}_k)$ given $f(\mathbf{x}_k|\mathscr{A}_{k-1})$. □

One primary issue is that for general nonlinear systems, a direct implementation of the above two steps is often difficult to do. Throughout this chapter, we will be focusing on additive noise perturbations. That is, we consider (1) specifically of the form

$$d\mathbf{x}(t) = f_c(t, \mathbf{x}(t))dt + d\mathbf{w}(t) \tag{5a}$$

$$\mathbf{y}(t) = h(\mathbf{x}(t)) + \mathbf{v}(t) \tag{5b}$$

and the corresponding discrete-time system

$$\mathbf{x}_{k+1} = f_d(k, \mathbf{x}_k) + \mathbf{w}_k \tag{6a}$$

$$\mathbf{y}_k = h(\mathbf{x}_k) + \mathbf{v}_k \tag{6b}$$

By way of the *Orthogonality Principle*, we can show that the *optimal estimator* $\hat{\mathbf{x}}(t)$ of $\mathbf{x}(t)$ is given by $\hat{\mathbf{x}}(t) = \mathbb{E}_P[\mathbf{x}(t)|\mathscr{A}(t)]$ and likewise, in the discrete-time case, $\hat{\mathbf{x}}_k = \mathbb{E}_P[\mathbf{x}_k|\mathscr{A}_k]$. Hence, we discuss the Orthgonality Principle in the following section.

2.1 The Orthogonality Principle

Define $\mathcal{L}^2 := \mathcal{L}^2(\Omega, \mathcal{F}, P)$ to be the space of all random variables on $\Omega \subseteq \mathbb{R}^n$ with finite second moments, i.e, the space of $\mathbf{x}$ such that $\mathbb{E}_P[\|\mathbf{x}\|^2] < \infty$. To maintain simplicity in this section, we use the notation $\mathbf{x}$, without mentioning the time index, to denote the random variable $\mathbf{x}(t)$ or $\mathbf{x}_k$ instead of the actual time-varying trajectory. Let $\mathcal{V}$ be a closed, linear subspace on $\mathcal{L}^2$.

With this setup, the estimation problem that the Bayesian filtering process posed to us in the previous section can be restated as follows: determine the estimator of $\mathbf{x} \in \mathcal{L}^2$ over all $\mathcal{V}$ with the least mean-squared error, i.e., we want to find a $\mathbf{z}^* \in \mathcal{V}$ such that $\mathbb{E}_P[\|\mathbf{x} - \mathbf{z}^*\|^2] \leq \mathbb{E}_P[\|\mathbf{x} - \mathbf{z}\|^2]$ for all $z \in \mathcal{V}$. Correspondingly, $\mathbb{E}_P[\|\mathbf{x} - \mathbf{z}^*\|^2]$ is known as the *minimum mean squared error (MMSE)*. See Figure 1 for visualization in the case where $\mathcal{V}$ is a line.

Theorem 1 (Orthogonality Principle)**.** The following conditions are equivalent:

1. There exists a **unique** element $\mathbf{z}^* \in \mathcal{V}$ which achieves the MMSE.

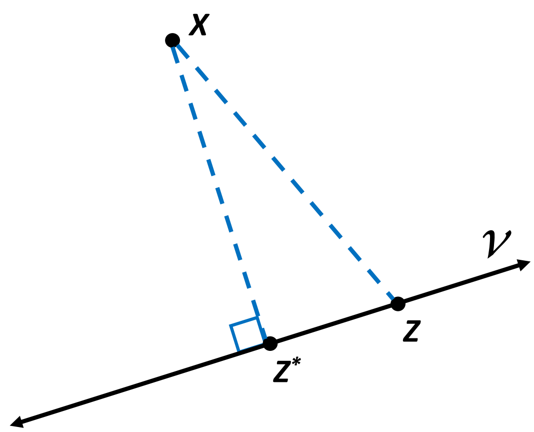

Figure 1: A visualization of the projection mechanism where $\mathcal{V}$ is a line and $\mathbf{x} \in \mathbb{R}^n$. Note that any other $\mathbf{z} \in \mathcal{V}$ does not achieve the minimum error.

2. Let $\mathbf{y} \in \mathcal{L}^2$. Then $\mathbf{y} = \mathbf{z}^*$ iff the following two conditions hold:
   a. $\mathbf{y} \in \mathcal{V}$
   b. $(\mathbf{x} - \mathbf{y}) \perp \mathbf{z}$ for all $\mathbf{z} \in \mathcal{V}$.
3. As a consequence of the above two conditions, the MMSE has a nice simplification:
$$\mathbb{E}_P[(\mathbf{x} - \mathbf{z}^*)^T(\mathbf{x} - \mathbf{z}^*)] = \mathbb{E}_P[\mathbf{x}^T\mathbf{x}] - \mathbb{E}_P[(\mathbf{z}^*)^T\mathbf{z}^*] \quad \text{since } (\mathbf{x} - \mathbf{z}^*) \perp \mathbf{z}^*$$

Furthermore, the MMSE is an *unbiased estimator*, meaning $\mathbb{E}_P[\mathbf{x} - \mathbf{z}^*] = 0$. $\mathbf{z}^*$ is alternatively known as the *projection* of x onto $\mathcal{V}$.

Example 1 (Specialization to Linear Subspaces)**.** One of the most common classes of estimators to constrain $\mathbf{z}^*$ to are linear estimators, which consequently gives rise to the *Kalman filtering* algorithm and its variants. We specialize the above analysis to the affine subspace $\mathcal{V} = \{c_0 + c_1 y_1 + \cdots + c_m y_m; c_i \in \mathbb{R}^n\}$ for a specific value $\mathbf{y} := (y_1, \cdots, y_m)$. The projection of $\mathbf{x}$ onto $\mathcal{V}$ is of the form $\mathbf{z}^* = A\mathbf{y} + \mathbf{b}$. The estimation error becomes $\mathbf{e} = \mathbf{x} - (A\mathbf{y} + \mathbf{b})$. Use the Orthogonality Principle to determine what the coefficients A and $\mathbf{b}$ should be. Namely, in order to have $\mathbf{e} \perp \mathbf{z}$ for all $\mathbf{z} \in \mathcal{V}$, we need:

- $\mathbb{E}_P[\mathbf{e}] = 0$, which implies that $\mathbf{b} = \mathbb{E}_P[\mathbf{x}] - A\mathbb{E}_P[\mathbf{y}]$. Thus, $z^* = \mathbb{E}_P[\mathbf{x}] + A(\mathbf{y} - \mathbb{E}_P[\mathbf{y}])$.
- $\mathrm{Cov}(\mathbf{e}, \mathbf{y}) = 0$. Combined with the previous expression, we get:
$$\begin{aligned}\mathrm{Cov}(\mathbf{x} - \mathbb{E}_P[\mathbf{x}] - A(\mathbf{y} - \mathbb{E}_P[\mathbf{y}]), \mathbf{y}) = 0 &\implies \mathrm{Cov}(\mathbf{x}, \mathbf{y}) - A\mathrm{Cov}(\mathbf{y}) = 0 \\ &\implies A = \mathrm{Cov}(\mathbf{x}, \mathbf{y})\mathrm{Cov}(\mathbf{y})^{-1}\end{aligned}$$

Combined together, we have the final expression for the MMSE. To distinguish the notation between the general and the linear cases, we denote the conditional expectation for the linear case with $\hat{E}$ instead of the usual $\mathbb{E}$.
$$\hat{\mathbf{x}} := \hat{E}[\mathbf{x}|\mathbf{y}] := \mathbb{E}_P[\mathbf{x}] + \mathrm{Cov}(\mathbf{x}, \mathbf{y})\mathrm{Cov}(\mathbf{y})^{-1}(\mathbf{y} - \mathbb{E}_P[\mathbf{y}]) \tag{7}$$

Moreover, we find that $\mathrm{Cov}(\mathbf{e})$ satisfies:
$$\mathrm{Cov}(\mathbf{e}) = \mathrm{Cov}(\mathbf{x}) - \mathrm{Cov}(\hat{\mathbf{x}}) = \mathrm{Cov}(\mathbf{x}) - \mathrm{Cov}(\mathbf{x}, \mathbf{y})\mathrm{Cov}(\mathbf{y})^{-1}\mathrm{Cov}(\mathbf{y}, \mathbf{x})$$

□

3 Change of Measure Formulas

3.1 For Brownian Motion Processes

Definition 1 (Absolute Continuity)**.** A measure Q is said to be *absolutely continuous* with respect to another measure P (also denoted mathematically by $Q << P$) if the null set of P is also under the null set of Q. That is, for any event set E, if $P(E) = 0$ then $Q(E) = 0$. □

Definition 2 (Radon-Nikodym Derivative)**.** The Radon-Nikodym derivative is defined as the multiplicative factor which transforms from measure P to measure Q, for $Q << P$:

$$Z(x) = \frac{dQ(x)}{dP(x)} \implies dQ(x) = Q(X = x)dx = Z(x)dP(x) \tag{8}$$

In terms of conditional expectations:

$$\mathbb{E}_P[Z(x)] = \int Z(x)dP(x) = \int dQ(x) = \mathbb{E}_Q[1]$$

and for general functions $h \in \mathcal{L}^1$, $\mathbb{E}_P[Z(x)h(x)] = \mathbb{E}_Q[h(x)]$.

In the case of the scalar Gaussian random variable, the interpretation of the Radon-Nikodym derivative goes as follows: with respect to measure P, X is normally distributed with mean μ, variance σ^2, and with respect to measure Q, X is normally distributed with mean 0 (same variance σ^2). □

There are two analogous change-of-measure formulas for the Gaussian stochastic random process:

1. Discrete-Time Case: We begin with $X_n = \sum_{i=1}^{n} \Delta X_i \sim \mathcal{N}[0, t_n]$ under measure P. The shift term is given by

   $$Y_n = \sum_{i=1}^{n} \mu_i \Delta t$$

   which corresponds to the Radon-Nikodym derivative

   $$Z_n = \frac{dQ}{dP} = e^{\sum\limits_{i=1}^{n} \mu_i \Delta x_i - \frac{1}{2} \sum\limits_{i=1}^{n} \mu_i^2 \Delta t} \tag{9}$$

   This causes the shift to $X_n - \sum_{i=1}^{n} \mu_i \Delta t \sim \mathcal{N}(0, t_n)$ under new measure Q.

2. Continuous-Time Case: We begin with $W(t)$, the standard Brownian motion process, under measure P. The shift term is given by

   $$Y(t) = \int_0^t \mu(s)ds$$

   which corresponds to the Radon-Nikodym derivative

   $$Z(t) = \frac{dQ}{dP} = e^{\int_0^t \mu(s)dW(s) - \frac{1}{2}\int_0^t \mu^2(s)ds} \tag{10}$$

   This causes the shift to $\tilde{W}(t) := W(t) - \int_0^t \mu(s)ds$ under new measure Q. Furthermore, the Radon-Nikodym derivative $Z(t)$ for this Brownian motion case is commonly referred to as the **Doléans-Dade exponential**. A further treatment of the continuous-time case can be found in [15].

Proposition 1. The Doléans-Dade exponential $Z(t)$ (10) is a martingale.

In summary, (10) is the solution to the SDE $dZ(t) = Z(t)\mu(t)dW(t)$, and encompasses the transformation $\tilde{W}(t) = W(t) - \mu(t)$. One can additionally extend the Doléans-Dade exponential to the vector case. Suppose $W(t) \in \mathbb{R}^d$ and $\overrightarrow{\mu}(t) \in \mathbb{R}^d$ is the desired shift in mean. Then (10) in the vector case is written as:

$$Z(t) = e^{\int_0^t \overrightarrow{\mu}^T(s)dW(s) - \frac{1}{2}\int_0^t \overrightarrow{\mu}^T(s)\overrightarrow{\mu}(s)ds} \tag{11}$$

and is the solution to the SDE $dZ(t) = Z(t)\mu^T(t)dW(t)$. Note that the Radon-Nikodym derivative $Z(t)$ is still a scalar quantity.

Lemma 1 (Bayes' Theorem for Conditional Expectations)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space and Q be such that $Q << P$. Let the multiplicative change of measure term be denoted by Z denote the Radon-Nikodym derivative transforming from P to Q. Then for any σ-algebra $\mathcal{G} \subset \mathcal{F}$, and any integrable random variable $X \in \mathbb{R}$ such that $\mathbb{E}_Q[|X|] < \infty$, the Bayes' formula holds:

$$\mathbb{E}_Q[X|\mathcal{G}] = \frac{\mathbb{E}_P[ZX|\mathcal{G}]}{\mathbb{E}_P[Z|\mathcal{G}]}$$

In the case of random vectors, $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbb{E}_Q[\|\mathbf{x}\|] < \infty$ where the norm can be any norm, Bayes' formula is:

$$\mathbb{E}_Q[\mathbf{x}|\mathcal{G}] = \frac{\mathbb{E}_P[Z\mathbf{x}|\mathcal{G}]}{\mathbb{E}_P[Z|\mathcal{G}]}$$

Note that the Radon-Nikodym derivative is still a scalar.

We can invoke **Lévy's Characterization of the Brownian motion** to show that the transformed white noise process is indeed a Brownian motion under the new measure Q, i.e., the new $\tilde{W}(t) := W(t) - \mu t$ is a martingale. The characterization consists of three parts:

1. $P(\tilde{W}(0) = 0) = 1$
2. $\tilde{W}(t)$ is a martingale with continuous sample paths
3. $\tilde{W}^2(t) - t$ is a martingale.

The proofs of this characterization and Bayes' rule can be found in multiple texts, so we do not discuss them here.

Now that we have discussed the change of measure between Brownian motion processes, we discuss how such a change of measure affects SDEs perturbed by Brownian motion processes. In particular, changing between two drift terms of the additive-noise SDE will be important for our discussion of measure transformations in the particle filter.

Example 2 (Drift-to-Drift Transformation)**.** Consider the following example transformations of the Brownian motion process. Let $\sigma(t) \not\equiv 0$ for all t.

1. Suppose we want to transform a drift-perturbed Brownian motion process to a process without drift.

$$dX(t) = \mu(t)dt + \sigma(t)dW(t) = \sigma(t)\left(\frac{\mu}{\sigma(t)}dt + dW(t)\right)$$

We want to find a $\tilde{W}(t)$ such that $dX(t) = \sigma(t)d\tilde{W}(t)$.

Choose the shift term $Y(t) = -(\mu(t)/\sigma(t))t$ so that $\tilde{W}(t) = W(t) + (\mu(t)/\sigma(t))t$. Then:

$$\tilde{Z}(t) = \frac{dQ(t)}{dP(t)} = e^{-\frac{\mu(t)}{\sigma(t)}W(t) - \frac{1}{2}\left(\frac{\mu(t)}{\sigma(t)}\right)^2 t}$$

2. Now suppose we want to transform a zero-drift Brownian motion process to one with nonzero drift.

$$dX(t) = \sigma(t)\tilde{W}(t) = \nu(t)dt + (-\nu(t)dt + \sigma(t)\tilde{W}(t)) = \nu(t)dt + \sigma(t)\left(-\frac{\nu(t)}{\sigma(t)}dt + \tilde{W}(t)\right)$$

We want to find $W'(t)$ such that $dX(t) = \nu(t)dt + \sigma(t)dW'(t)$.

Choose the shift term $\tilde{Y}(t) = (\nu(t)/\sigma(t))t$ so that $W'(t) = \tilde{W}(t) - (\nu/\sigma)t$. Then:

$$Z'(t) = \frac{dR(t)}{dQ(t)} = e^{\frac{\nu(t)}{\sigma(t)}\tilde{W}(t) - \frac{1}{2}\left(\frac{\nu(t)}{\sigma(t)}\right)^2 t}$$

Overall, the change of measure from drift $\mu(t)$ to drift $\nu(t)$ is performed using the Radon-Nikodym derivative

$$\frac{dR(t)}{dP(t)} = \frac{\cancel{dQ(t)}}{dP(t)} \cdot \frac{dR(t)}{\cancel{dQ(t)}} = \tilde{Z}(t)Z'(t)$$

with

$$dW'(t) = dW(t) - \frac{\nu(t) - \mu(t)}{\sigma(t)}dt$$

□

3.2 For Poisson and Semimartingale Processes

The expression for the Doléans-Dade exponential (10) might be familiar to the reader, especially when viewing it as an analogy to the well-known fact that the form of the solution to the equation $dZ(t) = Z(t)dt, Z(0) = 1$ is given by the exponential $Z(t) = e^t$. More generally, for ny scalar-valued, differentiable, deterministic function $X(t)$, the solution to the system $dZ(t) = Z(t)dX(t), Z(0) = 1$ is given by $Z(t) = e^{(X(t)-X(0))}$. Clearly, when $dX(t) = \mu(t)dW(t)$, we obtain (10) as the solution to $dZ(t) = Z(t)\mu(t)dW(t), Z(0) = 1$, which was shown in Proposition 1 when $\mu(t) = \mu$, a constant. Likewise, when $dX(t) = \mu^T(t)dW(t)$, we obtain (11), as the solution to $dZ(t) = Z(t)\mu^T(t)dW(t), Z(0) = 1$. Recall that the interpretation of $Z(t)$ is that it is the Radon-Nikodym derivative used to transform from $dW(t)$ to $d\tilde{W}(t) := dW(t) - \mu(t)dt$.

Proposition 2. The Doléans-Dade exponential for semimartingale X is given by:

$$Z(t) = e^{X(t) - \frac{1}{2}[X,X](t)} \prod_{0<s\leq t} (1 + \Delta X(s))e^{-\Delta X(s) + \frac{1}{2}\Delta X^2(s)} \tag{12}$$

Note that the jump part of the quadratic variation cancels out with the jumps parts in the product term. This allows us to rewrite the expression as:

$$Z(t) = e^{X(t) - \frac{1}{2}[X,X]^c(t)} \prod_{0<s\leq t} (1 + \Delta X(s))e^{-\Delta X(s)}$$

Proof of Proposition 2. The system $\Delta Z_n = Z_{n-1}\Delta X_n, Z_0 = 1$, with $\Delta Z_n := Z_n - Z_{n-1}$ can be recursively substituted backwards until the base case. We get:

$$Z_n = (1 + \Delta X_n)Z_{n-1} = \prod_{i=1}^{n}(1 + \Delta X_i)$$

This is easily extendable to the continuous-time case:

$$Z(t) = (1 + \Delta X(t))Z(t-) \text{ for all } t > 0$$

Apply Itô's formula for semimartingales:

$$d\left(\ln(Z(t))\right) = \frac{dZ(t)}{Z(t)} - \frac{1}{2Z^2(t)}d[Z,Z]^c(t) + \sum_{0<s\leq t}\left[\ln(Z(s)) - \ln(Z(s-)) - \frac{\Delta Z(s)}{Z(s-)}\right] \tag{13}$$

where $\Delta Z(s) := Z(s) - Z(s-)$.

We have $d[Z,Z]^c(t) = Z^2 d[X,X]^c(t)$ and

$$\begin{aligned}\ln(Z(s)) - \ln(Z(s-)) &= \ln\left(\frac{Z(s-)(1+\Delta X(s))}{Z(s-)}\right) = \ln(1+\Delta X(s))\\ \frac{\Delta Z(s)}{Z(s-)} &= \frac{Z(s-)\Delta X(s)}{Z(s-)} = \Delta X(s)\end{aligned}$$

Substituting, we get:

$$\begin{aligned}(13) &= dX(t) - \frac{1}{2}d[X,X]^c(t) + \sum_{0<s\leq t}\left[\ln(1+\Delta X(s)) - \Delta X(s)\right]\\ \implies Z(t) &= e^{X(t)-\frac{1}{2}[X,X]^c(t)}\prod_{0<s\leq t}(1+\Delta X(s))e^{-\Delta X(s)}\end{aligned}$$

which is exactly our desired formula. ∎

How does the Doléans-Dade exponential look like for specific Poisson processes? In the following two propositions, we show the form of the Radon-Nikodym derivatives needed for the standard Poisson process and the compound Poisson process.

Proposition 3 (CoM for the Standard Poisson Process)**.** Suppose $N(t)$ is a standard Poisson process with constant intensity λ according to measure P. Then the Radon-Nikodym derivative to transform $N(t)$ to a standard Poisson process with intensity $\tilde{\lambda}$ is expressed as

$$Z(t) = e^{-t(\lambda-\tilde{\lambda})}\left(\frac{\tilde{\lambda}}{\lambda}\right)^{N(t)} \tag{14}$$

Furthermore, (14) satisfies the SDE

$$dZ(t) = Z(t-)\left(\frac{\tilde{\lambda}-\lambda}{\lambda}\right)d(N(t) - \lambda t) \tag{15}$$

Proof. First, it is easy to verify (14) by writing out the probability distributions:

$$Q(N(t)=k) = e^{-t(\tilde{\lambda}-\lambda)}\left(\frac{\tilde{\lambda}}{\lambda}\right)^k P(N(t)=k) = e^{-t(\tilde{\lambda}-\lambda)}\left(\frac{\tilde{\lambda}}{\lambda}\right)^k \cdot e^{-\lambda t}\frac{(\lambda t)^k}{k!} = e^{-\tilde{\lambda}}\frac{\tilde{\lambda}^k t^k}{k!}$$

Thus, $N(t)$ is also Poisson under the measure Q, just with intensity $\tilde{\lambda}$ as opposed to λ.

Note that

$$X^c(t) = (\tilde{\lambda}-\lambda)t,\ X^d(t) = \left(\frac{\tilde{\lambda}-\lambda}{\lambda}\right)N(t),\ [X,X]^c(t) = 0$$

Hence, if there exists a jump at time t, $\Delta X(t) = \Delta X^d(t)$. Rearranging the terms yields

$$\frac{\tilde{\lambda}}{\lambda} = 1 + \Delta X(t)$$

Applying these to Proposition 2, we get

$$Z(t) = e^{\left(\frac{\tilde{\lambda}}{\lambda}-1\right)(N(t)-\lambda t)-0}\prod_{0<s\leq t}\left(\frac{\tilde{\lambda}}{\lambda}\right)e^{-\left(\frac{\tilde{\lambda}}{\lambda}\right)} = e^{(\lambda-\tilde{\lambda})t}\left(\frac{\tilde{\lambda}}{\lambda}\right)^{N(t)}$$

and we indeed obtain the desired expression (14). ■

Now consider the compound Poisson process $Y(t)$, written as follows:

$$Y(t) = \sum_{i=1}^{N(t)} \xi_i \tag{16}$$

where $N(t)$ is the standard Poisson process of intensity λ, and the jump sizes ξ_i are iid random variables.

It is of convenience to recall the moment-generating function for (16).

Proposition 4 (Moment-Generating Function of Compound Poisson Process)**.** The moment-generating function of $Y(t)$ from (16) is given by

$$\phi_Y(t) = e^{\lambda t\int_{-\infty}^{\infty}(e^{u\xi}-1)\nu(d\xi)}$$

Proof. This can be seen through the following argument. Since $N(t)$ has a Poisson distribution with parameter $t>0$ and is independent of $\{\xi_k\}$, we get

$$\mathbb{E}_P\left[e^{uY(t)}\right] = \mathbb{E}_P\left[e^{\left(u\sum_{k=1}^{N(t)}\xi_k\right)}\right] = \sum_{n=0}^{\infty}\mathbb{E}_P\left[e^{\left(u\sum_{k=1}^{n}\xi_k\right)}\middle| N(t)=n\right]P(N(t)=n) \tag{17}$$

where the last equality follows by definition of expectation. From the fact that the Poisson process has stationary increments:

$$
\begin{aligned}
(17) &= e^{-\lambda t}\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}\mathbb{E}_P\left[e^{\left(u\sum_{k=1}^{N(T)-N(t)}\xi_k\right)}\right] \\
&= e^{-\lambda t}\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}\prod_{k=1}^{n}\mathbb{E}_P\left[e^{u\xi_1}\right]^n \quad \text{since } \xi_k \text{ iid} \\
&= e^{-\lambda t}e^{\lambda t\mathbb{E}_P\left[e^{u\xi_k}\right]} \quad \text{by Taylor expansion} \\
&= e^{\lambda t\left(\mathbb{E}_P\left[e^{u\xi_1}\right]-1\right)} = e^{\lambda t\int_{-\infty}^{\infty}(e^{uy}-1)\nu(dy)}
\end{aligned}
$$

where the last equality comes from the fact that $\int_{-\infty}^{\infty}\nu(dy) = 1$. ∎

Now we investigate the change of measure formula for the compound Poisson process. We show below that it simply becomes a multiplication of standard Poisson processes grouped by jumps of the same height. The interpretation is that we are shifting each type of jump individually.

Proposition 5 (CoM for the Compound Poisson Process)**.** Suppose we are given a compound Poisson process of the form (16), with constant intensity λ and jumps $\xi_1, \xi_2, \cdots$ with distribution $f(\xi)$ (measure $\nu(d\xi) := f(\xi)d\xi$) under measure P. Then the Radon-Nikodym derivative used to transform (16) to a compound Poisson process with intensity $\tilde{\lambda}$ and jump distribution $\tilde{f}(\xi)$ (measure $\tilde{\nu}(d\xi) := \tilde{f}(\xi)d\xi$) is given by

$$
Z(t) = e^{(\lambda-\tilde{\lambda})t}\prod_{i=1}^{N(t)}\frac{\tilde{\lambda}\tilde{f}(\xi_i)}{\lambda f(\xi_i)} \tag{18}
$$

Moreover, the process (18) satisfies the SDE

$$
dZ(t) = Z(t-)\left(d\left(\sum_{i=1}^{N(t)}\frac{\tilde{\lambda}\tilde{f}(\xi_i)}{\lambda f(\xi_i)} - \tilde{\lambda}t\right) - d(N(t)-\lambda t)\right) \tag{19}
$$

Proof. Define the two quantities

$$
J(t) := \prod_{i=1}^{N(t)}\frac{\tilde{\lambda}\tilde{f}(\xi_i)}{\lambda f(\xi_i)}, \quad H(t) := \sum_{i=1}^{N(t)}\frac{\tilde{\lambda}\tilde{f}(\xi_i)}{\lambda f(\xi_i)}
$$

Then we can write the recursive relationships

$$
\begin{aligned}
&J(t) = J(t-)\cdot\frac{\tilde{\lambda}\tilde{f}(\xi_{N(t)})}{\lambda f(\xi_{N(t)})} = J(t-)\cdot\frac{\tilde{\lambda}\tilde{f}(\Delta Y(t))}{\lambda f(\Delta Y(t))} \\
\implies \quad &\Delta J(t) = J(t-)\left(\frac{\tilde{\lambda}\tilde{f}(\Delta Y(t))}{\lambda f(\Delta Y(t))} - 1\right)
\end{aligned}
$$

and

$$
H(t) = H(t-) + \frac{\tilde{\lambda}\tilde{f}(\Delta Y(t))}{\lambda f(\Delta Y(t))} \implies \Delta H(t) = \frac{\tilde{\lambda}\tilde{f}(\Delta Y(t))}{\lambda f(\Delta Y(t))}
$$

We then obtain a SDE for $J(t)$ as follows:

$$\Delta J(t) = J(t-)(\Delta H(t) - 1) = J(t-)(\Delta H(t) - \Delta N(t)) \; dJ(t) = J(t-)(dH(t) - dN(t))$$

We can now prove (19) using Itô's formula to get $dZ(t)$, with $Z(t) = e^{(\lambda-\tilde{\lambda})t}J(t)$.

$$\begin{aligned} Z(t) &= Z(0) + \int_0^t (\lambda - \tilde{\lambda})e^{(\lambda-\tilde{\lambda})s}J(s-)ds + \int_0^t e^{(\lambda-\tilde{\lambda})s}dJ(s) \\ &= Z(0) + (\lambda - \tilde{\lambda})\int_0^t \underbrace{e^{(\lambda-\tilde{\lambda})s}J(s-)}_{=:Z(s-)}ds + \int_0^t \underbrace{e^{(\lambda-\tilde{\lambda})s}J(s-)}_{=:Z(s-)}dH(s) - \int_0^t \underbrace{e^{(\lambda-\tilde{\lambda})s}J(s-)}_{=:Z(s-)}dN(s) \end{aligned}$$

and hence, we confirm (19):

$$dZ(t) = Z(t-)d\left((H(t) - \tilde{\lambda}t) - (N(t) - \lambda t)\right)$$

To show that the transformation indeed yields another Poisson process, we use characteristic functions. Define

$$G(t) := e^{uY(t) - \tilde{\lambda}t(\tilde{\phi}_\xi(u)-1)}$$

where $\tilde{\phi}_\xi(u) := \mathbb{E}_Q[e^{u\xi}] = \int_{-\infty}^{\infty} e^{u\xi}\tilde{\nu}(d\xi)$ is the characteristic function of the jumps ξ under measure Q. It is possible to show that $G(t)Z(t)$ is a martingale. This implies

$$\mathbb{E}_P[G(t)Z(t)] = 1 \implies e^{-\tilde{\lambda}t(\tilde{\phi}_\xi(u)-1)}\mathbb{E}_P\left[e^{uY(t)}Z(t)\right] = 1 \implies \mathbb{E}_Q\left[e^{uY(t)}\right] = \mathbb{E}_P\left[e^{uY(t)}Z(t)\right] = e^{\tilde{\lambda}t(\tilde{\phi}_\xi(u)-1)}$$

which shows that the distribution of $Y(t)$ is indeed a compound Poisson with intensity $\tilde{\lambda}$ and jumps $\tilde{f}(\xi)$ under measure Q. ■

4 The Particle Filter

Now we derive optimal estimators for the continuous-time system (5) in the case where $d\mathbf{w}(t) = \sigma(t)dW(t)$ for the standard Brownian motion process $W(t)$. When the dynamics are possibly nonlinear, it is common to use two kinds of filtering techniques: 1) particle filtering, or 2) linearizing the dynamics to take advantage of Gaussian properties. We begin by discussing particle filters in this section. It turns out that we can invoke measure transformation procedures from stochastic process theory to aid with the design of the particle filter.

4.1 Related to Gaussian Change of Measure

For the dynamics (5) with $d\mathbf{w}(t) = \sigma(t)dW(t)$, suppose we additionally keep track of an *importance process* of the form

$$d\mathbf{z}(t) = g(t, \mathbf{z}(t))dt + \theta(t)dW(t) \tag{20}$$

where the Brownian motion process $W(t)$ is the same as that of the state dynamics.

Theorem 2 (Transformation of Solutions via Change of Measure for Gaussian Noise Processes)**.** The process

$$d\mathbf{z}^*(t) = \sigma(t)\theta(t)^{-1}d\mathbf{z}(t), \quad \mathbf{z}(0) = \mathbf{x}_0 \tag{21}$$

yields a weak solution to (5) under the measure Q, where $Z(t)$ is the Radon-Nikodym derivative of the form

$$Z(t) = e^{\int_0^t \overrightarrow{\mu}(s)^T dW(s) + \frac{1}{2}\int_0^t \overrightarrow{\mu}(s)^T \overrightarrow{\mu}(s) ds}$$

Proof. Substitute (20) into (21) to get

$$d\mathbf{z}^*(t) = \sigma(t)\theta^{-1}(t)g(t, \mathbf{z}(t))dt + \sigma(t)dW(t)$$

Then solving for $dW(t)$ yields:

$$dW(t) = \sigma(t)^{-1}d\mathbf{z}^*(t) - \theta^{-1}(t)g(t, \mathbf{z}(t))dt \tag{22}$$

Define

$$\overrightarrow{\mu}(t) := \sigma(t)^{-1}f_c(t, \mathbf{z}^*(t)) - \theta(t)^{-1}g(t, \mathbf{z}(t)) \tag{23}$$

To provide intuition behind why we create such a definition, we take note of its analogy to the simple 1D case described in Example 2. The drift (mean) of the original importance process is $g(t, \mathbf{z}(t))$, but it needs to be transformed to the new mean $f_c(t, \mathbf{z}^*(t))$.

Then we can define the new process as follows

$$d\tilde{W}(t) = dW(t) - \overrightarrow{\mu}(t)dt$$

and we can derive the corresponding form of the Doléans-Dade exponential. Under the transformed measure Q, this noise process is a Brownian motion process. Simplifying the expression by substituting in $\overrightarrow{\mu}(t)$ and (22):

$$\begin{aligned} d\tilde{W}(t) &= dW(t) - \sigma(t)^{-1}f_c(t, \mathbf{z}^*(t))dt + \theta(t)^{-1}g(t, \mathbf{z}(t))dt \\ &= \sigma(t)^{-1}d\mathbf{z}^*(t) - \sigma(t)^{-1}f_c(t, \mathbf{z}^*(t))dt \end{aligned}$$

and rearranging the equation yields the transformed SDE:

$$d\mathbf{z}^*(t) = f_c(t, \mathbf{z}^*(t))dt + \sigma(t)d\tilde{W}(t)$$

This tells us that $\mathbf{z}^*(t)$ is the solution to (5) under the new measure Q. ∎

Example 3 (Linear Case)**.** We can adapt this analysis specially to the case of linear dynamics:

$$\begin{aligned} d\mathbf{x}(t) &= A\mathbf{x}(t)dt + \sigma(t)dW(t), \quad \mathbf{x}(0) = \mathbf{x}_0 \\ \mathbf{y}(t) &= C\mathbf{x}(t) + \mathbf{v}(t) \end{aligned} \tag{24}$$

We take the importance process to be

$$d\mathbf{z}(t) = B\mathbf{z}(t)dt + \theta(t)dW(t)$$

and consider again the transformation

$$d\mathbf{z}^*(t) = \sigma(t)\theta^{-1}(t)d\mathbf{z}(t) = \sigma(t)\theta^{-1}(t)B\mathbf{z}(t)dt + \sigma(t)dW(t)$$

Define again the difference of the two drift terms:

$$\overrightarrow{\mu}(t) = \sigma^{-1}(t)A\mathbf{z}^*(t) - \theta^{-1}(t)B\mathbf{z}(t)$$

Then the new Brownian motion process can be constructed as

$$d\tilde{W}(t) = dW(t) - \overrightarrow{\mu}(t)dt = \sigma^{-1}(t)\left(d\mathbf{z}^*(t) - A\mathbf{z}^*(t)dt\right)$$

and when this is rearranged:

$$d\mathbf{z}^*(t) = A\mathbf{z}^*(t)dt + \sigma(t)d\tilde{W}(t)$$

which implies that $\mathbf{z}^*(t)$ is a solution to (24) under the new measure Q. The corresponding Doléans-Dade exponential is:

$$Z(t) = \exp\left\{\int_0^t \left[\sigma^{-1}(s)A\mathbf{z}^*(s) - \theta^{-1}(s)B\mathbf{z}(s)\right]^T dW(s)\right.$$
$$\left. -\frac{1}{2}\int_0^t \left[\sigma^{-1}(s)A\mathbf{z}^*(s) - \theta^{-1}(s)B\mathbf{z}(s)\right]^T \left[\sigma^{-1}(s)A\mathbf{z}^*(s) - \theta^{-1}(s)B\mathbf{z}(s)\right]\right\}$$

□

4.2 Related to Poisson Change of Measure

Now consider instead, the dynamics (5) with $d\mathbf{w}(t) = \xi(t)dN(t)$. We additionally keep track of an importance process of the form

$$d\mathbf{z}(t) = g(t, \mathbf{z}(t))dt + \alpha(t)dN(t) \tag{25}$$

Note that in contrast to the discussion of Section 3.2, both ξ and α are vector-valued functions that map from $\mathbb{R}^+$ to $\mathbb{R}^{n_x}$.

We then have the following analogous theorem.

Theorem 3 (Transformation of Solutions via Change of Measure for Poisson Noise Processes)**.** The process

$$d\mathbf{z}^*(t) = \xi(t)\alpha(t)^{-1}d\mathbf{z}(t), \quad \mathbf{z}(0) = \mathbf{x}_0 \tag{26}$$

yields a weak solution to (5) under the measure Q, under which the shot noise process is transformed to $\tilde{N}(t)$ with intensity $\tilde{\lambda}$ and jump distribution $\tilde{f}$. The corresponding Radon-Nikodym derivative $Z(t)$ is of the form

$$Z(t) = e^{(\lambda-\tilde{\lambda})t}\prod_{i=1}^{N(t)}\frac{\tilde{\lambda}\tilde{f}(\xi(T_i))}{\lambda f(\xi(T_i))}$$

T_i represents the ith arrival time of the Poisson process, and f and $\tilde{f}$ map vector-valued $\xi(T_i)$ to scalar probability values.

Proof. As in the previous theorem, we substitute (25) into (26):

$$d\mathbf{z}^*(t) = \xi(t)\alpha^{-1}(t)g(t, \mathbf{z}(t))dt + \xi(t)dN(t)$$

Then solving for $dN(t)$ yields:

$$dN(t) = \xi(t)^{-1}d\mathbf{z}^*(t) - \alpha^{-1}(t)g(t, \mathbf{z}(t))dt \tag{27}$$

■

5 The Kalman Filter

In the previous section, we discussed the relationship of probability measure transformations, a well-known concept in the study of stochastic differential equations, to particle filtering. When the dynamics are linear and perturbed by an additive Gaussian noise process, as in Example 3, two additional properties make the filtering approach more straightforward than the techniques we've seen previously. Namely,

- **Affine combinations of Gaussian random vectors are still Gaussian-distributed.** So, if our initial condition $\mathbf{x}(0)$ has a prior distribution which is Gaussian, all future state variables $\mathbf{x}(t)$ (or $\mathbf{x}_k$, in the discrete-time case) will be Gaussian-distributed too.
- **Gaussian distributions are fully characterized by its mean and covariance matrix.** As a consequence of this property, determining the mean and the covariance of the true state is enough to know everything about the full distribution.

With these two properties, state estimation can be achieved by using a type of transformation different from the measure transformation approach seen in Example 3. It is based off of the construction of *innovation processes*. The primary distinction is that the space of measurements $\mathscr{A}(t)$ is the quantity that is changed instead of changing $\mathbf{x}(t)$ via multiplication of the Radon-Nikodym derivative $Z(t)$. This gives rise to the *Kalman filtering algorithm*, which is the focus of this present section.

5.1 Linear Innovations Sequence

Before we derive the Kalman filtering process, we first establish some necessary background about innovation processes, particularly in the case where the conditioned space is linear.

The motivation behind the construction of innovation sequences is as follows: it will often be the case that a new observation $\mathbf{y}_k$ at time k is not totally new if we've already observed previous values $\mathbf{y}_1, \cdots, \mathbf{y}_{k-1}$. The only innovation will come from the component that is orthogonal to the linear span of all the previous observations:

$$\tilde{\mathbf{y}}_k = \mathbf{y}_k - \hat{E}[\mathbf{y}_k|\mathscr{A}_{k-1}] \tag{28}$$

where $\hat{E}$ is the notation derived from (7), and $\mathscr{A}_l$ represents the sigma algebra $\sigma(\mathbf{y}_1, \cdots, \mathbf{y}_l)$ spanned by the random vectors $\mathbf{y}_1, \cdots, \mathbf{y}_l$ for any $l \in \mathbb{Z}$. Furthermore, with this definition, $\mathbb{E}[\tilde{\mathbf{y}}_k] = 0$.

For this simplicity, we will often condition our estimate $\hat{\mathbf{x}}$ around this *linear innovations sequence* $\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \cdots$ as opposed to the original observation sequence $\mathbf{y}_1, \mathbf{y}_2, \cdots$.

Theorem 4. The estimate of $\mathbf{x}$ based on the sequence of orthogonal observations $\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \cdots, \tilde{\mathbf{y}}_n$ is given by the joint projection:

$$\hat{E}[\mathbf{x}|\tilde{\mathscr{A}_n}] = \overline{\mathbf{x}} + \sum_{i=1}^{n} \hat{E}[\mathbf{x} - \overline{\mathbf{x}}|\tilde{\mathbf{y}}_i] \tag{29}$$

where we denote $\tilde{\mathscr{A}_n} = \sigma(\tilde{\mathbf{y}}_1, \cdots, \tilde{\mathbf{y}}_n)$ and $\overline{\mathbf{x}} := \mathbb{E}[\mathbf{x}]$.

Refer to [16] for the proof.

Remark 3. Although both the innovation and observation sequences span the same space ($\tilde{\mathscr{A}_n} = \mathscr{A}_n$), the elements of the innovations sequence are all orthogonal to each other. One might note the similarity of this construction process to the Gram-Schmidt orthonormalization process used in the context of linear algebra. $\square$

Remark 4. We can simplify the expression (29) as follows:

$$\begin{aligned} \hat{E}[\mathbf{x}|\tilde{\mathscr{A}_n}] = \overline{\mathbf{x}} + \sum_{i=1}^{n} \hat{E}[\mathbf{x} - \overline{\mathbf{x}}|\tilde{\mathbf{y}}_i] &= \overline{\mathbf{x}} + \sum_{i=1}^{n} \mathrm{Cov}(\mathbf{x}, \tilde{\mathbf{y}}_i)\mathrm{Cov}(\tilde{\mathbf{y}}_i)^{-1}\tilde{\mathbf{y}}_i \\ &= \hat{E}[\mathbf{x}|\tilde{\mathscr{A}}_{n-1}] + \hat{E}[\mathbf{x} - \overline{\mathbf{x}}|\tilde{\mathbf{y}}_n] \end{aligned} \tag{30}$$

This gives us a recursive formula in terms of each new observation $\tilde{\mathbf{y}}_n$. $\square$

5.2 The Discrete-Time Case

For the sake of simplicity, we will initially consider the case where there is no control input.

$$\mathbf{x}_{k+1} = A_k\mathbf{x}_k + \mathbf{w}_k \tag{31a}$$

$$\mathbf{y}_k = C_k\mathbf{x}_k + \mathbf{v}_k \tag{31b}$$

where $\mathbf{x}_k, \mathbf{w}_k \in \mathbb{R}^n, \mathbf{y}_k, \mathbf{v}_k \in \mathbb{R}^m$, and $A_k \in \mathbb{R}^{n\times n}, C_k \in \mathbb{R}^{m\times n}$ are known for all $k > 0$.

We make the following assumptions: $\mathbf{x}_0, \mathbf{v}_0, \mathbf{v}_1, \cdots, \mathbf{w}_0, \mathbf{w}_1, \cdots$ are all pairwise uncorrelated Gaussian-distributed random variables, and $\mathbb{E}[\mathbf{w}_k] = 0, \mathrm{Cov}(\mathbf{w}_k) = Q_k, \mathbb{E}[\mathbf{v}_k] = 0, \mathrm{Cov}(\mathbf{v}_k) = R_k$ are known constant matrices. Further assume that $\mathbf{x}_0$ comes from a known Gaussian distribution with $\mathbb{E}[\mathbf{x}_0] = \bar{x}_0, \mathrm{Cov}(\mathbf{x}_0) = P_0$, and that it is pairwise uncorrelated with $\mathbf{w}_k$ and $\mathbf{v}_k$ for all k. Denote $\overline{\mathbf{x}}_k = \mathbb{E}[\mathbf{x}_k]$ and $P_k = \mathrm{Cov}(\mathbf{x}_k)$. These quantities are recursively determined for $k \geq 1$ by:

$$\begin{aligned} \mathbb{E}[\mathbf{x}_{k+1}] = A_k\mathbb{E}[\mathbf{x}_k] &\implies \overline{\mathbf{x}}_{k+1} = A_k\overline{\mathbf{x}}_k \\ P_{k+1} &= A_kP_kA_k^T + Q_k \end{aligned}$$

As before, we will define $\mathscr{A}_k = \sigma(\mathbf{y}_0, \mathbf{y}_1, \cdots, \mathbf{y}_k)$ to represent the observations up until time k. Then we define $\hat{\mathbf{x}}_{i|j} := \hat{E}[\mathbf{x}_i|\mathscr{A}_j]$ for nonnegative integers i, j, where $\hat{E}[\mathbf{x}_i|\mathscr{A}_j]$ is the linear MMSE given by

$$\begin{aligned} \hat{E}[\mathbf{x}_i|\mathbf{y}_j] &= \mathbb{E}[\mathbf{x}_i] + \mathrm{Cov}(\mathbf{x}_i, \mathbf{y}_j)\mathrm{Cov}(\mathbf{y}_j)^{-1}(\mathbf{y}_j - \mathbb{E}[\mathbf{y}_j]) \\ \implies \hat{E}[\mathbf{x}_i|\mathscr{A}_j] &= \hat{E}[\mathbf{x}_i|\mathscr{A}_{j-1}] + \mathrm{Cov}(\mathbf{x}_i, \mathbf{y}_j)\mathrm{Cov}(\mathbf{y}_j)^{-1}(\mathbf{y}_j - \mathbb{E}[\mathbf{y}_j]) \text{ by Remark 4} \end{aligned}$$

Finally, denote the associated covariance of error matrices $\Sigma_{i|j} := \mathrm{Cov}(\mathbf{x}_i - \hat{\mathbf{x}}_{i|j})$ for nonnegative integers i, j.

The goal is to compute an estimate of $\mathbf{x}_k$ at each timestep k. We will do this by deriving a recursive relationship between successive state estimates $\hat{\mathbf{x}}_{k-1|k-1}$ and $\hat{\mathbf{x}}_{k|k}$.

The filtering process takes the same two steps as in the Bayesian framework from Section 2:

1. **Prediction**: we predict the value of $\hat{\mathbf{x}}_{k|k-1}$ given $\hat{\mathbf{x}}_{k-1|k-1}$. To do this, directly use the equations **(*)** and the fact that the noise random variables are uncorrelated from all the system variables. Equations yield:

$$\hat{\mathbf{x}}_{k|k-1} = A_{k-1}\hat{\mathbf{x}}_{k-1|k-1} \tag{32a}$$

$$\Sigma_{k|k-1} = A_{k-1}\Sigma_{k-1|k-1}A_{k-1}^T + Q_{k-1} \tag{32b}$$

2. **Measurement Update**: we modify our prediction from $\hat{\mathbf{x}}_{k|k-1}$ to $\hat{\mathbf{x}}_{k|k}$ in order to take into account a new observation $\mathbf{y}_k$. Because we are able to predict a part of $\mathbf{y}_k$ through the linear MMSE, the only innovation comes from the orthogonal component $\tilde{\mathbf{y}}_k = \mathbf{y}_k - \hat{E}[\mathbf{y}_k|\mathscr{A}_{k-1}]$. Alternatively written, this is:

$$\tilde{\mathbf{y}}_k = \mathbf{y}_k - C_k\hat{\mathbf{x}}_{k|k-1} \tag{33}$$

Derive $\hat{\mathbf{x}}_{k|k}$ from $\hat{\mathbf{x}}_{k|k-1}$:

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathrm{Cov}(\mathbf{x}_k, \tilde{\mathbf{y}}^k)\mathrm{Cov}(\tilde{\mathbf{y}}_k)^{-1}\tilde{\mathbf{y}}_k \tag{34}$$

Furthermore, the covariance of error is updated:

$$\Sigma_{k|k} = \Sigma_{k|k-1} - \mathrm{Cov}(\mathbf{x}_k, \tilde{\mathbf{y}}_k)\mathrm{Cov}(\tilde{\mathbf{y}}_k)^{-1}\mathrm{Cov}(\tilde{\mathbf{y}}_k, \mathbf{x}_k) \tag{35}$$

Intuitively, the use of the new observation $\tilde{\mathbf{y}}_k$ reduces the covariance of error for predicting $\mathbf{x}_k$ from $\Sigma_{k|k-1}$ by the covariance matrix of the innovative part of the estimator.

Let us define the gain $L_k := \mathrm{Cov}(\mathbf{x}_k, \tilde{\mathbf{y}}^k)\mathrm{Cov}(\tilde{\mathbf{y}}_k)^{-1}$ so that we can simplify the information update equations as:

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + L_k\tilde{\mathbf{y}}_k \tag{36a}$$

$$\Sigma_{k|k} = \Sigma_{k|k-1} - L_k\mathrm{Cov}(\tilde{\mathbf{y}}_k, \mathbf{x}_k) \tag{36b}$$

We can further calculate

$$\mathrm{Cov}(\mathbf{x}_k, \tilde{\mathbf{y}}_k) = \Sigma_{k|k-1}C_k^T$$
$$\mathrm{Cov}(\tilde{\mathbf{y}}_k) = C_k\Sigma_{k|k-1}C_k^T + R_k$$

so that:

$$L_k = \Sigma_{k|k-1}C_k^T(C_k\Sigma_{k|k-1}C_k^T + R_k)^{-1} \tag{37}$$

Combining the results of the information update together with the time update, we obtain our final equations:

$$\hat{\mathbf{x}}_{k|k} = A_{k-1}\hat{\mathbf{x}}_{k-1|k-1} + L_k\tilde{\mathbf{y}}_k \tag{38a}$$

$$\Sigma_{k|k} = \Sigma_{k|k-1} - L_kC_k\Sigma_{k|k-1} = (I - L_kC_k)(A_{k-1}\Sigma_{k-1|k-1}A_{k-1}^T + Q_{k-1}) \tag{38b}$$

5.3 The Continuous-Time Case

If we extend the DTKF to the setting of continuous time, we derive the famous *Kalman-Bucy filter*. Like its discrete-time analog, the continuous-time Kalman-Bucy filter is widely used for many applications in the field of control theory and signal processing. Detailed treatments of the Kalman-Bucy filter can be found in standard texts such as Chapter 9 of [17] and Chapter 6 of [15].

Now consider the continuous-time dynamics

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B_w\mathbf{w}(t) \tag{CT}$$

$$\mathbf{y}(t) = C\mathbf{x}(t) + \mathbf{v}(t) \tag{39}$$

where $\mathbf{x}(t), \mathbf{w}(t) \in \mathbb{R}^n, \mathbf{y}(t), \mathbf{v}(t) \in \mathbb{R}^m$, and $A, B_w \in \mathbb{R}^{n\times n}, C \in \mathbb{R}^{m\times n}$ are known for all $t > 0$.

As in the discrete-time case, we make the following assumptions: $\mathbf{x}(0) := \mathbf{x}_0$, $\{\mathbf{w}(t)\}$, and $\{\mathbf{v}(t)\}$ are pairwise uncorrelated for all $t > 0$, and $\mathbb{E}[\mathbf{w}(t)] = 0, \mathrm{Cov}(\mathbf{w}(s), \mathbf{w}(t)) = \mathbb{E}[\mathbf{w}(s)\mathbf{w}(t)] = Q\delta(t-s), \mathbb{E}[\mathbf{v}(t)] = 0, \mathrm{Cov}(\mathbf{v}(s), \mathbf{v}(t)) = R\delta(t-s)$, where Q and R are known constant matrices. Further assume that $\mathbf{x}_0$ comes from a known Gaussian distribution with $\mathbb{E}[\mathbf{x}_0] = 0$ (assumed for simplicity), $\mathrm{Cov}(\mathbf{x}_0) = \Sigma_0$. We will define $\mathscr{A}(t) = \sigma\{\mathbf{y}(s) \ : \ 0 \leq s < t\}$ to represent the observations made until time t.

The goal is to estimate $\hat{\mathbf{x}}(t)$ of $\mathbf{x}(t)$ given the observations $\mathscr{A}(t)$ such that the MSE:

$$J := \mathbb{E}\left[\mathrm{tr}((\mathbf{x}(t) - \hat{\mathbf{x}}(t))(\mathbf{x}(t) - \hat{\mathbf{x}}(t))^T)\right] \tag{40}$$

is minimized. In analogue to the DTKF, the MMSE $\hat{\mathbf{x}}(t)$ is given by $\mathbb{E}[\mathbf{x}(t) \mid \mathscr{A}(t)]$. This implies that $\hat{\mathbf{x}}(0) = \mathbb{E}[\mathbf{x}_0] = 0$. Additionally, one can derive the dynamics:

$$\dot{\hat{\mathbf{x}}}(t) = A\hat{\mathbf{x}}(t) + K(\mathbf{y} - C\hat{\mathbf{x}}(t))$$

For the moment, we will take this as given, and use it to derive the covariance equation and the optimal Kalman filter gain.

Define the error vector $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$. Its dynamics are given by:

$$\dot{\mathbf{e}} = \dot{\mathbf{x}} - \dot{\hat{\mathbf{x}}} = A\mathbf{x} + B_w\mathbf{w} - A\hat{\mathbf{x}} - K(\mathbf{y} - C\hat{\mathbf{x}}) = (A - KC)\mathbf{e} + B_w\mathbf{w} - K\underbrace{(\mathbf{y} - C\mathbf{x})}_{\mathbf{v}}$$

Define the error covariance $\Sigma := \mathbb{E}[\mathbf{e}\mathbf{e}^T]$. Then note that the MSE at time t is exactly $\mathrm{tr}(P)$. Derive the dynamics of Σ as follows:

$$\begin{aligned}
\dot{\Sigma} &= \mathbb{E}[\dot{\mathbf{e}}\mathbf{e}^T + \mathbf{e}\dot{\mathbf{e}}^T] \\
&= \mathbb{E}[(A - KC)\mathbf{e}\mathbf{e}^T + \mathbf{e}\mathbf{e}^T(A^T - C^TK^T)] + \mathbb{E}[B_w\mathbf{w}\mathbf{e}^T + \mathbf{e}\mathbf{w}^TB_w^T] + \mathbb{E}[-K\mathbf{v}\mathbf{e}^T - \mathbf{e}\mathbf{v}^TK] \\
&= (A - KC)\Sigma + \Sigma(A^T - C^TK^T) + \mathbb{E}[B_w\mathbf{w}\mathbf{e}^T + \mathbf{e}\mathbf{w}^TB_w^T] + \mathbb{E}[-K\mathbf{v}\mathbf{e}^T - \mathbf{e}\mathbf{v}^TK^T]
\end{aligned} \tag{41}$$

Denote the state transition matrix $\Phi(t_0, t) := e^{(A-KC)(t-t_0)}$. Then we can write $\mathbf{e}(t)$ as:

$$\mathbf{e}(t) = \Phi(0, t)\mathbf{e}_0 + \int_0^t \Phi(s, t)B_w\mathbf{w}(s)ds - \int_0^t \Phi(s, t)K\mathbf{v}(s)ds \tag{42}$$

which implies that

$$\begin{aligned}\mathbb{E}[\mathbf{e}\mathbf{w}^T(t)B_w^T] &= \Phi(0,t)\mathbb{E}[\mathbf{e}_0\mathbf{w}^T]B_w^T + \int_0^t \Phi(s,t)B_w\mathbb{E}[\mathbf{w}(s)\mathbf{w}^T(t)]B_w^T ds - \int_0^t \Phi(s,t)K\mathbb{E}[\mathbf{w}(s)\mathbf{v}^T(t)]B_w^T ds \\ &= \int_0^t \Phi(s,t)B_w Q\delta(t-s)B_w^T ds \\ &= \frac{1}{2}B_w Q B_w^T \text{ since } \Phi(t,t) = I\end{aligned}$$

One can make a symmetric argument for $\mathbb{E}[B_w\mathbf{w}\mathbf{e}^T] = \frac{1}{2}B_wQB_w^T$.

For the last line, we also have the following formula:

$$\int_a^b f(x)\delta(b-x)dx = \int_{b-a}^0 f(b-u)\delta(u)du = \frac{1}{2}f(b)$$

Informally speaking, since the delta function occurs at one of the endpoints of the integral, only half the total weight gets integrated over.

Similarly, (42) implies that $\mathbb{E}[-\mathbf{e}\mathbf{v}^T(t)K^T] = \mathbb{E}[-K\mathbf{v}\mathbf{e}^T] = (1/2)KRK^T$ since the minus signs cancel.

Substituting everything back int (41) yields:

$$\begin{aligned}\dot{\Sigma} &= (A-KC)\Sigma + \Sigma(A^T - C^TK^T) + B_wQB_w^T + KRK^T \\ &= A\Sigma + \Sigma A^T + B_wQB_w^T - KC\Sigma - \Sigma C^TK^T + KRK^T \qquad (43) \\ &= A\Sigma + \Sigma A^T + B_wQB_w^T + (KR - \Sigma C^T)R^{-1}(KR-\Sigma C^T)^T - \Sigma C^T RC\Sigma \qquad (44)\end{aligned}$$

where the last line follows from completing the square (exercise). To minimize $J = \mathrm{tr}(\Sigma(t))$, we can minimize $\Sigma(t)$ by choosing K so that $\dot{P}(t)$ decreases by the maximum amount possible at time t. This happens when the term $(KR - \Sigma C^T)R^{-1}(KR + \Sigma C^T)^T$ is equal to 0, since it can never be negative (like a square term).

This implies:

$$K(t)R - \Sigma(t)C^T = 0 \Longrightarrow K(t) = \Sigma(t)C^TR^{-1}$$

Substituting this back into (43) yields the final covariance equation.

6 The CTKBF for Lévy Noise Systems

[8] considers a type of single-target Kalman filtering in a setting where both the system dynamics and the observation process were perturbed by Lévy noise processes. We will describe this approach to filtering in this section.

The system process is defined as

$$d\mathbf{x}(t) = A(t)\mathbf{x}(t-)dt + B(t)dL_1(t), \quad \mathbf{x}(0) = \mathbf{x}_0 \qquad (45)$$

with $\mathbf{x}(t) \in \mathbb{R}^n$, while the corresponding observation process is

$$d\mathbf{y}(t) = C(t)\mathbf{x}(t)dt + D(t)dL_2(t), \quad \mathbf{y}(0) = 0 \text{ a.s.} \qquad (46)$$

with $\mathbf{y}(t) \in \mathbb{R}^m$, where L_1, L_2 are two independent Lévy processes on the probability space Ω, taking values in $\mathbb{R}^{\ell_1}$ and $\mathbb{R}^{\ell_2}$. Note that the solution to the system dynamics can be easily determined by linearity

$$\mathbf{x}(t) = e^{A(s)ds}\mathbf{x}(0) + \int_0^t e^{\int_s^t A(u)du} B(s) dL_1(s) \tag{47}$$

Unlike the Hilbert space case, the Orthogonality principle cannot be applied in the proof of optimality of a filter. This is because the setting is now in the more general Banach space. Nevertheless, a statement of optimality can be given.

For a specific given square-integrable stochastic process $\mathbf{z} := \{\mathbf{z}(s)\}$ in $\mathbb{R}^m$, denote $\mathcal{P}_z$ to mean the operator for the orthogonal projection from $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{R}^n)$, abbreviated $\mathcal{L}^2$, to the space

$$\mathcal{L}(\mathbf{z}) := \{c_0 + c_1\mathbf{z}(t_1) + \cdots c_M\mathbf{z}(t_M) | c_0 \in \mathbb{R}^n, c_i \in \mathbb{R}^{n \times m} i \geq 1\} \subseteq \mathcal{L}^2 \tag{48}$$

Define $\hat{\mathbf{x}}(t) := \mathcal{P}_y(\mathbf{x}(t))$ to be the *best linear estimator* of $\mathbf{x}(t)$ given the observation process $\mathbf{y}$. Further define the *innovations process* $\{\tilde{\mathbf{y}}(t), 0 \leq t \leq T\}$ to be

$$\tilde{\mathbf{y}}(t) := \mathbf{y}(t) - \int_0^t C(s)\hat{\mathbf{x}}(s-)ds \tag{49}$$

As seen before, note that $\tilde{\mathbf{y}}$ is a process with orthogonal increments.

Now define the process $\{\mathbf{z}(s)\} \subseteq \mathbb{R}^m$ to be the unique solution to the SDE

$$d\mathbf{z}(t) = G(t)d\tilde{\mathbf{y}}(t), \quad \mathbf{z}(0) = 0 \text{ a.s.} \tag{50}$$

where $G(t) := (D(t)D(t)^T)^{-1/2}$ for all $0 \leq t \leq T$. Note that $\mathbf{z}$ is centered and has orthogonal increments with covariance

$$K_2(s \wedge t) := \mathbb{E}\left[\mathbf{z}(s)\mathbf{z}(t)^T\right] := \int_0^{s \wedge t} G(u)D(u)\Lambda_2 D(u)^T G(u)^T du \tag{51}$$

where Λ_2 is the covariance matrix corresponding to the jump part of L_2.

This gives us an alternative way of representing $\hat{\mathbf{x}}$, similar to (7):

$$\hat{\mathbf{x}}(t) = \mathbb{E}[\mathbf{x}(t)] + \int_0^t \frac{\partial}{\partial s}\mathbb{E}\left[\mathbf{x}(s)\mathbf{z}(s)^T\right] K_2(s)^{-1} ds \tag{52}$$

We now define the *mean-squared error* (MSE) matrix as

$$\Sigma(t) = \mathbb{E}\left[(\mathbf{x}(t) - \hat{\mathbf{x}}(t))(\mathbf{x}(t) - \hat{\mathbf{x}}(t))\right], \; 0 \leq t \leq T \tag{53}$$

The MSE satisfies the Riccati equation as

$$\begin{aligned} \frac{d\Sigma(t)}{dt} &= A(t)\Sigma(t) + \Sigma(t)A(t)^T + B(t)\Lambda_1 B(t)^T - \Sigma(t)C(t)^T G(t)^T K_2(t)^{-1} G(t)C(t)\Sigma(t)^T \\ \Sigma(0) &= \mathrm{Cov}(\mathbf{x}_0) \end{aligned} \tag{54}$$

The *Kalman-Bucy filter* is then given as follows. The linear estimate $\hat{\mathbf{x}}(t)$ solves the SDE

$$d\hat{\mathbf{x}}(t) = A(t)\hat{\mathbf{x}}(t)dt + \Sigma(t)C(t)^T G(t)^T K_2(t)^{-1} G(t)(d\mathbf{y}(t) - C(t)\hat{\mathbf{x}}(t)dt) \tag{55}$$

In the case where the Lévy process L_2 corresponding to the observation process (46) is not square-integrable, we can take a sequence of square-integrable processes to the limit as $k \to \infty$.

Set up a sequence of observation processes which abide by the following SDEs

$$d\mathbf{y}_k(t) = C(t)\mathbf{x}(t-)dt + D(t)dL_2^{(k)}(t), \quad k \in \mathbb{N} \tag{56}$$

Proposition 6. The sequence $\{\mathbf{y}_k(t)\}_{k\in\mathbb{N}}$ converges to $\mathbf{y}(t)$ for all $t \geq 0$ in the $\mathcal{L}^1$ sense.

Define the following space in a way similar to (48)

$$\mathcal{L}(\mathbf{z}_k) := \{c_0 + c_1\mathbf{z}_k(t_1) + \cdots c_M\mathbf{z}_k(t_M) | c_0 \in \mathbb{R}^n, c_i \in \mathbb{R}^{n\times m} i \geq 1\} \subseteq \mathcal{L}^2 \tag{57}$$

and $\mathcal{P}_k$ to be orthogonal projection from $\mathcal{L}^2$ to $\mathcal{L}(\mathbf{z}_k)$. Write $\hat{\mathbf{x}}_k(t) := \mathcal{P}_k(\mathbf{x}(t))$ to be the projection of the true state onto the sequence of observations $\{\mathbf{y}_k\}_{k\in\mathbb{N}}$.

The MSE of $\hat{\mathbf{x}_k(t)}$ is denoted by $\Sigma_k(t)$ and defined similarly to (53). We will assume the sequence $\{K_s^{(k)}(t)^{-1}\}$ converges to $\Phi(t)$.

References

[1] J. Yang, A. Dani, S.-J. Chung, and S. Hutchinson, “Vision-based localization and robot-centric mapping in riverine environments,” *J. Field Robot*, vol. 34, no. 3, pp. 429–450, 2017.

[2] S. Bandyopadhyay and S.-J. Chung, “Distributed bayesian filtering using logarithmic opinion pool for dynamic sensor networks,” *Automatica*, vol. 97, pp. 7–17, Nov 2018.

[3] A. Patel and B. Kosko, “Stochastic resonance in continuous and spiking neuron models with Lévy noise,” *IEEE Transactions on Neural Networks*, vol. 19, pp. 1993–2008, Jan 2009.

[4] R. Cont and P. Tankov, *Financial Modelling With Jump Processes*. London, UK: CRC Press UK, 2004.

[5] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[6] K. Reif, S. Gunther, E. Yaz, and R. Unbehauen, “Stochastic stability of the discrete-time extended Kalman filter,” *IEEE Transactions on Automatic Control*, vol. 44, no. 4, pp. 714–728, Apr 1999.

[7] E. A. Wan and R. Van Der Merwe, “The unscented Kalman filter for nonlinear estimation,” in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*, Oct 2000, pp. 153–158.

[8] D. Applebaum and S. Blackwood, “The Kalman-Bucy filter for integrable Lévy processes with infinite second moment,” *ArXiv preprint, arXiv:1306.5103*, Apr 2014.

[9] R. Olfati-Saber, “Distributed Kalman filtering for sensor networks,” in *2007 46th IEEE Conference on Decision and Control*, Dec 2007, pp. 5492–5498.

[10] ——, “Kalman-consensus filter : Optimality, stability, and performance,” in *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, Dec 2009, pp. 7036–7042.

[11] M. S. Arulampalam, S. Maskell, and N. Gordon, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, pp. 174–188, 2002.

[12] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: fifteen years later," 2011.

[13] T. Yang, P. G. Mehta, and S. P. Meyn, "Feedback particle filter with mean-field coupling," in *IEEE Conference on Decision and Control and European Control Conference*, Dec 2011, pp. 7909–7916.

[14] H. Qiao and J. Duan, "Nonlinear filtering of stochastic dynamical systems with Lévy noises," *Advances in Applied Probability*, vol. 47, no. 3, p. 902–918, 2015.

[15] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, ser. Universitext. Berlin, Germany: Springer Berlin, 2010. [Online]. Available: https://books.google.com/books?id=EQZEAAAAQBAJ

[16] B. Hajek, "Notes for ECE534: An exploration of random processes for engineers," 2006. [Online]. Available: http://www.ifp.uiuc.edu/~hajek/Papers/randomprocJuly06.pdf

[17] J. Xiong, *An Introduction to Stochastic Filtering Theory.* Oxford University Press, 2008.