Introduction to Filtering Theory with Applications to Finance

Thesis · August 2008

DOI: 10.13140/RG.2.2.22620.28806

1 author:

Mohamed Abdelghani
Morgan Stanley

Introduction to Filtering Theory with Applications to Finance

by

Mohamed Abdelghani

Submitted to the Department of Mathematics and Statistics
in partial fulfillment of the requirements for the degree of

Master of Art

at the

YORK UNIVERSITY

August 2008

Signature of Author ....

Department of Mathematics and Statistics
29 July 2008

Certified by ....

Tom Salisbury
Professor, Stochastic calculus and Financial mathematics
Thesis Supervisor

Accepted by ....

Neal Madras
Chairperson, Department Committee on Graduate Students

Introduction to Filtering Theory with Applications to Finance

by

Mohamed Abdelghani

Submitted to the Department of Mathematics and Statistics
on 29 July 2008, in partial fulfillment of the
requirements for the degree of
Master of Art

Abstract

I survey the development of filtering theory from its inception at the hands of Kolmogorov and Wiener to its current form. Also, I will discuss approximation methods for solving equations of nonlinear filters and some applications of filtering to finance.

Thesis Supervisor: Tom Salisbury
Title: Professor, Stochastic calculus and Financial mathematics

Chapter 1

Introduction

Filtering theory is concerned with the estimation of an unknown signal process X_t from a related known process Y_t. It started with the method of least squares estimation first used by Gauss. Kolmogorov [35] and Wiener [57] extended least square method to estimation in stochastic processes. Initially Kolmogorov and Wiener worked on the prediction problem. Later, Wiener [57] formulated the filtering problem; for it had many engineering applications.

Wiener formulation

Wiener considered estimating a signal process X_t that is corrupted by white noise W_t. X_t and W_t are uncorrelated. $Y_t = X_t + W_t$ is the observed signal during the time interval $t_0 \leq t \leq t_f$. The optimum linear least square estimate of X_t is give by

$$\widehat{X}_t = \int_{t_0}^{t_f} H(t,\tau) Y_\tau d\tau \tag{1.1}$$

which is the minimum of $E\{(X_t - Z)'(X_t - Z)\}$ where Z is square integrable and measurable with respect to the filtration generated by Y_t, $t_0 \leq t \leq t_f$. $H(t,\tau)$ is the filter transfer function which is determined by the orthogonality property $E\{(X_t - \widehat{X}_t)Y_s'\} = 0$ for all $t_0 \leq s \leq t_f$. A simple calculation shows that the optimum filter $H(\cdot,\cdot)$ is determined by solving the equation

$$H(t,s) + \int_{t_0}^{t_f} H(t,\tau) K(\tau,s) d\tau = E\{X_t X_s' + X_t W_s'\} \tag{1.2}$$

for $t_0 \leq s \leq t_f$ and $t = t_f$. Note that if $t > t_f$ we get the optimal prediction problem and if $t < t_f$ we get the smoothing problem. $K(t,s) = E\{X_t X_s' + X_t W_s' + W_t X_s'\}$ is the kernel of integration. If we assume that the future values of the signal X_t is uncorrelated with past noise W_s $(s \leq t)$ then we can write equation 1.2 as follows

$$H(t,s) + \int_{t_0}^{t_f} H(t,\tau)K(\tau,s)d\tau = K(t,s). \tag{1.3}$$

Equation 1.3 is known as the Fredholm integral equation of the second kind and is more difficult to solve than 1.2 because of the added constraint $(s \leq t)$. Wiener looked at a particular case of equation 1.3 when the observation is a scalar process under a semi-infinite observation interval $(t_0 = -\infty, t_f = 0)$ with jointly stationary signal and noise processes. He discovered what was known for physicist as the Hopf equation

$$H(t) + \int_0^{\infty} H(\tau)K(t-\tau)d\tau = K(t) \qquad 0 < t < \infty \text{ and } s = 0 \tag{1.4}$$

where $K(t-\tau) = \sum_{i=1}^{n} \alpha_i \exp(-\beta_i |t-\tau|)$. Wiener and Hopf [58] solved for $H(t)$ using spectral decomposition. Soon after, Wiener-Hopf equation was extended to the estimation of stationary processes over a finite observation interval $(t_0 < \infty)$, and nonstationary processes (see survey paper by Kaliath [36]).

After all of the work done on the Wiener formulation there was not a general method for solving 1.3 but many complicated techniques and special results. Because of these difficulties, Kalman changed the formulation of the filtering problem.

Kalman formulation

Kalman [37], [38], [39], simplified the problem by considering a particular *model* for the signal and observation processes. He assumed that the signal and observation can be described by

$$dX_t = a(t)X_t dt + b(t)dB_t, \tag{1.5}$$

$$dY_t = c(t)X_t dt + dW_t \tag{1.6}$$

where a, b, c, and d are known functions. $E\{X_0 X_0'\}$ a constant initial state variance. B_t and W_t are independent and jointly Gaussian. $E\{B_t B_s'\} = E\{W_t W_s'\} = t \wedge s$, and $E\{B_t W_s'\} = E\{B_t X_0'\} = E\{W_t X_0'\} = 0$. The estimated state is given by

$$d\widehat{X}_t = a(t)\widehat{X}_t dt + k(t)(dY_t - c(t)\widehat{X}_t dt) \tag{1.7}$$

where $k(t) = r(t)c'(t)$ and $r(t)$ is the estimate of the error covariance $(X_t - \widehat{X}_t)(X_t - \widehat{X}_t)'$. $r(t)$ is computed by the Riccati equation

$$\dot{r}(t) = a(t)r(t) + r(t)a'(t) - k(t)k'(t) + b(t)b'(t). \tag{1.8}$$

The Kalman filter was successful in many applications but limited by assumption that the signal and observation equations are linear. To deal with nonlinear filtering problems which are the most common a different approach was developed.

History of nonlinear formulation

For the estimation of a nonlinear signal process corrupted by any type of noise from observations it is not enough to find the mean and covariance of the signal as is done in the Kalman filter. In fact, all the moments are needed or the conditional probability of the signal given the observations.

The nonlinear filtering problem has gone through decades of development. Stratonovich [53] and Kushner [40] formulated a nonlinear stochastic partial differential equation for the evolution of the conditional probability density of the signal process given the observation for a Markov process. It was later generalized by Fujisaka, Kunita, and Kallianpur [27]. The reference measure approach to nonlinear filtering was developed by Mortensen [50], Duncan [23] and Zakai [60]. The reference measure approach lead to Duncan–Mortensen–Zakai (DMZ) equation for nonlinear filtering of diffusion processes. It is a linear stochastic partial differential equation which describes the evolution of the unormalized conditional probability density of the signal process given the observations. The robust version of the DMZ equation which uses the robust transform was developed by Clark [12] and Davis [20]. It reduced the DMZ equation to an ordinary PDE with random coefficients.

While the linear filtering problem lead to finite dimensional statistics (*e.g.* mean and covariance estimators) the nonlinear filtering problem is often infinite dimensional; the conditional statistics of the signal could depend on the full history of the observation process. The infinite dimensionality of the conditional density has posed a problem for any practical use of the nonlinear filtering approach. This has motivated the use of algebraic methods in nonlinear filtering to find conditions whereby a finite dimensional recursive filter exits. Brockett [6], [8] and Brockett and Clark [9] were the first to apply nonlinear system theory and Lie algebras to nonlinear estimation problems. Mitter [48], [49] emphasized functional integration and group representations in nonlinear filtering. These methods lead to a better understanding of the structure of nonlinear filtering problems, provided guidance in the search for finite dimensional filters, the classification of filtering problems, and the design of useful approximate filters [46].

Below is a table showing the main developments of the filtering problem [10].

Author(s) (year)	Method	Solution	Comment
Kolmogorov (1941)	innovations	exact	linear, stationary
Wiener (1942)	spectral factorization	exact	linear, stationary, infinite memory
Levinson (1947)	lattice filter	approximate	linear, stationary, finite memory
Bode & Shannon (1950)	innovations, whitening	exact	linear, stationary
Zadeh & Ragazzini (1950)	innovations, whitening	exact	linear, non-stationary
Kalman (1960)	orthogonal projection	exact	LQG, non-stationary, discrete
Kalman & Bucy (1961)	recursive Riccati equation	exact	LQG, non-stationary, continuous
Stratonovich (1960)	conditional Markov process	exact	nonlinear, non-stationary
Kushner (1967)	PDE	exact	nonlinear, non-stationary
Zakai (1969)	PDE	exact	nonlinear, non-stationary
Handschin & Mayne (1969)	Monte Carlo	approximate	nonlinear, non-Gaussian, non-stationary
Bucy & Senne (1971)	point-mass, Bayes	approximate	nonlinear, non-Gaussian, non-stationary
Kailath (1971)	innovations	exact	linear, non-Gaussian, non-stationary
Bene's (1981)	Bene's	exact solution of Zakai eqn.	nonlinear, finite-dimensional
Daum (1986)	Daum, virtual measurement	exact solution of FPK eqn.	nonlinear, finite-dimensional
Gordon, Salmond, & Smith (1993)	bootstrap, sequential Monte Carlo	approximate	nonlinear, non-Gaussian, non-stationary
Julier & Uhlmann (1997)	unscented transformation	approximate	nonlinear, non-Gaussian, derivative-free

Outline

In chapter 2, I will derive the general filtering problem in probability measure setup, the reference measure approach, the Fujisaki, Kallianpur and Kunita (FKK) nonlinear filtering PDE for Markov processes and the derivation of the DMZ or Zakai equation for unormalized density. I will also describe some finite dimensional filters and Lie algebraic methods for nonlinear filtering. In chapter 3, I will describe a number of numerical approximation methods for solving the filtering problem particle filters, spectral decomposition and neural network methods. Finally in chapter 4, I will present some applications of the nonlinear filtering to Finance.

Chapter 2

Filtering Theory

There are many possible filtering problems. They result from the different types of signals, observation processes, how the signal and observation relate, and the different types of underlying noise affecting the signal and observation. For example, the signal process may be continuous but the observation process discrete, the signal and observation are discrete, the signal or observation processes are not Markovian, the underlying disturbances are not Brownian but jump diffusion. In the next section, I will only consider the special case of nonlinear Markov processes closely following the presentation given by Davis [20].

2.1 Filtering of Markov Processes

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{F}_t \subset \mathcal{F}$ be a filtration in $\mathcal{F}$. Suppose Φ_t is a *signal process* in $\mathcal{F}_t$ which we can not observe. But, we are able to *observe* a *related* process Y_t. Our aim is to get the *best estimate* of Φ_t given the history of Y_t up to time t.

The signal process Φ_t is a semimartingale with respect to $\mathcal{F}_t$ with $E\Phi_0^2 < \infty$ and Φ_0 is $\mathcal{F}_0$ measurable. Φ_t satisfies the stochastic differential equation

$$d\Phi_t = \alpha_t dt + dM_t \tag{2.1}$$

α_t is an $\mathcal{F}_t$ measurable process with $E\{\int_0^t \alpha_s^2 ds\} < \infty$ and M_t is an $\mathcal{F}_t$ martingale with $d\langle M, M\rangle_t = m_t dt$ where m_t is $\mathcal{F}_t$ measurable.

The observation process Y_t is a semimartingale with respect to $\mathcal{F}_t$ such that $EY_0^2 < \infty$ and Y_0 is $\mathcal{F}_0$ measurable. Y_t satisfies the stochastic differential equation

$$dY_t = \zeta_t dt + dN_t \tag{2.2}$$

where N_t is an $\mathcal{F}_t$ martingale with $d\langle N, N\rangle_t = n_t dt$ and n_t is an $\mathcal{F}_t$ measurable. ζ_t is an $\mathcal{F}_t$ measurable process with $E\{\int_0^t \zeta_s^2 ds\} < \infty$. Y_t generates the sigma algebra $\mathcal{F}_t^Y = \sigma(Y_s, s \leq t) \subset \mathcal{F}_t$.

The best estimate of Φ_t given the history of Y_t in $\mathcal{F}_t^Y$ is in the sense of minimizing the cost function $J(Z) = E\{(\Phi_t - Z)^2\}$ over all $Z \in \mathcal{H}_t^Y$:=$\{\mathcal{F}_t^Y$ measurable and $L^2(dP)$ square integrable functions$\}$. $J(Z)$ is a minimum if and only if $Z = E\{\Phi_t|\mathcal{F}_t^Y\}$.

Let $\widehat{\Phi}_t = E\{\Phi_t|\mathcal{F}_t^Y\}$ which is the projection of Φ_t on the Hilbert space $\mathcal{H}_t^Y$. The filtering problem is to derive the stochastic differential equation that characterize the evolution of $\widehat{\Phi}_t$. And this is how it is done:

First we take the conditional expectation of both sides of equation 2.1 given $\mathcal{F}_t^Y$;

$$\begin{aligned} E\{d\Phi_t|\mathcal{F}_t^Y\} &= E\{\alpha_t dt|\mathcal{F}_t^Y\} + E\{dM_t|\mathcal{F}_t^Y\}, \\ dE\{\Phi_t|\mathcal{F}_t^Y\} &= E\{\alpha_t|\mathcal{F}_t^Y\}dt + dE\{M_t|\mathcal{F}_t^Y\}, \\ d\widehat{\Phi}_t &= \widehat{\alpha}_t dt + d\widehat{M}_t. \end{aligned} \tag{2.3}$$

We do the same for the observation process Y_t;

$$\begin{aligned} E\{dY_t|\mathcal{F}_t^Y\} &= E\{\zeta_t dt|\mathcal{F}_t^Y\} + E\{dN_t|\mathcal{F}_t^Y\}, \\ dY_t &= \widehat{\zeta}_t dt + d\widehat{N}_t. \end{aligned} \tag{2.4}$$

The above argument is not rigorous. It can in fact be made rigorous if equations 2.1 and 2.2 are put in their integral form. But the results of computing conditional expectation are the same as equations 2.3 and 2.4.

Second, we write

$$d\nu_t = dY_t - \widehat{\zeta}_t dt \tag{2.5}$$

($i.e.$ $\nu_t \equiv \widehat{N}_t$) and call it the *innovation process.* Note that $d\left\langle \nu,\nu\right\rangle_t = d\left\langle Y,Y\right\rangle_t = d\left\langle N,N\right\rangle_t = n_t dt$. Now, we assume that Y_t is continuous. Then Allinger-Mitter [20] proved that the martingale representation property holds for all $\mathcal{F}_t^Y$ martingales: all $\mathcal{F}_t^Y$ martingales H_t have the form $dH_t = \eta_t d\nu_t$ for some process η_t adapted to $\mathcal{F}_t^Y$.

In the signal process equation 2.3 M_t is an $\mathcal{F}_t$ martingale then $\widehat{M_t} = E\{M_t|\mathcal{F}_t^Y\}$ is $\mathcal{F}_t^Y$ martingale. To simplify the estimate of the signal process equation 2.3 and relate it to Y_t then we want to write $\widehat{M_t}$ in terms of Y_t. But Y_t is a semimartingale with respect to $\mathcal{F}_t^Y$ and we can't use the martingale representation property to express $\widehat{M_t}$ in terms of Y_t. One way around this problem is to use the innovation process ν_t equation 2.5, which is a martingale with respect to $\mathcal{F}_t^Y$, to write an expression of $\widehat{M_t}$ in terms of Y_t. Therefore, by the martingale representation property there exits a process η_t adapted to $\mathcal{F}_t^Y$ such that

$$d\widehat{M_t} = \eta_t d\nu_t = \eta_t(dY_t - \widehat{\zeta}_t dt). \tag{2.6}$$

Hence equation 2.3 for $\widehat{\Phi}_t$ is reduced to

$$\begin{aligned} d\widehat{\Phi}_t &= \widehat{\alpha}_t dt + d\widehat{M_t} \\ &= \widehat{\alpha}_t dt + \eta_t d\nu_t. \end{aligned} \tag{2.7}$$

Our objective now is to find η_t.

We use the projection lemma, $tr(E\{E_t Z_s'\}) = 0\ \forall s \leq t\ Z_s \in \mathcal{F}_s^Y$ where $E_t = \Phi_t - \widehat{\Phi}_t$. So, take derivatives of $\Phi_t Y_t$ and $\widehat{\Phi}_t Y_t$ and compare their difference under the conditional expectation with respect to $\mathcal{F}_s^Y$;

$$\begin{aligned} d\left(\Phi_t Y_t\right) &= d\Phi_t Y_t + \Phi_t dY_t + d\left\langle \Phi, Y\right\rangle_t \\ &= \left(\alpha_t dt + dM_t\right) Y_t + \Phi_t(\zeta_t dt + dN_t) + \beta_t dt \\ &= \left(\alpha_t Y_t + \Phi_t \zeta_t + \beta_t\right) dt + \Phi_t dN_t + dM_t Y_t \end{aligned} \tag{2.8}$$

and

$$
\begin{aligned}
d(\widehat{\Phi}_t Y_t) &= d\widehat{\Phi}_t Y_t + \widehat{\Phi}_t dY_t + d\langle \widehat{\Phi}, Y\rangle_t \\
&= (\widehat{\alpha}_t dt + d\widehat{M}_t)Y_t + \widehat{\Phi}_t(\widehat{\zeta}_t dt + d\widehat{N}_t) + \eta_t n_t dt \\
&= (\widehat{\alpha}_t dt + \eta_t d\nu_t)Y_t + \widehat{\Phi}_t(\widehat{\zeta}_t dt + d\nu_t) + \eta_t n_t dt \\
&= (\widehat{\alpha}_t Y_t + \widehat{\Phi}_t \widehat{\zeta}_t + \eta_t n_t)dt + \widehat{\Phi}_t d\nu_t + \eta_t d\nu_t Y_t. \qquad (2.9)
\end{aligned}
$$

Integrate equations 2.8 and 2.9

$$
\Phi_t Y_t = \Phi_0 Y_0 + \int_0^t (\alpha_\tau Y_\tau + \Phi_\tau \zeta_\tau + \beta_\tau)\, d\tau + \int_0^t \Phi_\tau dN_\tau + \int_0^t dM_\tau Y_\tau, \qquad (2.10)
$$

$$
\widehat{\Phi}_t Y_t = \widehat{\Phi}_0 Y_0 + \int_0^t (\widehat{\alpha}_\tau Y_\tau + \widehat{\Phi}_\tau \widehat{\zeta}_\tau + \eta_\tau n_\tau) d\tau + \int_0^t \widehat{\Phi}_\tau d\nu_\tau + \int_0^t \eta_\tau d\nu_\tau Y_\tau. \qquad (2.11)
$$

Subtract equation 2.11 from equation 2.10 and take the conditional expectation with respect to $\mathcal{F}_s^Y$ for all $s \le t$, knowing that $E\{\Phi_t Y_t - \widehat{\Phi}_t Y_t | \mathcal{F}_s^Y\} = 0$ from the projection lemma. So,

$$
\begin{aligned}
&E\{\int_s^t (\widehat{\Phi_\tau \zeta_\tau} - \widehat{\Phi}_\tau \widehat{\zeta}_\tau + \widehat{\beta}_\tau - \eta_\tau n_\tau) d\tau | \mathcal{F}_s^Y\} \\
&= \int_{A\times(s,t]} (\widehat{\Phi_\tau \zeta_\tau} - \widehat{\Phi}_\tau \widehat{\zeta}_\tau + \widehat{\beta}_\tau - \eta_\tau n_\tau) d\tau dP = 0, \quad \forall s, t \ge s, \text{ and } A \in \mathcal{F}_s.
\end{aligned}
$$

Therefore, $\widehat{\Phi_\tau \zeta_\tau} - \widehat{\Phi}_\tau \widehat{\zeta}_\tau + \widehat{\beta}_\tau - \eta_\tau n_\tau = 0$ *a.e.* $d\tau dP$ or $\eta_t = (\widehat{\Phi_t \zeta_t} - \widehat{\Phi}_t \widehat{\zeta}_t + \widehat{\beta}_t) n_t^{-1}$ *a.e.* $dtdP$. Now, we write the SPDE for $\widehat{\Phi}_t$ as

$$
d\widehat{\Phi}_t = \widehat{\alpha}_t dt + (\widehat{\Phi_t \zeta_t} - \widehat{\Phi}_t \widehat{\zeta}_t + \widehat{\beta}_t) n_t^{-1} d\nu_t. \qquad (2.12)
$$

Formula 2.12 is not an explicit recursive equation in $\widehat{\Phi}_t$. To obtain an explicit recursive form we need to make more assumptions about the signal and observation processes. We will do this in section 2.2. But now I will introduce the reference measure approach.

Reference measure approach

In the *Reference probability* setup of the filtering problem we begin with *Girsanov Transformation* of the observation process, equation 2.2. We let P_o be a new probability measure on

$(\Omega, \mathcal{F})$ with $t \in [0, T]$. The Girsanov transform Λ_t given by

$$\begin{aligned}
\Lambda_t &= \exp(\int_0^t n_s^{-1}\zeta_s dN_s + \frac{1}{2}\int_0^t n_s^{-1}\zeta_s^2 ds) \\
&= \exp(\int_0^t n_s^{-1}\zeta_s(dY_s - \zeta_s ds) + \frac{1}{2}\int_0^t n_s^{-1}\zeta_s^2 ds) \\
&= \exp(\int_0^t n_s^{-1}\zeta_s dY_s - \frac{1}{2}\int_0^t n_s^{-1}\zeta_s^2 ds) \qquad (2.13)
\end{aligned}$$

transforms P to P_o in the following way $dPo = \Lambda_T^{-1}dP$. Since ζ_t is bounded for $t \in [0, T]$, P_o is a probability measure and Y_t is a P_o martingale. Let E_o be the expectation under P_o. Then we can write

$$\widehat{\Phi}_t = E\{\Phi_t|\mathcal{F}_t^Y\} = \frac{E_o\{\Phi_t\Lambda_t|\mathcal{F}_t^Y\}}{E_o\{\Lambda_t|\mathcal{F}_t^Y\}}. \qquad (2.14)$$

$E_o\{\Phi_t\Lambda_t|\mathcal{F}_t^Y\}$ is the unormalized conditional distribution of Φ_t given $\mathcal{F}_t^Y$ and $E_o\{\Lambda_t|\mathcal{F}_t^Y\}$ is the normalization factor.

To obtain a stochastic differential equation for $E_o\{\Phi_t\Lambda_t|\mathcal{F}_t^Y\}$ we need to know $E_o\{\Lambda_t|\mathcal{F}_t^Y\}$. Since $d\Lambda_t = \Lambda_t n_t^{-1}\zeta_t dY_t$ then

$$\begin{aligned}
\overline{\Lambda}_t = E_o\{\Lambda_t|\mathcal{F}_t^Y\} &= 1 + \int_0^t E_o\{\Lambda_s n_s^{-1}\zeta_s|\mathcal{F}_t^Y\}dY_s \\
&= 1 + \int_0^t \overline{\Lambda_s n_s^{-1}\zeta_s} dY_s \qquad (2.15)
\end{aligned}$$

where $\overline{(\ )}$ means to take the conditional expectation under P_o. Since $\overline{\Lambda_s n_s^{-1}\zeta_s} = E\{n_s^{-1}\zeta_s|\mathcal{F}_t^Y\}E_o\{\Lambda_t|\mathcal{F}_t^Y\}$ we are able to write equation 2.15 as

$$\overline{\Lambda}_t = 1 + \int_0^t \overline{\Lambda}_s\widehat{n_s^{-1}\zeta_s} dY_s. \qquad (2.16)$$

$\widehat{(\ )}$ is the expectation under P. Equation 2.16 has the unique solution

$$\overline{\Lambda}_t = \exp(\int_0^t \widehat{n_s^{-1}\zeta_s} dY_s - \frac{1}{2}\int_0^t \widehat{n_s^{-1}\zeta_s}^2 ds). \qquad (2.17)$$

Now, we are able to find the stochastic differential equation for $E_o\{\Phi_t\Lambda_t|\mathcal{F}_t^Y\}$. We rewrite equation 2.14 in the following way

$$\overline{\Phi_t\Lambda_t} = E_o\{\Phi_t\Lambda_t|\mathcal{F}_t^Y\} = E\{\Phi_t|\mathcal{F}_t^Y\}E_o\{\Lambda_t|\mathcal{F}_t^Y\} = \widehat{\Phi}_t\overline{\Lambda}_t. \tag{2.18}$$

By applying Ito's rule on $\widehat{\Phi}_t\overline{\Lambda}_t$, and recalling that n_t is $\mathcal{F}_t^Y$ measurable we get

$$\begin{aligned}
d(\overline{\Phi_t\Lambda_t}) &= d(\widehat{\Phi}_t\overline{\Lambda}_t) = \widehat{\Phi}_t d\overline{\Lambda}_t + \overline{\Lambda}_t d\widehat{\Phi}_t + d\langle\overline{\Lambda},\widehat{\Phi}\rangle_t \\
&= \widehat{\Phi}_t\overline{\Lambda}_t n_t^{-1}\widehat{\zeta}_t dY_t + \overline{\Lambda}_t(\widehat{\alpha}_t dt + (\widehat{\Phi_t\zeta_t} - \widehat{\Phi}_t\widehat{\zeta}_t + \widehat{\beta}_t)n_t^{-1}(dY_t - \widehat{\zeta}_t dt)) \\
&+ \overline{\Lambda}_t n_t^{-1}\widehat{\zeta}_t(\widehat{\Phi_t\zeta_t} - \widehat{\Phi}_t\widehat{\zeta}_t + \widehat{\beta}_t)dt \\
&= \{\overline{\Lambda}_t\widehat{\alpha}_t + \overline{\Lambda}_t n_t^{-1}\widehat{\zeta}_t(\widehat{\Phi_t\zeta_t} - \widehat{\Phi}_t\widehat{\zeta}_t + \widehat{\beta}_t) - \overline{\Lambda}_t n_t^{-1}\widehat{\zeta}_t(\widehat{\Phi_t\zeta_t} - \widehat{\Phi}_t\widehat{\zeta}_t + \widehat{\beta}_t)\}dt \\
&+ \{\overline{\Lambda}_t(\widehat{\Phi_t\zeta_t} - \widehat{\Phi}_t\widehat{\zeta}_t + \widehat{\beta}_t)n_t^{-1} + \overline{\Lambda}_t\widehat{\Phi}_t n_t^{-1}\widehat{\zeta}_t\}dY_t \\
&= \overline{\Lambda}_t\widehat{\alpha}_t dt + \overline{\Lambda}_t(\widehat{\Phi_t\zeta_t} + \widehat{\beta}_t)n_t^{-1}dY_t \\
&= \overline{\Lambda_t\alpha_t}dt + (\overline{\Lambda_t\Phi_t\zeta_t} + \overline{\Lambda\beta}_t)n_t^{-1}dY_t.
\end{aligned} \tag{2.19}$$

2.2 Fujisaki-Kallianpur-Kunita Equation

Here we derive *Fujisaki-Kallianpur-Kunita* (FKK) equation, a stochastic partial differential equation for filtering of Markov process X_t with a generator A (see Example 1). Let $f \in \mathcal{D}(A)$ be any function in the domain of A and define

$$M_{[s,t]}^f = f(X_t) - f(X_s) - \int_s^t Af(X_\tau)d\tau. \tag{2.20}$$

f is sometimes called the test function. $M_{[s,t]}^f$ is a martingale with respect to the filtration $\mathcal{F}_t^X$.

Example 1 *Let X_t be the Ito diffusion*

$$dX_t = g(X_t)dt + \sigma(X_t)dB_t \tag{2.21}$$

then the generator of this process is

$$A = \textstyle\sum_i g_i(x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{ij}(\sigma\sigma')_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_i} \tag{2.22}$$

To derive the FKK equation, we set $\Phi_t = f(X_t)$, $M_t = M^f_{[0,t]}$ and $\alpha_t = Af(X_t)$ in the signal process equation 2.1. Then the FKK signal process is described by

$$df(X_t) = Af(X_t)dt + dM_t. \tag{2.23}$$

And for the observation process, equation 2.2, we let $\zeta_t = h(X_t)$ and $N_t = W_t$ a Brownian motion with a quadratic variation $d\langle W, W\rangle_t = dt$, that is $n_t = 1$. Then in the FKK setup the observation process is given by

$$dY_t = h(X_t)dt + dW_t. \tag{2.24}$$

Note that $d\langle M, W\rangle_t = \beta_t dt$. Now, we introduce the notation

$$\pi_t(f) = \widehat{f(X_t)} = E\{f(X_t)|\mathcal{F}^Y_s\} = \int f(x)P_Y(dx, t, \omega) \tag{2.25}$$

and substitute the relevant variables in equation 2.12 to get the FKK filtering equation

$$d\pi_t(f) = \pi_t(Af)dt + (\pi_t(fh) - \pi_t(f)\pi_t(h) + \pi_t(\beta))d\nu_t. \tag{2.26}$$

We can also write the FKK equation in terms of the observation process Y_t as

$$d\pi_t(f) = \pi_t(Af)dt + (\pi_t(fh) - \pi_t(f)\pi_t(h) + \pi_t(\beta))(dY_t - \pi_t(h)dt). \tag{2.27}$$

In the case where the signal is independent of W_t, $\langle M, W\rangle_t = 0$ equation 2.27 becomes

$$\begin{aligned} d\pi_t(f) &= \pi_t(Af)dt + (\pi_t(fh) - \pi_t(f)\pi_t(h))(dY_t - \pi_t(h)dt) \\ &= (\pi_t(Af) - \pi_t(fh)\pi_t(h) + \pi_t(f)\pi_t(h)^2)dt + (\pi_t(fh) - \pi_t(f)\pi_t(h))dY_t. \end{aligned} \tag{2.28}$$

Equation 2.28 is known as the Kushner-Stratonovich (KS) equation.

The test function f used to derive FKK and KS equations is useful for computing conditional density if it is smooth, moments, or characteristic function when the probability density does not exit.

Density result

When the density exits and is smooth we can compute the conditional density directly without having to use a test function f. First note that we are able to write the conditional density $P_Y(dx, t, \omega)$ in many ways; $P_Y(dx, t, \omega) = \pi_t(dx) = \pi_t(x)dx = d\pi_t(x)$. Then, we can write $\pi_t(f)$ as

$$\pi_t(f) = \langle f, \pi_t \rangle = \int f(x)\pi_t(x)dx = \int f(x)\pi_t(dx) = \int f(x)d\pi_t(x). \tag{2.29}$$

$\langle f, \pi_t \rangle$ is the inner product of the function f with π_t. The inner product form makes it easy to derive the SPDE for the conditional density π_t.

To obtain a stochastic partial differential equation for the evolution of the conditional density $\pi_t(x)$ substitute $\pi_t(f)$ by $\langle f, \pi_t \rangle$ in equation 2.27

$$d\langle f, \pi_t \rangle = \langle Af, \pi_t \rangle dt + (\langle fh, \pi_t \rangle - \langle f, \pi_t \rangle \langle h, \pi_t \rangle)(dY_t - \langle h, \pi_t \rangle dt). \tag{2.30}$$

Since f is constant with respect to the differential d then

$$\begin{aligned} d\langle f, \pi_t \rangle = \langle f, d\pi_t \rangle &= \left\langle f, A'\pi_t \right\rangle dt + (\left\langle f, h'\pi_t \right\rangle - \langle f, \pi_t \rangle \langle h, \pi_t \rangle)(dY_t - \langle h, \pi_t \rangle dt) \\ &= (\left\langle f, A'\pi_t \right\rangle - \left\langle f, h'\pi_t \right\rangle \langle h, \pi_t \rangle + \langle f, \pi_t \rangle \langle h, \pi_t \rangle^2)dt + (\left\langle f, h'\pi_t \right\rangle - \langle f, \pi_t \rangle \langle h, \pi_t \rangle)dY_t \\ &= (\left\langle f, A'\pi_t \right\rangle - \left\langle f, h'\pi_t \langle h, \pi_t \rangle \right\rangle + \left\langle f, \pi_t \langle h, \pi_t \rangle^2 \right\rangle)dt + (\left\langle f, h'\pi_t \right\rangle - \langle f, \pi_t \langle h, \pi_t \rangle \rangle)dY_t \end{aligned} \tag{2.31}$$

where A' is the adjoint of the operator A and h' is the transpose of h.

Example 2 *The adjoint of Ito diffusion operator in example 1 is*

$$A' = -\textstyle\sum_i \dfrac{\partial(g_i(x)\cdot)}{\partial x_i} + \dfrac{1}{2}\textstyle\sum_{ij} \dfrac{\partial^2((\sigma\sigma')_{ij}(x)\cdot)}{\partial x_i \partial x_i}. \tag{2.32}$$

After rearranging terms of equation 2.31 we get

$$\langle f, d\pi_t - \{(A'\pi_t - h'\pi_t \langle h, \pi_t \rangle + \pi_t \langle h, \pi_t \rangle^2)dt + (h'\pi_t - \pi_t \langle h, \pi_t \rangle)dY_t\} \rangle = 0 \tag{2.33}$$

which implies that

$$d\pi_t = (A'\pi_t - h'\pi_t \langle h, \pi_t \rangle + \pi_t \langle h, \pi_t \rangle^2)dt + (h'\pi_t - \pi_t \langle h, \pi_t \rangle)dY_t. \tag{2.34}$$

2.3 Zakai Equation

The $\langle h, \pi_t \rangle$ and $\langle h, \pi_t \rangle^2$ factors complicates the use of 2.34. I now describe an approach which avoid them. Using the reference measure approach I will derive a linear stochastic partial differential equation of *Zakai* which describe the evolution of the unormalized conditional distribution of the signal given the observation.

Suppose the signal and observation are given by equations 2.23 and 2.24 respectively and that $\beta_t = 0$. Let $\mu_t(f) = \overline{f(X_t)\Lambda_t} = E_o\{f(X_t)\Lambda_t|\mathcal{F}_t^Y\}$ and substitute $\Phi_t = f(X_t)$, $\alpha_t = Af(X_t)$, $\zeta_t = h(X_t)$, and $n_t = 1$ in equation 2.19 to get

$$d\mu_t(f) = \mu_t(Af)dt + \mu_t(fh)dY_t. \tag{2.35}$$

Density result

Following the same lines of reasoning for the derivation of the density of the FKK equation in section 2.2 we can write the unormalized unconditional density of the Zakai equation as

$$d\mu_t = A'\mu_t dt + h'\mu_t dY_t. \tag{2.36}$$

Robust form

If $\frac{\partial h}{\partial t}$, $\frac{\partial h}{\partial x}$ and $\frac{\partial^2 h}{\partial x^2}$ are bounded then we can apply the following transformation to equation 2.36

$$\rho_t(x) = \mu_t(x) \exp(-h'Y_t). \tag{2.37}$$

This change of μ_t to ρ_t is known as the robust transformation. It was first introduced by Doss and Sussmann in order to reduce Ito's equations to ordinary differential equations. It is used extensively in nonlinear filtering by Lipster, Shiryaev, Clark and Davis. It allows us to reduce the Zakai equation to an ordinary PDE with random coefficients depending on the path

$Y_t(\omega)$. The solutions are then defined pathwise for $\rho_t(x)$. So we are not able to find information about expected values (for further details see [20],[2]).

2.4 Finite Dimensional Filters

Because $\pi_t(f)$ and $\mu_t(f)$ are a result of measure valued stochastic process equations 2.26 and 2.35 which depends on $\mathcal{F}_t^Y$ they are regarded as *infinite dimensional* stochastic partial differential equations. It means that we need an infinite number of parameters to characterize $\pi_t(f)$ or $\mu_t(f)$.

In general, it is not possible to derive a finite dimensional recursive filter, for the conditional normalized or unormalized expectation. In some special cases a finite dimensional filter exist. Some examples of finite dimensional filters are finite state Markov process, Kalman-Bucy [39], Bene's [1], and Daum [19].

Now, I will show how to get the Kalman filter from the nonlinear filtering problem and the conditional density for a finite Markov process.

Kalman filter

The Kalman filter [37] was the first finite dimensional filter. We are only required to compute two variables; the conditional mean and covariance of the error to fully characterize the conditional density of the process given the observation. From the nonlinear formulation of the filtering problem I will derive the conditional mean and covariance and show it is similar to the original derivation equations 1.7 and 1.8. Suppose that the signal process and observation satisfy the equations 1.5 and 1.6. To obtain the conditional mean we substitute $\widehat{X}_t$ for $\widehat{\Phi}_t$, $\widehat{\alpha}_t = a(t)\widehat{X}_t$, $\widehat{\beta}_t = 0$, $\zeta_t = c(t)X_t$, $d\nu_t = dY_t - c(t)\widehat{X}_t dt$ and $n_t = 1$. Also we let $r(t) = E\{(X_t - \widehat{X}_t)(X_t - \widehat{X}_t)'|\mathcal{F}_t^Y\}$ and $k(t) = r(t)c'(t)$ as we have defined them in the introduction. We compute the estimate of the signal

$$
\begin{aligned}
d\widehat{X}_t &= a(t)\widehat{X}_t dt + (\widehat{X_t X_t'} - \widehat{X}_t \widehat{X}_t')c'(t)(dY_t - c(t)\widehat{X}_t dt) \\
&= a(t)\widehat{X}_t dt + E\{(X_t - \widehat{X}_t)(X_t - \widehat{X}_t)' | \mathcal{F}_t^Y\} c'(t)(dY_t - c(t)\widehat{X}_t dt) \\
&= a(t)\widehat{X}_t dt + r(t)c'(t)(dY_t - c(t)\widehat{X}_t dt) \\
&= a(t)\widehat{X}_t dt + k(t)(dY_t - c(t)\widehat{X}_t dt). \qquad (2.38)
\end{aligned}
$$

The equation above is precisely the Kalman equation for the estimate of the signal equation 1.7. But since the estimate of the signal depends on $r(t)$ our next step is to compute $r(t)$. To obtain the conditional variance of the error, we let $\Phi_t = (X_t - \widehat{X}_t)(X_t - \widehat{X}_t)'$ and write

$$
\begin{aligned}
d\Phi_t &= d(X_t - \widehat{X}_t)(X_t - \widehat{X}_t)' + (X_t - \widehat{X}_t)d(X_t - \widehat{X}_t)' + d\langle (X_t - \widehat{X}_t), (X_t - \widehat{X}_t)' \rangle \\
&= (a(t)(X_t - \widehat{X}_t)dt + b(t)(dB_t - d\widehat{B}_t))(X_t - \widehat{X}_t)' \\
&+ (X_t - \widehat{X}_t)(a(t)(X_t - \widehat{X}_t)dt + b(t)(dB_t - d\widehat{B}_t))' \\
&+ d\langle (X_t - \widehat{X}_t), X_t' \rangle - d\langle (X_t - \widehat{X}_t), \widehat{X}_t' \rangle.
\end{aligned}
$$

Since $X_t - \widehat{X}_t$ is independent of $\widehat{X}_t$ then $d\langle (X_t - \widehat{X}_t), \widehat{X}_t' \rangle = 0$. Furthermore, $d\langle X_t, X_t' \rangle = b(t)b(t)'dt$. Substituting these values and Φ_t for $(X_t - \widehat{X}_t)(X_t - \widehat{X}_t)'$ we get

$$
\begin{aligned}
d\Phi_t &= (a(t)\Phi_t + \Phi_t a'(t) + b(t)b(t)')dt - d\langle \widehat{X}_t, X_t' \rangle \\
&\quad + b(t)(dB_t - d\widehat{B}_t)(X_t - \widehat{X}_t)' + (X_t - \widehat{X}_t)(dB_t - d\widehat{B}_t)'b'(t).
\end{aligned}
$$

Now we take the conditional expectation with respect to $\mathcal{F}_t^Y$ to get

$$
d\widehat{\Phi}_t = (a(t)\widehat{\Phi}_t + \widehat{\Phi}_t a'(t) + b(t)b'(t) - k(t)k'(t))dt. \qquad (2.39)
$$

The $E\{b(t)d(B_t - \widehat{B}_t)(X_t - \widehat{X}_t)' + (X_t - \widehat{X}_t)d(B_t - \widehat{B}_t)'b'(t) | \mathcal{F}_t^Y\} = 0$ because $X_t - \widehat{X}_t$ and $B_t - \widehat{B}_t$ are independent of $\mathcal{F}_t^Y$. Also, $E\{d\langle \widehat{X}_t, X_t' \rangle | \mathcal{F}_t^Y\} = d\langle \widehat{X}_t, E\{X_t' | \mathcal{F}_t^Y\} \rangle = k(t)k'(t)dt$.

Writing equation 2.39 in terms of $r(t)$

$$\dot{r}(t) = a(t)r(t)' + r(t)a'(t) + b(t)b'(t) - k(t)k'(t). \tag{2.40}$$

Equations 2.38 and 2.40 are exactly those of the Kalman filter.

We have obtained the Kalman filter from the derivation procedure of nonlinear filtering. But why is the Kalman filter a finite dimensional filter? The reason is that the estimate of the error covariance $r(t)$ does not depend on higher order moments and is a deterministic process, independent of any underlying noise. Therefore, the Kalman filter is a finite dimensional filter.

It is interesting to write out the FKK equation for a simple Kalman filter where X_t is a scalar process and $a = b = c = 1$. The equation is

$$d\pi_t = (\frac{1}{2}\frac{\partial^2 \pi_t}{\partial x^2} - x\frac{\partial \pi_t}{\partial x} - \pi_t(1 + x\widehat{X}_t - \widehat{X}_t^2))dt + (x - \widehat{X}_t)\pi_t dY_t. \tag{2.41}$$

For this simple Kalman filter we know that $d\widehat{X}_t = \widehat{X}_t(1 - r(t))dt + r(t)dY_t$ and $\dot{r}(t) = 2r(t) + 1 - r(t)^2$ are the estimate mean and covariance of X_t given Y_t. Furthermore, we also know that the π_t is Gaussian with mean $\widehat{X}_t$ and variance $r(t)$. We use this fact to solve equation 2.41.

Let

$$\pi_t = \frac{1}{\sqrt{2\pi r}}\exp(-\frac{(x - \widehat{X}_t)^2}{2r}).$$

By Ito's lemma for a constant x (*i.e.* $dx = 0$) we get

$$\begin{aligned} d\pi_t &= (\frac{1}{2}\frac{\partial^2 \pi_t}{\partial \widehat{x}^2}r^2 + \frac{\partial \pi_t}{\partial r}\dot{r})dt + \frac{\partial \pi_t}{\partial \widehat{x}}d\widehat{X}_t \\ &= (\frac{1}{2}\frac{\partial^2 \pi_t}{\partial \widehat{x}^2}r^2 + \widehat{X}_t(1 - r)\frac{\partial \pi_t}{\partial \widehat{x}} + \frac{\partial \pi_t}{\partial r}\dot{r})dt + r\frac{\partial \pi_t}{\partial \widehat{x}}dY_t. \end{aligned} \tag{2.42}$$

Since

$$\begin{aligned} \frac{\partial \pi_t}{\partial x} &= \frac{-(x - \widehat{X})}{r}\pi_t, \\ \frac{\partial^2 \pi_t}{\partial x^2} &= \left(\frac{(x - \widehat{X})^2}{r^2} - \frac{1}{r}\right)\pi_t \end{aligned}$$

and

$$\frac{\partial \pi_t}{\partial \widehat{x}} = \frac{(x-\widehat{X})}{r}\pi_t = -\frac{\partial \pi_t}{\partial x},$$
$$\frac{\partial^2 \pi_t}{\partial \widehat{x}^2} = \left(\frac{(x-\widehat{X})^2}{r^2} - \frac{1}{r}\right)\pi_t = \frac{\partial^2 \pi_t}{\partial x^2},$$
$$\frac{\partial \pi_t}{\partial r} = \left(\frac{1}{2}\frac{(x-\widehat{X})^2}{r^2} - \frac{1}{2r}\right)\pi_t = \frac{1}{2}\frac{\partial^2 \mu_t}{\partial x^2}.$$

By substituting these partial derivative in equation 2.42 and after a series of algebraic transformations

$$\begin{aligned}
d\pi_t &= (\frac{1}{2}\frac{\partial^2 \pi_t}{\partial x^2}r^2 + \widehat{X}_t(r-1)\frac{\partial \pi_t}{\partial x} + \frac{1}{2}\frac{\partial^2 \mu_t}{\partial x^2}(2r+1-r^2))dt + (x-\widehat{X})\pi_t dY_t \\
&= (\frac{1}{2}\frac{\partial^2 \pi_t}{\partial x^2} + \widehat{X}_t(r-1)\frac{\partial \pi_t}{\partial x} + \frac{\partial^2 \pi_t}{\partial x^2}r)dt + (x-\widehat{X})\pi_t dY_t \\
&= (\frac{1}{2}\frac{\partial^2 \pi_t}{\partial x^2} + \widehat{X}_t(r-1)\frac{\partial \pi_t}{\partial x} - (x-\widehat{X}_t)\frac{\partial \pi_t}{\partial x} - \pi_t)dt + (x-\widehat{X})\pi_t dY_t \\
&= (\frac{1}{2}\frac{\partial^2 \pi_t}{\partial x^2} - x\frac{\partial \pi_t}{\partial x} + \widehat{X}_t r\frac{\partial \pi_t}{\partial x} - \pi_t)dt + (x-\widehat{X})\pi_t dY_t \\
&= (\frac{1}{2}\frac{\partial^2 \pi_t}{\partial x^2} - x\frac{\partial \pi_t}{\partial x} - \widehat{X}_t(x-\widehat{X})\pi_t - \pi_t)dt + (x-\widehat{X})\pi_t dY_t \\
&= (\frac{1}{2}\frac{\partial^2 \pi_t}{\partial x^2} - x\frac{\partial \pi_t}{\partial x} - \pi_t(1 + x\widehat{X}_t - \widehat{X}_t^2))dt + (x-\widehat{X})\pi_t dY_t
\end{aligned}$$

we get the FKK equation 2.41 for the simple Kalman filter. This means that $\pi_t = \frac{1}{\sqrt{2\pi r}}\exp(-\frac{(x-\widehat{X}_t)^2}{2r})$ solves equation 2.41. And it is in fact the conditional density of X_t given the observations Y_t.

Finite Markov process

An interesting finite filtering problem is that in which a scalar valued X_t is a finite state Markov process [46] taking values in the set $\{x_i\}_{i=1}^n$. Let $p_t^i = P(X_t = x_i)$, and $p_t = (p_t^1, p_t^2, ..., p_t^n)'$ which evolves according to

$$\frac{dp_t}{dt} = A'p_t. \tag{2.43}$$

A is the generator of X_t. Let the test function $f^i(X_t) = \mathbf{1}_{\{X_t=x_i\}}$ and use equation 2.35 to get

$$\begin{aligned} d\mu_t^i &= \langle f^i, d\mu_t \rangle = \langle \mathbf{1}_{\{X_t=x_i\}}, A'\mu_t \rangle \, dt + \langle \mathbf{1}_{\{X_t=x_i\}} h, \mu_t \rangle \, dY_t \\ &= (A'\mu_t)_i dt + h(x_i)\mu_t^i dY_t \\ &= (\textstyle\sum_{j=1}^n A'_{ij}\mu_t^j) dt + h(x_i)\mu_t^i dY_t. \end{aligned} \tag{2.44}$$

Let the function $f = (f^1, f^2, ..., f^n)'$, $\mu_t = (\mu_t^1, \mu_t^2, ..., \mu_t^n)'$ and $H = diag(h(x_1), h(x_2), ..., h(x_n))$ then we can write the above set of equations 2.44 in a vector form as follows

$$\begin{aligned} d\mu_t &= A'\mu_t dt + H\mu_t dY_t \\ &= (A'dt + HdY_t)\mu_t. \end{aligned} \tag{2.45}$$

This has the solution

$$\mu_t = \exp((A' - \frac{1}{2}H^2)t + HY_t)\mu_0. \tag{2.46}$$

Equation 2.45 is a finite n-dimensions bilinear stochastic differential equations for the unormalized conditional probabilities μ_t^i and equation 2.46 is its solution. The conditional probability density is given by $\pi_t = \frac{\mu_t}{\mu_t'\mu_t}$.

Other finite dimensional filters are those developed by Bene's [1], Duam [19] and Yau [59]. All finite dimensional filters involve some restrictions imposed on the dynamical equations governing the evolution of the signal and observation processes. So, does there exit a general criteria by which we can find finite dimensional filters? Algebraic methods, the topic of the next section, addresses this question.

2.5 Algebraic Methods

The purpose of this section is to shed some light on the algebraic structure of the filtering problem. Brockett, Clark and Mitter were the first to apply methods from Lie algebra, group representations, and functional integration to the filtering problem. This led to criteria to find the existence of finite dimensional filters, to efficient methods for computing these filters and showed that some filtering problems are inherently infinite dimensional.

The object of study in these approaches is the unormalized density equation in the Fisk-Stratonovich representation

$$d\mu_t(x) = (A' - \frac{1}{2}h^2(x))\mu_t(x)dt + h'(x)\mu_t(x) \circ dY_t. \tag{2.47}$$

This is done for two reasons: First, the unormalized density equation is simpler and is in the bilinear form where Y_t is its input and $\mu_t(f)$ is the output; Second stochastic differentials are treated as ordinary differentials which makes it easier to use Lie Algebraic methods.

Let $\mathcal{L} = \{A' - \frac{1}{2}h^2, h'\}_{LA}$ the Lie Algebra generated by $A' - \frac{1}{2}h^2$ and h' on smooth functions C^∞. $\mathcal{L}$ is known as the *estimation algebra*. Using the estimation algebra the following questions were addressed: Given a filtering problem, do there exist a low finite dimensional filter that computes the conditional distribution? Does there exist a classification of filtering problems based on an order of complexity in some sense? Are there "equivalent" filtering problems and are there useful invariants associated with each equivalence class?

Finite estimation algebras and finite filters

Definition 3 *An estimation algebra $\mathcal{L}$ is finite if and only if* $\dim \mathcal{L} < \infty$ *where* $\dim \mathcal{L} \geq \dim \mathcal{L}f$ *for all $f \in C^\infty$.*

Definition 4 *A Lie algebra $\mathcal{L}$ (e.g. estimation algebra) is solvable if $\mathcal{L}^{(n)} = 0$ for some $n \in \mathbb{N}$. $\mathcal{L}^{(n)} = \{[u, v] : u, v \in \mathcal{L}^{(n-1)}\}$ and $\mathcal{L}^{(0)} = \mathcal{L}$.*

If $\mathcal{L}$is finite dimensional and solvable then a finite dimensional filter exits.

The filter can usually be computed by the Wei-Norman method [56]. If $\{\Lambda_i\}_{i=1}^n$ are the basis of the estimation algebra $\mathcal{L}$ then Wei-Norman solution of the finite dimensional Zakai equation is

$$\begin{aligned} \mu_t &= \left(\exp \lambda_t^1 \Lambda_1\right) \left(\exp \lambda_t^2 \Lambda_2\right) ... \left(\exp \lambda_t^n \Lambda_n\right) \mu_0 \\ \lambda_t^i &= \textstyle\int_0^t \alpha^i(\lambda_s, Y_s(\omega))ds \qquad\qquad i = 1, 2, ..., n \end{aligned} \tag{2.48}$$

The fact that a finite dimensional filter exits when the estimation algebra is finite and solvable is an important result. The second question is; does there exits a filter with lower than n dimensions for an n dimensional filtering problem? The answer to this question is yes. In fact if $\dim \mathcal{L} = n \geq 2$ then there is a smallest filter of dimension 2 and a transformation ψ from the 2 dimensional filter to the n dimensional unormalized density. In other words, the estimation algebra $\mathcal{L}$ is isomorphic to $gl(2)$. (For a proof see [46])

Example 5 *Estimation algebra for finite Markov process*

$$d\mu_t = (A' - \frac{1}{2}H^2)\mu_t dt + H\mu_t \circ dY_t. \tag{2.49}$$

The estimation algebra $\mathcal{L} = \{A' - \frac{1}{2}H^2, H\}$. *The* $\dim \mathcal{L} \leq n$.

Example 6 *Estimation algebra for scalar linear Kalman problem. Suppose* $a = b = c = 1$ *and* X_t *is scalar valued then the generator for the linear problem is*

$$A' = \frac{1}{2}\frac{\partial^2}{\partial x^2} - x\frac{\partial}{\partial x} - 1. \tag{2.50}$$

$\mathcal{L} = \{\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}x^2 - x\frac{\partial}{\partial x} - 1, x\}_{LA}$ *and* $\dim \mathcal{L} = 4$ *spanned by* $\{\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}x^2, \frac{\partial}{\partial x}, x, 1\}$.

Isomorphic estimation algebras

The estimation algebra can be useful in recognizing "equivalent" filtering problems; that is, when the filtering problem is invariant under certain transformations. In [46], it is shown that if $\psi : \mathbb{R} \to \mathbb{R}$ is smooth and positive then the Zakai equation and its associated Lie Algebra $\mathcal{L}$ is isomorphic to $\tilde{\mathcal{L}} = \{\psi A' \psi^{-1} - \frac{1}{2}h^2, h'\}_{LA}$. The implication of this is that a solution of one filtering problem can be used to solve another filtering problems just by a simple transformation of the operators $A' - \frac{1}{2}h^2$ and h'. The transformation ψ is called the gauge transformation. Another related method is when one performs a smooth nonsingular change of variables $z = \alpha(x)$ with inverse $x = \beta(z)$. Again $\mathcal{L}$ is isomorphic to $\{A' \circ \beta - \frac{1}{2}(h \circ \beta)^2, (h \circ \beta)'\}_{LA}$. An example of a nonlinear filtering problem isomorphic to the Kalman problem and that can be solved with the Kalman filter with a simple change of coordinates between x and z is

$$dZ_t = \frac{1}{2}(\alpha'' \circ \beta)(Z_t)dt + (\alpha' \circ \beta)(Z_t)dB_t \tag{2.51}$$

$$dY_t = (h \circ \beta)(Z_t)dt + dW_t. \tag{2.52}$$

The set of all transformations consisting of the gauge transform and nonsingular change of variables described above forms a group under composition called the estimation equivalence group. Two estimation problems are equivalent if their estimation algebras can be transformed into one another by elements of this group [8].

Homomorphism principle

A finite dimensional recursive filter that computes $\mu_t(f)$ exits if it has the form

$$\xi_t = \xi_0 + \int_0^t a(\xi_s)ds + \int_0^t b(\xi_s) \circ dY_s \tag{2.53}$$

$$\mu_t(f) = \Psi(\xi_t) \tag{2.54}$$

where ξ_t evolves on a finite dimensional Euclidean space or on an analytic manifold and a, b and Ψ are analytic [46]. It is important to note that the filter algebra $\{a, b\}_{LA}$ of the finite dimensional recursive filter 2.53 and 2.54 may be infinite dimensional. Consider this example,

$$d\xi_t = (\xi_t)^2 \circ dY_t^1 + (\xi_t)^3 \circ dY_t^2$$
$$\mu_t = \xi_t$$

$a(\xi_t) = 0$, $b^1(\xi_s) = (\xi_t)^2$, $b^2(\xi_s) = (\xi_t)^3$ which generates an infinite estimation Lie algebra of polynomials [46]. So, it is possible for estimation algebra to be infinite dimensional even on a one-dimensional manifold. Therefore, an infinite dimensional estimation algebra does not mean that there is not a finite recursive filter to be found. Brockett [6], [7], [8] realized that if a finite dimensional recursive filter 2.53 and 2.54 were to solve 2.47, then there must be a homomorphism from the estimation algebra of 2.47 into the filter algebra $\{a, b\}_{LA}$ even when both the estimation and filter algebras may be infinite dimensional.

This view has lead to the *homomorphism principle*; If there is a homomorphism ϕ from the estimation algebra of 2.47 into a Lie algebra generated by two complete analytic vector fields on a finite dimensional manifold $\mathcal{M}$, and if ϕ maps the isotropy subalgebra at the initial density into the isotropy subalgebra at a point of $\mathcal{M}$, then this is an indication that some conditional statistic may be computable by a finite recursive filter of the form 2.53 and 2.54. An important corollary of homomorphism principle is that for a class of infinite dimensional Lie algebras, the *Weyl algebras*, for which there are no nonconstant homomorphisms into Lie algebras of analytic vector fields on C^∞ [6]. Therefore, if the estimation algebra of a filtering problem is a Weyl algebra then there is no nonconstant statistic of the conditional density can be computed with a recursive finite dimensional filter.

The homomorphism principle has made it possible to prove that some filtering problems are truly infinite dimensional. For example, the cubic sensor problem [33]

$$dX_t = dB_t \tag{2.55}$$

$$dY_t = (X_t)^3 dt + dW_t. \tag{2.56}$$

Its estimation algebra is the Weyl algebra W_n (*i.e.* the algebra of all polynomial differential operator on $\mathbb{R}^n$).

Final comment

How can the algebraic methods be of use for a particular nonlinear filtering problem? First an attempt should be made to compute the estimation algebra; if it is finite dimensional and solvable, then a finite dimensional filter can usually be computed by the Wei-Norman method; If it is Weyl algebra then there are no exact finite dimensional recursive filters for conditional statistics and approximation methods should be used; If it is infinite dimensional but not a Weyl algebra then the estimation algebra is not of use, because even if there is homomorphism from estimation algebra to Lie algebras of vector fields on finite dimensional manifolds, it is difficult to determine precisely which conditional statistic the corresponding filter computes [46].

Chapter 3

Computational Approximations

Computational approximations are used when the nonlinear filtering problem is infinite dimensional and a finite dimensional recursive filter doesn't exit. The goal from the approximation is to get a simple, fast, nearly efficient and nearly optimal finite dimensional filter. There are two ways the approximation can be done. The first approach is to approximate the equations of the state and observation processes and then compute the estimate of the signal given the observation; The extended Kalman filter is of this type. I will refer to these methods as the *direct approximation* methods. The other approach is to approximate the conditional normalized or unormalized expectation of the signal given the observation; Particle filters are in this category. I will refer to these as *approximation of random measures*.

3.1 Direct approximation

In these methods the equations of the state process and observation process are approximated and then estimates are computed. There are many filters in this category: the extended Kalman filter [34]; second order method [55]; geometrically intrinsic filter [16]; Monte-Carlo simulation filter [47]; unscented filter [52]; Gauss–Hermite quadrature filter [41]; and probably many more. These methods are favoured by engineers designing real-time applications over conditional density approximations. Because, it is computationally expensive to store and update conditional densities when computational resources are limited and fast estimation is required. However, It is known that applying approximation procedures to signal and observation equations leads to biased filtering estimates [54].

Here I will describe briefly a version of the extended Kalman filter and the Monte-Carlo simulation filter. The other filters described in literature are enhancement and variants of the above two.

3.1.1 Extended Kalman filter

The extended Kalman filter also known as first order method works by linearizing the state process and observation function around a current guessed state. Then apply the Kalman filter algorithm to obtain the estimate of the signal [34]. This method is used widely by engineers because it involves only update of conditional mean and covariance, and can be easily programmed. Picard has given a rigorous analysis of its efficiency [51]. Unfortunately, the filtered estimates depends on the coordinate systems used for state and observation.

Suppose the signal X_t and Y_t are real valued processes satisfying the nonlinear stochastic differential equations

$$
\begin{aligned}
dX_t &= A(X_t)dt + b(X_t)dB_t, \\
dY_t &= C(X_t)dt + dW_t.
\end{aligned}
$$

Introduce $\bar{X}_t$ to be our guess of the state at time t. We prefer a deterministic guess; so we choose $\bar{X}_t$ to be the solution of the deterministic differential equation [34]

$$
\frac{d\bar{X}_t}{dt} = A(\bar{X}_t).
$$

But this is not the only possible choice. In the discrete version of the extended Kalman filters $\bar{X}_t$ is chosen to be the predicted state given the observation $\bar{X}_t = E\{X_t|\mathcal{F}_{t-1}^Y\}$ [54].

Now that we know what $\bar{X}_t$ is, we approximate X_t and Y_t as a perturbation from $\bar{X}_t$ within a small window of time $(0, \delta t)$

$$
\begin{aligned}
dX_t &\simeq (A(\bar{X}_t) + \frac{dA}{dx}(\bar{X}_t)(X_t - \bar{X}_t))dt + b(\bar{X}_t)dB_t \qquad (3.1) \\
&\simeq (a(t)X_t + d(t))dt + b(t)dB_t,
\end{aligned}
$$

where $a(t) = \frac{dA}{dx}(\bar{X}_t)$, $d(t) = A(\bar{X}_t) - \frac{dA}{dx}(\bar{X}_t)\bar{X}_t$ and $b(t) = b(\bar{X}_t)$; the observation is

$$\begin{aligned} dY_t &\simeq (C(\bar{X}_t) + \frac{dC}{dx}(\bar{X}_t)(X_t - \bar{X}_t))dt + dW_t \\ &\simeq (c(t)X_t + \kappa(t))dt + dW_t, \end{aligned} \tag{3.2}$$

where $c(t) = \frac{dC}{dx}(\bar{X}_t)$, $\kappa(t) = C(\bar{X}_t) - \frac{dC}{dx}(\bar{X}_t)\bar{X}_t$.

Using the Kalman filter we get the approximate state estimate as

$$\begin{aligned} d\widehat{X}_t &= (a(t)\widehat{X}_t + d(t))dt + r(t)c'(t)(dY_t - (c(t)\widehat{X}_t + \kappa(t))dt) \\ &= ((a(t) - r(t)c'(t)c(t))\widehat{X}_t + d(t) - r(t)c'(t)\kappa(t))dt + r(t)c'(t)dY_t, \end{aligned}$$

where $r(t)$ is given by equation 2.40.

If the initial position of the signal is well approximated, the coefficient $A(X_t)$, $C(X_t)$ and $b(X_t)$ are only slightly nonlinear and $C(X_t)$ is one to one and the system is stable then the approximation will be good.

In the extended Kalman filter the state and observation processes are linearized up to the first term of the Taylor series. An extension of this first order approximation is the second order filter [55] where the first and second terms of the Taylor series are used. The problem with these two filters is that the linearization depends on the coordinate system used which changes the accuracy and bias of the estimates. The geometric intrinsic filter is another linearization method equivalent to the second order filter. But unlike the extended Kalman filter and its second order extension it is coordinate invariant. It applies to the estimation of a Markov diffusion process subject to discrete time observations [16].

The geometric intrinsic filter is based on the fact that the signal and observation covariances induce geometries on state space and observation space, respectively. And, they give rise to geometric objects known as intrinsic location parameters from which a quadratic coordinate free filter update formula is derived [18]. The interested reader can find details of the derivation and examples in the following list of papers by Darling [16], [17] and [18].

3.1.2 Monte-Carlo simulation filter

Tanizaki and Mariano [54], [47] developed the Monte-Carlo simulation filter. They used Monte-Carlo stochastic simulation to compute the mean and variance of the state given the observation for the approximated signal and observation processes, equations 3.1 and 3.2. The samples for the simulation are drawn from a normal distribution whose mean and variance are adapted based on previous state estimates and on new observations.

Tanizaki and Mariano's derivation of the filter is in discrete time. Here, I will give the derivation in continuous time with samples drawn from a sum of Gaussians whose parameters are updated based on the arrival of new observations. Suppose the observation is given by $Y_t = h(X_t, W_t)$ where X_t, is the signal and W_t is Brownian motion independent of X_t. Let $\{X_{i,t}\}_{i=1}^n$ be samples from our current *guess* of what the distribution of X_t is. The samples are drawn from a sum of Gaussian distributions of different means and variances

$$\{X_{i,t}\}_{i=1}^n \sim p(x; K, w, m, v) = \textstyle\sum_{k=1}^K w_k N(m_k, v_k)$$

where $N(m_k, v_k)$ is normal and w_k, m_k, and v_k are updated according to Y_t.

The goal is to compute $E\{f(X_t)|\mathcal{F}_t^Y\}$ for a test function f. We can approximate $E\{f(X_t)|\mathcal{F}_t^Y\}$ by its empirical estimate

$$E\{f(X_t)|\mathcal{F}_t^Y\} \approx \frac{1}{n}\textstyle\sum_{i=1}^n f(X_{i,t})$$

if our guess distribution of X_t given $\mathcal{F}_t^Y$ is good.

To check if our guess distribution of X_t is correct we will have to compute the resulting values of Y_t's for every sample $X_{i,t}$, and their empirical mean

$$\begin{aligned} \tilde{Y}_{i,t} &= h(X_{i,t}, W_t), \\ \widetilde{Y}_t &= \frac{1}{n}\textstyle\sum_{i=1}^n \tilde{Y}_{i,t}. \end{aligned}$$

where $\widetilde{Y}_t$ is the observation arising from our guess of X_t given $\mathcal{F}_t^Y$. If $\widetilde{Y}_t$ is close to the actual observed measurements Y_t for all t then we are done. But what if $\widetilde{Y}_t$ is different from Y_t then we set $e_t = \widetilde{Y}_t - Y_t$ and adjust the parameters m, w,and v of the distribution $p(x; K, w, m, v)$ as to minimize the error e_t. Gradient descent is a possible method to minimize e_t. Simply we write

$$\begin{aligned}\dot{w}_k &= -\eta\frac{\partial E\{e_t e_t'\}}{\partial w_k} \\ &= -\eta E\{e_t\frac{\partial h'}{\partial x}\frac{\partial p}{\partial w_k}\},\end{aligned}$$

where $E\{\}$ is with respect to the distribution of W_t. And for the other parameters we get

$$\begin{aligned}\dot{m}_k &= -\eta E\{e_t\frac{\partial h'}{\partial x}\frac{\partial p}{\partial m_k}\}, \\ \dot{v}_k &= -\eta E\{e_t\frac{\partial h'}{\partial x}\frac{\partial p}{\partial v_k}\}.\end{aligned}$$

Deterministic variants of Monte-Carlo simulation filters are the unscented and Gauss-Hermite quadrature filters. Instead of randomly sampling from a distribution, the basic idea of the unscented filter is to choose a deterministic sample of points known as sigma points with associated fixed weights to capture the mean and covariance of a density. And when the sample points are propagated through a nonlinear observation function, they will still capture the mean and covariance of that function up to second order term. The Gauss-Hermite quadrature filter also use a deterministic sample of points known as quadrature points to approximate expectations. For example, if $X_{i,t}$ were quadrature points and α_i are their associated weights then

$$E\{f(X_t)|\mathcal{F}_t^Y\} \approx \textstyle\sum_{i=1}^{n}\alpha_i f(X_{i,t}).$$

3.2 Approximation of random measures

In general computing π_t or μ_t explicitly in closed form as a function of the observation process is rare. Often, π_t and μ_t are infinite dimensional and not computable. In practice, π_t or μ_t would have to be approximated by a suitable finite description to run on a computer. A class of known measures $\{\xi_t^i\}$ on $\mathcal{F}_t^Y$ are used to approximate π_t or μ_t as follows

$$\mu_t(f) \approx \mu_t^n(f) = \textstyle\sum_{i=1}^{n(t)} w_i \xi_t^i(f)$$

Examples of sets of possible approximation measures is the exponential family for the projection filter and Gaussian sum filter, the set discrete measures for particle filters, the set of densities with fixed length for spectral decompositions or densities with finite spatial and temporal extent for grid approximation methods such Markov chains and numerical PDE methods.

In addition to measure approximation we need an efficient algorithm to update the finite dimensional representation as observations accumulates and in the limit the algorithm must converge to the infinite dimensional SPDE.

Here I will describe briefly a version of particle filters to solve the Zakai equation, the spectral decomposition method and neural network filters.

3.2.1 Particle methods

Particle filters approximate π_t or μ_t by discrete random measures. They are among the most successful and versatile methods for solving nonlinear filtering problems. For an in depth review of particle filters see [10], [54], and [13].

Here I will summarize results for the particle approximation of the Zakai equation [14]. Let μ_t be the unormalized density whose evolution is give by the Zakai equation

$$d\mu_t(f) = \mu_t(Af)dt + \mu_t(fh)dY_t,$$

and $\pi_t = \frac{\mu_t}{\mu_t(1)}$ the corresponding density. Particle filters approximate $\mu_t(f)$ by the empirical measure

$$\mu_t^n(f(x)) = \frac{1}{n}\sum_{k=1}^{m_n(t)} f(x)\delta_{Z_t^k}(x)$$

where $\{Z_t^k\}_{k=1}^{m_n(t)}$ is a sequence of identically distributed random variables with distribution π_t, $m_n(t) \in \mathbb{N}$ is the number of particles and $\frac{1}{n}$ is the mass of each particle. The goal of any particle filter is

$$\lim_{n\to\infty} \mu_t^n(f) = \mu_t(f).$$

But π_t is unknown, so how is it possible to construct a sequence of random variables Z_t^k having π_t as their distribution? Here is how it is done.

The particle filter is begun by taking $Z_0^1, Z_0^2, ..., Z_0^{m_0}$ samples from π_0, which is known.

Compute the empirical estimate of $f(X_0)$ as

$$\mu_0^n(f(x)) = \frac{1}{n}\textstyle\sum_{k=1}^n f(x)\delta_{Z_0^k}(x),$$

which is our initial empirical estimate. Now we discretize the process μ_t^n in time and consider its evolution on the intervals $[\frac{i}{n}, \frac{i+1}{n}]$, $i = 0, 1, 2, ..., n-1$. Let $\mu_i^n = \mu_t^n|_{t=\frac{i}{n}}$ and $m_i = m_n(t)|_{t=\frac{i}{n}}$. At $i = 0$, $\mu_i^n = \mu_0^n$ and $m_0 = n$. During the interval $t \in [\frac{i}{n}, \frac{i+1}{n})$ the particles Z_t^k move independently with the same law as that of the process X_t (*i.e.* the dynamical equations or the transition probability density of X_t). At the end of the interval $\frac{i+1}{n}$, a branching process commences and m_i is updated to m_{i+1}. Each particle Z_t^k is branched into a random number of new particles such that the second moment of the branching process is finite. The initial state of the new particles at the beginning of the next interval $[\frac{i+1}{n}, \frac{i+2}{n})$ is set to the final state of their parent particle $Z_t^k|_{t=\frac{i+1}{n}}$. Basically, the branching process make q_{i+1}^k copies of the particle Z_t^k at the end of the interval $[\frac{i}{n}, \frac{i+1}{n})$. Thus, the total number of particles at the beginning of each interval is $m_{i+1} = \sum_{k=1}^{m_i} q_{i+1}^k$. After the particles have evolved and branched μ_i^n is updated to μ_{i+1}^n,

$$\mu_i^n = \frac{1}{n}\textstyle\sum_{k=1}^{m_i} f(x)\delta_{Z_{i/n}^k}(x) \longrightarrow \mu_{i+1}^n = \frac{1}{n}\textstyle\sum_{k=1}^{m_{i+1}} f(x)\delta_{Z_{i+1/n}^k}(x).$$

During the interval $t \in [\frac{i}{n}, \frac{i+1}{n}]$ the real unormalized density would have evolved in the following way

$$\mu_t(f)|_{t=\frac{i+1}{n}} = \mu_t(f)|_{t=\frac{i}{n}} + \textstyle\int_{i/n}^{i+1/n}\mu_t(Af)dt + \textstyle\int_{i/n}^{i+1/n}\mu_t(fh)dY_t,$$
$$\mu_{i+1}(f) = \mu_i(f) + \textstyle\int_{i/n}^{i+1/n}\mu_t(Af)dt + \textstyle\int_{i/n}^{i+1/n}\mu_t(fh)dY_t.$$

Furthermore, we have $E\{\mu_i^n(f)\} = \mu_t(f)|_{t=\frac{i}{n}} = \mu_i(f)$ and $E\{\mu_i^n(f)|\mathcal{F}_{i-1}^Y\} = \mu_{i-1}^n(Af)$. Using this evolving-branching algorithm one can recursively construct μ_t^n, such that the relative entropy $\mathsf{H}(\mu_t^n|\mu_t)$ is minimal for all t if the mean number of the offsprings is

$$\Lambda_i^n(Z_t^k) = \exp(\textstyle\int_{\frac{i}{n}}^{\frac{i+1}{n}} h'(Z_t^k)(dY_t - \frac{1}{2}h(Z_t^k)dt))$$

and the resulting minimum variance is

$$v_i^n(Z_t^k) = (\Lambda_i^n(Z_t^k) - \lfloor \Lambda_i^n(Z_t^k) \rfloor)(\lfloor \Lambda_i^n(Z_t^k) \rfloor - \Lambda_i^n(Z_t^k) + 1).$$

3.2.2 Spectral approach

The spectral filter is based on the fact that the unormalized conditional density of the Zakai equation admits a decomposition in terms of Wick polynomials of Wiener integrals with respect to the observation process, and a deterministic Hermite-Fourier coefficients in the Cameron Martin orthogonal polynomial [43]. This expansion, a form of Wiener chaos expansion, separates the observation process from the signal process. The Wick polynomials are determined by the observation process, but the Hermite-Fourier coefficients are determined by the generator of the signal process X_t, the law of X_0, and the observation function h. The computation of the Hermite-Fourier polynomial can be performed off-line, while the updating of the Wick polynomials are performed in real time.

The decomposition of μ_t is

$$\mu_t = \sum\nolimits_\alpha \frac{1}{\sqrt{\alpha!}} \varphi_\alpha(t,x) \xi_\alpha(Y_t) \tag{3.3}$$

where $\xi_\alpha(Y_t(\omega))$ are Wick polynomials; a product of Hermite polynomials of Wiener integrals. $\varphi_\alpha(t,x)$ are deterministic Hermite-Fourier coefficients. $\alpha = (\alpha_k^l)_{1 \leq l \leq d, k > 1}$ is a d-dimensional multi-index meaning that only finitely many α_k^l are different from zero. $\alpha! = \prod_{k,l} \alpha_k^l!$.

Let $\xi_{k,l} = \int_0^t m_k(s) dY_s^l$, where $\{m_k(t)\}$ form a complete orthonormal system in $L^2([0,t])$. $\xi_{k,l}$ are independent identically distributed Gaussian variable. Let $(H_n)_{n \geqslant 1}$ be Hermite polynomials.

$$\xi_\alpha = \prod\nolimits_{k,l} \Big(\frac{H_{\alpha_k^l}(\xi_{k,l})}{\sqrt{\alpha_k^l!}}\Big)$$

is the Wick polynomial. The sequence $(\xi_\alpha)_\alpha$ form a complete orthonormal system in $L_2(\Omega, F_t^Y, \tilde{P})$. The corresponding coefficients $\varphi_\alpha(t,x)$ in the decomposition of μ_t satisfy a system of deterministic partial differential equations

$$\frac{d}{dt}\varphi_\alpha(t,x) = A'\varphi_\alpha(t,x) + \textstyle\sum_{k,l}\alpha_k^l m_k(t)h^l(x)\varphi_{\alpha(k,l)}(t,x)$$
$$\varphi_\alpha(0,x) = \pi(0,x)\mathbf{1}_{\{|\alpha|=0\}},\ |\alpha| = \textstyle\sum_{k,l}\alpha_k^l$$

Under certain technical assumptions one could prove that the expansion $\sum_\alpha \frac{1}{\sqrt{\alpha!}}\varphi_\alpha\xi_\alpha$ converges almost surely to μ_t [43].

To achieve fast computational speeds and efficiency one need to truncate the series 3.3. If we let $J_N^n = \{\alpha||\alpha| \le N, d(\alpha) \le n\}$, where $d(\alpha) = \max\{k \ge 1 : \alpha_k^l > 0$ for some $1 \le l \le d\}$ and choose

$$m_1(s) = 1/\sqrt{t}$$
$$m_k(s) = \frac{2}{\sqrt{t}}\cos(\frac{\pi(k-1)s}{t})$$

for $k \ge 1$ and $0 \le s \le t$ then one can prove that $\tilde{E}\{||\mu_t^{n,N} - \mu_t||_{L_2}^2\} \le C_t^1/(N+1)! + C_t^2/n$ where C_t^1 and C_t^2 are independent of n and N and $\mu_t^{n,N}$ is the truncated series 3.3, [43].

3.2.3 Neural Networks

Neural networks are general function approximators that can be used to approximate π_t or μ_t. But, they require that π_t and μ_t be approximately stationary, smooth, and essentially bounded.

There are many neural network methods applied to nonlinear filtering [44], [61], [22] and [32]. Here I will discuss the work of Haykin on nonlinear filtering [32].

Let $Z_t \in \mathbb{R}^r$ and $X_t \in \mathbb{R}^d$. Replace the observation process Y_t in equation 2.2 by Z_t as follows

$$dY_t = h(X_t)dt + dW_t = (h(X_t) + \dot{W}_t)dt \tag{3.4}$$
$$dY_t = Z_t dt$$
$$\therefore Z_t = h(X_t) + \dot{W}_t.$$

Let $S_n = \{(X_i, Z_i) \in \mathbb{R}^d \times \mathbb{R}^r\}_{i=1}^n$ be the neural network training set. S_n is obtained either by direct measurements performed on the physical devices giving rise to the signal and observation processes or by simulating equation 2.1 and 3.4.

Our neural network model approximates $\pi_t(f)$ as a weighted sum of radial basis functions.

$$\pi_t(f) \approx \textstyle\sum_{i=1}^{n} \lambda_i(f) g_i(Z_t),$$
$$g_i(\cdot) = \exp(-(\cdot - z_i)'\Sigma(\cdot - z_i))$$

where $\{z_i\}_{i=1}^n$ are chosen centers for the radial basis functions and Σ is weighting positive definite matrix. Let $\mathbf{g} = (g_1, ..., g_n)'$ and $\Lambda = (\lambda_1, \lambda_2, ..., \lambda_n)'$. Using the training set S_n we can solve for the matrix of weights Λ by variational interpolation

$$\Lambda_f = \arg\min_{\Lambda} \textstyle\sum_{i=1}^{n} [(f(X_i) - \Lambda \mathbf{g}(Z_i))^2 + \rho |\Lambda \mathbf{g}_z(Z_i)|^2].$$

$\mathbf{g}_z$ is the derivative of the radial basis function with respect to z and $\rho > 0$ is the regularization parameter. Λ_f minimizes both the error $(f(X_i) - \Lambda \mathbf{g}(Z_i))$ and the derivative $\mathbf{g}_z$. Therefore, we obtain a smooth approximation of $\pi_t(f)$ from the training set S_n. Knowing Λ_f, the approximate value of $\pi_t(f)$ is

$$\pi_t(f) \approx \Lambda_f \mathbf{g}(Z_t).$$

Λ_f incorporates information about the space spanned by the set of training inputs $\{Z_i \in \mathbb{R}^r\}_{i=1}^n$ and the test function f.

Chapter 4

Financial Applications

The role of filtering theory in finance is to uncover hidden parameters in financial systems such as volatility, drift, interest rate, mean interest rate, etc. that influence the evolution of observed market variables like assets or derivatives.

There is a growing literature on the application of filtering theory to finance. For example, fund managers are faced with the problem of investing capital in various assets in an optimal way. The fund manager keeps track of readily observable quantities such as assets prices, interest rates, inflation and industrial indices. But there are unobserved factors such as expected stock returns, psychology of the market and actions of competitors. Elliott and van der Hoek [24] represented the unknown factors by a hidden Markov chain whose states have to be estimated from observable quantities. Fischer [26], investigated the problem of hedging strategy by minimizing risk under incomplete information. Elliott, Fischer and Platen [25] looked at the estimation of the mean in the mean reverting interest rate model. In this case the mean is modeled as a finite state continuous time hidden Markov chain. And in bond pricing Landen [42] considered the short rate pricing of bonds driven by a finite state Markov process. The Markov process influences both the interest rate and price of bond. Gombani and Runggaldier [30], considered a filtering approach to pricing in multifactor term structure. Frey and Runggaldier [28] [29], looked at estimation of volatility in high frequency market data. The Heath-Jarrow-Morton (HJM) term structure family of interest rate models allows a finite dimensional Markov representation of the stochastic dynamics. The volatility in this case is a function of time to maturity, the instantaneous spot rate and one fixed maturity forward rate. Chiarella [11] considered an additional factor in the volatility function; the market price of interest rate risk, that connects the historical and the HJM martingale measures and produced a recursive estimation procedure for the three factors influencing the volatility function. Bhar [3] used Kalman filter to infer the forward looking equity risk premium from derivative prices.

In the following sections I will explain three filtering problems arising in finance: tracking volatility from asset prices observed at random times [15], inferring the forward looking equity risk premium from derivative prices [3] and hedging under partial observation [21].

4.1 Tracking volatility

The stock price is modeled as geometric Brownian motion

$$dS_t = mS_t dt + vS_t dB_t$$

with diffusion coefficient equal to vS_t. In this basic model v is a constant known as the volatility parameter.

The volatility parameter is important especially for option traders, investment banks, economic analysts. But direct observation of the volatility is not possible; its determination is not easy; and it is not really a constant parameter. It could depend on the underlying stock price, Brownian noise, and even other market variables such as interest rate, psychology of investors, political trends etc.

There are many models that generalize the constant volatility model to a stochastic volatility, where v itself is a stochastic process. Two basic classes of models emerged. One are complete models where v is a function of the stock price. The others are the incomplete models in which the volatility v is a function of some other sources possibly correlated with the driving Brownian motion.

Here I will discuss the incomplete model by Cvitanić, Liptser and Rozovskii [15]. They assumed that the stock price is given by

$$dS_t = m(\theta_t)S_t dt + v(\theta_t)S_t dB_t,$$

where θ_t the volatility process, is independent of the driving Brownian motion and is a strong Markov process with the generator A. θ_t is not observed but the functions $v(\cdot)$ and $m(\cdot)$ are known. $X_k = \log S_{\tau_k}$ is observed at random times $\{\tau_k\}_{k\geq 0}$. The process $(\tau_k, X_k)_{k\geq 0}$ is a multivariate point process with measure $\kappa(dt, dx)$.

The estimate of θ_t given $\mathcal{G}_t = \sigma(\kappa([0,r], B), r \leq t, B \in \mathcal{B}(\mathbb{R}))$; $\pi_t(f) = E\{f(\theta_t)|\mathcal{G}_t\}$ is given by the following result

$$\pi_{\tau_{k+1}}(f) = \left.\frac{\pi_{\tau_k}(\psi_k(f;t,y))}{\pi_{\tau_k}(\psi_k(1;t,y))}\right|_{(t=\tau_{k+1}, y=X_{k+1})} - \mathcal{M}_k(f;t,\pi_t)\big|_{t=\tau_{k+1}} \Phi(\{\tau_{k+1}\}),$$

$$d\pi_t(f) = \pi_t(Af)dt - \mathcal{M}_k(f;t,\pi_t)\Phi(dt) \quad \text{for } t \in (\tau_k, \tau_{k+1}).$$

where ψ_k is known as the structure constant, $\mathcal{M}_k$ and Φ are all defined in [15].

From the above equations we see that estimating volatility from observed stock prices is not a trivial task. But the system of equations give a closed form solution to the problem.

4.2 Inferring risk premium

Bhar [3] proposed a solution to the problem of inferring the forward looking equity risk premium from derivative prices. The equity risk premium is the difference between the expected rate of return on stock(s) investment and the risk-free rate.

Bhar, Chiarella, and Runggaldier [3] used a system of equations relating the index price, the futures on the index, options on those futures and implied volatility to estimate the unobserved risk premium. The risk premium is modelled by a mean-reverting process.

Let S_t be the index value, F_t a future contract value on the index, and C_t an option on the future. All these quantities are observed signals. λ_t is the instantaneous market price of risk.

It is the signal we want to estimate. The system of equations that link all these variables are

$$
\begin{aligned}
d\lambda_t &= k(\bar{\lambda} - \lambda_t) + v_\lambda dB_t \\
d\log S_t &= (r - q + \lambda v - \frac{1}{2}v^2)dt + vdW_t \\
d\log F_t &= (\lambda v - \frac{1}{2}v^2)dt + vdW_t \\
d\log C_t &= (r + \lambda v_c - \frac{1}{2}v_c^2)dt + v_c dW_t
\end{aligned}
$$

v is the volatility of the index. r is the risk-free interest rate. q is the continuous dividend yield on the index. v_c is the option return volatility and is calculated from the Black-Scholes model. $\bar{\lambda}$ is the long-run value of λ, k is the speed of mean reversion and v_λ is the volatility of λ.

The volatility v is not a constant parameter. A possible way to cope with its non-constancy is to develop a stochastic volatility model. But then we will not have the simple option pricing model. A practical solution is to use implied volatility calculated from market prices using Black's model.

Given that we know all the parameters in the above system of equations then we just use the Kalman filter to estimate λ.

4.3 Hedging under partial observation

Here I discuss an optimal hedging strategy under imprecise knowledge of the asset price [4]. Consider a financial market over the time interval $[0, T]$ with a risky asset price S_t, a bond price $D_t = 1$ $\forall t \in [0, T]$. and portfolio H_t. The stochastic model of the problem

$$
\begin{aligned}
dS_t &= \sigma(Z_t)S_t dB_t, \\
d\eta_t &= \alpha_t S_t dt + \beta_t dW_t, \\
H_t &= \psi_t S_t + \varphi_t D_t = \psi_t S_t + \varphi_t \\
X_t &= H_t - \textstyle\int_0^t \psi_u dS_u
\end{aligned}
$$

The process Z_t represent the state of the economy and is modelled by Markov process with transition matrix A and N_t number of jumps in the economy. The process η_t is the observed asset price under additive noise. $Y_t = (\eta_t, N_t)'$ is the observed process, S_T is also known. (ψ_t, φ_t) is our hedging strategy. And the contingent claim price at time T is H.

The goal is to find hedging strategies (ψ_t, φ_t) such that $H_T = H$ and the risk $E\{(X_T - X_t)^2 | \mathcal{F}_t^Y\}$ is minimum for all t. The solution of this second problem was given by Di Masi et al. [21].

Acknowledgement 7 *I took ideas, text, and topics organization from the following sources [36], [20], [46], [4], [3], [54], and [45]. Also, I would like to extend my gratitude to Tom Salisbury who facilitated the completion of this survey paper.*

Bibliography

[1] V. E. Bene's, Exact finite dimensional filters for certain diffusions with nonlinear drift, Stochastics, 5 (1981), pp. 65-92.

[2] A. Bensoussan, Stochastic Control with Partial Information, Cambridge University Press, 1992.

[3] Ramaprasad Bhar, Carl Chiarella, and Wolfgang J. Runggaldier (2004) "Inferring the Forward Looking Equity Risk Premium from Derivative Prices", Studies in Nonlinear Dynamics & Econometrics: Vol. 8: No. 1, Article 3. http://www.bepress.com/snde/vol8/iss1/art3

[4] D. Brigo, and B. Hanzon, On some Filtering Problems arising in Mathematical Finance, Insurance : Mathematics and Economics 22 (1998), 53-64.

[5] Brigo D., Hanzon, B. and Le Gland, F. A differential geometric appraoch to nonlinear filtering: the projection filter, IEEE Tans. on Automatic Control, 43. No. 2, 247-252 (1998)

[6] R.W. Brockett, Remarks on finite dimensional nonlinear estimation, Analyse des syetémes, C. Lobry, ed., Bordeaux, 1978; also, Astérisque, 76 (1980), Soc. Math. de France.

[7] R.W. Brockett, Classification and equivalence in estimation theory, Proc. 18th IEEE Conference on Decision and Control, Ft. Lauderdale, FL, 1979, p. 172-175.

[8] R.W. Brockett, Nonlinear systems and nonlinear estimation theory, Stochastic Systems: The Mathematics of Filtering and Identification and Applications, M. Hazewinkel and J. C. Willems, eds., Reidel, Dordrecht, The Netherlands, 1981, p. 441-478.

[9] R. W. Brockett and J. M. C. Clark, The geometry of the conditional density equation, Analysis and Optimization of Stochastic Systems, O. L. R. Jacobs et al., eds., New York, 1980, p. 299-309.

[10] Zhe Chen, Bayesian Filtering: From Kalman Filters to particle Filters, and Beyond, http://users.isr.ist.utl.pt/~jpg/tfc0607/chen_bayesian.pdf

[11] C.Chiarella, S.Pasquali, W.J. Runggaldier, On Filtering in Markovian Term Structure Models (an approximation approach). Advances in Applied Probability 33 (2001), pp.794-809.

[12] J.M.C. Clark, The design of robust approximations to the stochastic differential equation of nonlinear filtering, in Communication Systems and Random Process Theory, ed. J.K Skwirzynski, NATO Advanced Study Institute Series, Sijth off and Noordhoff, Alphen aan den Rijn 1978.

[13] Dan Crisan, Terry Lyons, A particle approximation of the solution of the Kushner–Stratonovitch equation, Probab. Theory Relat. Fields 115, 549–578 (1999).

[14] Dan Crisan, Jessica Gaines, and Terry Lyons Convergence of branching particle method to the solution of the Zakai equation. SIAM J. APPL. MATH. Vol. 58, No. 5, pp. 1568-1590, October 1998

[15] Cvitanic-Liptser-Rozovski (2006) "A filtering approach to tracking volatility from prices observed at random times". The Annals of Applied Probability, 16, 1633-1652.

[16] R. W. R. Darling (1998), Geometrically intrinsic nonlinear recursive filters I: Algorithms. Report 493, available at http://www.stat.berkeley.edu/tech-reports/

[17] R. W. R. Darling (1998), Geometrically intrinsic nonlinear recursive filters II: foundations. Report 512, available at http://www.stat.berkeley.edu/tech-reports/

[18] R. W. R. Darling (1997), Intrinsic location parameter of a diffusion process. Report 493, Report 494, available at http://www.stat.berkeley.edu/tech-reports/

[19] F. Daum, Exact Finite-Dimensional Nonlinear Filters, lEEE transactions on automatic control, vol. AC-31. NO. 7, JULY 1986

[20] M. E. A. Davis, Lectures on stochastic control and nonlinear filtering, Springer-Verlag, Berlin Heidelberg New York Tokoyo 1984.

[21] G.B. Di Masi, Yu. M. Kabanov, W.J. Runggaldier, Mean-variance hedging of options on stocks with Markov volatilities, Theory of Probability and Its Applications, 39 (1994), pp. 211-222

[22] Jun Morimoto, Kenji Doya, Reinforcement Learning State Estimator, Neural Computation 19(3): 730-756 (2007)

[23] T. E. Duncan, Probability densities for diffusion processes with applications to nonlinear filtering theory, Ph.D. dissertation, Stanford Univ., Stanford, CA, 1967.

[24] Robert J. Elliott, and John Van Der Hoek, An Application of Hidden Markov Models to Asset Allocation Problems FINANCE AND STOCHASTICS, Vol. 1 No. 3, 1997

[25] R. Elliott, P. Fischer and Eckhard Platen, Filtering and Parameter Estimation for a Mean Reverting Interest Rate Model, No 17, Research Paper Series from Quantitative Finance Research Centre, University of Technology, Sydney 1999.

[26] Fischer, P., Platen, E., Runggaldier, W.J., 1996. Risk-minimizing hedging strategies under partial observation. Proceedings of the 1996 Ascona Seminar on Stochastic Analysis, Random Fields and Applications.

[27] M. Fujisaki, G. Kallianpur & H. Kunita, Stochastic differential equations for the non linear filtering problem, Osaka J. Math. vol. 9, p. 19-40, 1972.

[28] R. Frey, W.J. Runggaldier, Risk-minimizing hedging strategies under restricted information : The case of stochastic volatility models observable only at discrete random times. Mathematical Methods of Operations Reasearch 50 (1999), pp. 339-350.

[29] R. Frey, W.J. Runggaldier, A Nonlinear Filtering Approach to Volatility Estimation with a View Towards High Frequency Data International Journal of Theoretical and Applied Finance 4 (2001), pp.199-210.

[30] A.Gombani, W.J. Runggaldier, A filtering approach to pricing in multifactor term structure models. International Journal of Theoretical and Applied Finance 4 (2001), pp.303-320.

[31] Harvey A.C., Time series models, Published 1981 by Philip Allen Publisher Limited, Oxford

[32] S. Haykin, P. Yee, E. Derbez, Optimal nonlinear filtering, IEEE transaction of signal processing, vol 45, no. 11 Nov 1997.

[33] M. Hazenwinkel, S. I. Marcus, On Lie algebras and finite dimensional filtering, Stochastics, 7 (1982), pp. 29-62.

[34] A. H. Jazwinski, Stochastic processes and filtering theory, volume 64 of Mathematics in science and engineering. Academic Press, 1970.

[35] A. N. Kolmogorov, Sur l'interpolation et extrapolatton dessuites stationnaires, C. R. Acud. Sci., vol. 208, p. 2043, 1939.

[36] Thomas Kailath, A View of Three Decades of Linear Filtering Theory, IEEE Transactions on Information Theory, vol. IT-20, no.2, March 1974.

[37] R. E. Kalman, A new approach to linear filtering and prediction problems, J. Basic Eng., vol. 82, p. 3445, Mar. 1960.

[38] R. E. Kalman, New methods of Wiener filtering theory, in Proc. 1st Symp. Engineering Applications of Random Function Theory and Probability, J. L. Bogdanoff and F. Kozin, Eds. New York: Wiley, 1963, p. 270-388; also, RIAS, Baltimore, Md., Tech. Rep. 61-1, 1961.

[39] R. E. Kalman and R. S. Bucy, New results in linear filtering and prediction theory, Trans. ASME, Ser. D, J. Basic Eng., vol. 83, p. 95-107, Dec. 1961.

[40] H. J. Kushner, On the differential equations satisfied by conditional probability densities of Markov processes with applications, SIAM J. Contr., vol. 2, p. 106–119, 1965.

[41] L. Arasaratnam, I. Haykin, S. Elliott, R.J., Discrete-Time Nonlinear Filtering Algorithms Using Gauss–Hermite Quadrature, IEEE Proceedings, vol 95, issue 5, page 953-977, May 2007

[42] C. Landen, Bond Pricing in a Hidden Markov Model of the Short Rate, Finance and Stochastics, Vol. 4, Issue 4.

[43] S. Lototsky, R. Mikulevicius, and B. L. Rozovskii, Nonlinear Filtering Revisited: A Spectral Approach, SIAM J. on Control and Optim., 1997.

[44] James Ting-Ho Lo, Neural network approach to optimal filtering, Proceedings of the first world congress of nonlinear analysts 92, volume IV Tampa, Florida, United States Pages: 3705 - 3716 Year of Publication: 1995.

[45] T. J. Lyons T. S. Salisbury, Numerical methods and stochastics, American Mathematical society. 2002.

[46] Steven I. Marcus Algebraic and Geometric Methods in nonlinear filtering, SIAM J. CONTROL AND OPTIMIZATION Vol. 22, No. 6, Nov 1984.

[47] Hisashi Tanizaki & Roberto S. Mariano, Nonlinear and Non-Gaussian State-Space Modeling with Monte-Carlo Simulations, Journal of Econometrics, Vol.83, No.1&2, pp.263-290, 1998.

[48] S. K. Mitter, Filtering theory and quantum fields, Analyse des syst6mes, C. Lobry, eds., Bordeaux, 1978; also, Ast6risque, 76 (1980).

[49] S. K. Mitter, On the analogy between mathematical problems of non-linear filtering and quantum physics, Richerche di Automatica, 10 (1980), p. 163-216.

[50] R. E. Mortensen, Optimal control of continuous time stochastic systems, Ph.D. dissertation, Univ. Calif., Berkeley, CA, 1966.

[51] Jean Picard, Efficiency of the extended Kalman filter for nonlinear systems with small noise, SIAM Journal on Applied Mathematics archive vol 51 , issue 3, Pages: 843 - 885 June 1991

[52] A New Extension of the Kalman Filter to Nonlinear Systems, Simon J. Julier Jeffrey K. Uhlmann, SPIE Proceedings, Vol. 3068 pp.182-93, 1997

[53] R. L. Stratonovich, Conditional Markov processes, Theor. Prob. Appl. (USSR), vol. 5, p. 156–178, 1960.

[54] Hisashi Tanizaki, Nonlinear Filters: Estimation and Applications, Published 1993 Springer

[55] R.P. Wishner, J.A. Tabaczyski, and M. Athans. A Comparison of Three Nonlinear Filters. Automatica, 5:487–496, 1969.

[56] J. Wei and E. Norman, On the global representation of the solutions of linear differential equations as a product of exponentials, Proc. Am. Math. Soc., 15 (1964), pp. 327-334.

[57] N. Wiener, Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications, New York: Technology Press and Wiley, 1949. (Originally issued in February 1942, as a classified Nat. Defense Res. Council Rep.)

[58] N. Wiener and E. Hopf, On a class of singular integral equations, Proc. Prussian Acad., Math.-Phys. Ser., p. 696, 1931.

[59] S.S.-T. Yau, Finite dimensional filters with nonlinear drift I: A class of filters including both Kalman-Bucy filters and Benes filters, J. Math. Systems, Estimation and Control 4, no. 2 (1994), 181-203.

[60] M. Zakai, On the optimal filtering of diffusion processes, Z. Wahrsch, Verw. Geb., vol. 11, p. 230–243, 1969.

[61] T. Parisini, R. Zoppoli, Neural networks for nonlinear state estimation, International Journal of Robust and Nonlinear Control Volume 4, Issue 2 , Pages 231 - 248, Mar 2007.