BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 48, Number 2, April 2011, Pages 293–305
S 0273-0979(2010)01311-2
Article electronically published on October 26, 2010

Fundamentals of stochastic filtering, by Alan Bain and Dan Crisan, Stochastic Modelling and Applied Probability, 60, Springer, New York, 2009, xiv+390 pp., ISBN 978-0-387-76895-3

1. Introduction

Unless you, the reader, work in signal processing or in probability theory, it is likely that you have never heard the term *filtering* or the rather strange and sophisticated probabilistic adjective *stochastic*.

Nowadays, the term stochastic is often reduced to reflect the uncertainty or the randomness of a given object or probabilistic model. In fact, this term is derived from the ancient Greek word $\sigma\tau\acute{o}\chi o\varsigma$ (stókhos) which means aim, guess, target, or able to conjecture. In this sense, stochastic models are better defined as random processes or strategies based on random guesses used to predict, to infer, or to capture the unknown evolution of some physical or biological quantities.

In this interpretation, stochastic filtering can be defined as a natural online and automated probabilistic or statistical inference technique based on learning, using predictions, the information encapsulated in a series of noisy and partial observations delivered by some sensor. For instance, in signal processing, filtering a given process consists in removing some unwanted components, such as noisy inaccuracies coming from sensor perturbations. As an example of an application, consider the problem of determining the precise trajectory and the kinetic parameters of a noncooperative aircraft target given a series of noisy distance measurements delivered by some radar type sensor. In this situation, the filtering problem consists in estimating the true trajectory, including the velocity and the acceleration of the target, given these noisy and partial observations.

This book is one of the few books dealing with both the theoretical foundations and modern stochastic particle techniques in stochastic filtering through the entire text. It also provides a rather complete historical account of the origins of filtering, starting with the celebrated works of famous researchers, such as Wiener, Stratonovich, Kolmogorov, Krein, Kushner, Kalman, and Bucy in the 1950s and the mid-1960s, on discrete and continuous time estimation problems of dynamical systems in the presence of noisy and partial observations. These pioneering works were followed from the late-1960s to the late 1970s by a series of theoretical works by another series of famous mathematicians, including Shiryaev, Kailath, Fujisaki, Kunita, Zakai, Krylov, Rosovskii, Pardoux, Duncan, Clark, and Davis, on the rigorous derivations of the optimal filtering equations in terms of stochastic partial differential equations or by functional representation of conditional expectations of stochastic processes.

Unfortunately, there are only very few examples of filtering problems for which the optimal filter is finite dimensional, in the sense that the optimal solution can be solved by a computer. To name a few, finite state space filters, the Kalman–Bucy filter discussed above, the Beneš filter. More precisely, in the mid-1980s, the work of two French probabilists, Chaleyat-Maurel and Michel, shows that it was hopeless to try to find a finite-dimensional representation of the optimal filter! Generically,

2010 *Mathematics Subject Classification.* Primary 60G35, 93E11.

and apart from a few exceptions, the filtering problem is infinite dimensional and it cannot be solved without approximations on a real-world computer.

Much of the work carried out since the 1990s has focused on numerical solutions of the filtering problem. Six important classes of numerical methods are nowadays available: the extended Kalman filter based on local linearization methods, approximations by the exact finite-dimensional filters, projection techniques and the moment methods, spectral approximations, partial differential equation methods, and the more recent and advanced stochastic particle interpretation models.

2. Stochastic filtering and Bayesian inference

To better understand these models and the numerical problems discussed in this book, let us present more formally these probabilistic models. The central filtering problem is to compute recursively in time the conditional distributions of the random possible trajectories of a given target signal given the series of observations delivered by some measurement sensor. Of course, the rigorous mathematical description of these conditional distributions requires that we identify precisely the different sources of randomness in these physical and engineering models. This modeling strategy is well known in Bayesian statistics. Firstly, before estimating the realization of some random state variable, say X, we need to have some existing *prior* belief on the possible random values it may take. Given some correlated observation random variable, say Y, these prior distributions $p(x)$ are updated to give conditional distributions $p(x|y)$, also called *posterior* distributions. Up to some normalizing constant, this updating step amounts to multiplying the prior distribution by some likelihood function $p(y|x)$ that measures the adequacy of the prior first guesses w.r.t. the observations; that is, we have that $p(x|y) \propto p(y|x)p(x)$. This elementary formula is sometimes called the Bayes' theorem or the Bayes' rule. Of course, this heuristic type rule is intuitively convincing, but its rigorous derivation requires a little more probabilistic technicalities.

In signal processing and stochastic filtering, these prior distributions and the observation likelihood functions are dictated by the physical model at hand. The sources of randomness in the target evolution probabilistic model comes from both the unknown control strategies driving the target and the mathematical model uncertainties. In much the same way, the random sources in the observation model mainly come from the probabilistic model uncertainties, the environmental perturbations as well as internal sensor perturbations, such as thermic noise in electronic type sensor devices.

Now, the reader should be convinced of the huge potential range of applications of stochastic filtering. To name a few traditional and some more recent applications, this stochastic learning technique plays a central role in the following scientific fields: object tracking and vehicle positioning using radar, sonar, or GPS sensors, satellite navigation systems, volatility estimation in financial markets, weather forecasting and turbulent fluid motion tracking, speech recognition, computer vision and pattern recognition, time series estimations in econometrics, and many others.

3. The Kalman filter

Due to the applicability of stochastic filtering in such a diverse range of applications, this subject has been the focus of enormous amount of research in the last fifty years. The origins of this research can be traced back to the pioneering and

famous article of Rudolf E. Kalman, *A new approach to linear filtering and prediction problems*, published in 1960 in the Journal of Basic Engineering 82 (1): 35–45. Briefly speaking, the central idea is to conduct inductively the following fact.

Suppose that X_0 is a Gaussian state variable with mean $\widehat{X}_0^-$ and covariance matrix $\widehat{P}_0^-$. We also assume that the associated observation has the form, $Y_0 = C\ X_0 + V_0$, for some independent Gaussian perturbation V_0 and some given matrix C. In this case, the conditional distribution of X_0 given Y_0 is again Gaussian with a mean $\widehat{X}_0$ and a covariance matrix $\widehat{P}_0$ that can be explicitly computed in terms of Y_0 and the pair of parameters $(\widehat{X}_0^-, \widehat{P}_0^-)$. This first step is often called the updating stage with respect to the current observation Y_0.

In the same way, if the state variable X_0 changes into a new random state $X_1 = AX_0 + BW_1$, for some pair of matrices (A, B), and some independent Gaussian perturbation W_1, then the conditional distribution of X_1 given Y_0 is again Gaussian with a mean $\widehat{X}_1^-$ and a covariance matrix $\widehat{P}_1^-$ that can be explicitly computed in terms of Y_1 and the previous pair of parameters $(\widehat{X}_0, \widehat{P}_0)$. This second step is named the prediction stage in the sense that we estimate the current state X_1 using the last observation Y_0. Iterating this updating-prediction learning transition, in this linear Gaussian world we easily derive the Kalman filter recursion.

This rather simple algorithm is surprisingly extremely powerful. For instance, it has been applied by NASA to the problem of trajectory estimation for the Apollo program to get the missions off the ground and on the moon. It is still nowadays the subject of extensive research projects and applications, particularly in the area of autonomous or assisted navigation systems, as well as in tracking problems using GPS, radar, or sonar measurements. Now, a series of important and natural questions arise:

- What can we do if the random states and/or the observation sensors are neither linear nor Gaussian? Can we solve numerically on some computer the filtering problem for more general models?
- How do we choose in practice the initial parameters of the recursion? What happens if we start with some erroneous initial condition?
- Is there some continuous time version of the above filtering models?
- Are all of these stochastic filtering models stable and robust w.r.t. the observation measurements?

Most of the answers to these central questions can be found in this book, including a rather detailed historical account on the main scientists that contributed to this research project since the mid-1990s.

4. Discrete time filters

In the first part of the book, the authors give a theoretical presentation of both the discrete time filtering equation as well as the continuous time filtering equation for general diffusion type stochastic processes.

The discrete time probabilistic model can be described rather easily. Suppose we are given a pair signal-observation Markov chain $(X_n, Y_n)_{n\geq 0}$ on some product space, with some initial distribution and Markov transitions of the form

$$\begin{aligned} \mathbb{P}\left((X_0, Y_0) \in d(x,y)\right) &= \eta_0(dx)\ g_0(x,y)\ \lambda_0(dy), \\ \mathbb{P}\left((X_{n+1}, Y_{n+1}) \in d(x,y) | (X_n, Y_n)\right) &= M_{n+1}(X_n, dx)\ g_{n+1}(x,y)\ \lambda_{n+1}(dy) \end{aligned}$$

with some reference probability measures λ_n on the observation state space. The prototype stochastic model is the one given by the $\mathbb{R}^2$-recursive formula

$$\left\{ \begin{array}{rcl} X_n & = & a_{n-1}(X_{n-1}) + W_n, \\ Y_n & = & b_n(X_n) + V_n, \end{array} \right.$$

with some regular functions (a_n, b_n) and some independent and centered Gaussian random variables (W_n, V_n) with unit variance. In this elementary case, we have

$$M_{n+1}(X_n, dx) = \frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}(x-a_n(X_n))^2} dx$$

and

$$g_n(x,y) = e^{b_n(x)y - \frac{1}{2}b_n^2(x)} \quad \text{with} \quad \lambda_n(dy) = \frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}y^2} dy. \tag{4.1}$$

For a given sequence of observations $Y = y$, the optimal one step predictor η_n and the optimal filter $\widehat{\eta}_n$ are defined by

$$\begin{array}{rcl} \eta_n & = & \text{Law}(X_n | Y_0 = y_0, \ldots, Y_{n-1} = y_{n-1}), \\ \widehat{\eta}_n & = & \text{Law}(X_n | Y_0 = y_0, \ldots, Y_{n-1} = y_{n-1}, Y_n = y_n). \end{array} \tag{4.2}$$

For $n = 0$, we use the convention $\eta_0 = \text{Law}(X_0)$. The updating of the measure η_0 w.r.t. the current observation $Y_0 = y_0$ is simply given by the Bayes' formula

$$\widehat{\eta}_0(dx) := \frac{g_0(x, y_0)}{\int \eta_0(dx')\ g_0(x', y_0)}\ \eta_0(dx).$$

Notice that the probability mass variation is dictated here by the so-called likelihood function g_0 and the observation data y_0. During the prediction transition, we predict the new value of the random state X_1 using the following linear transport equation dictated by the a priori Markov transitions M_1 of the signal

$$\eta_1(dx') := \int\ \widehat{\eta}_0(dx)\ M_1(x, dx').$$

Iterating these two procedures, we update the measure η_1 w.r.t. the current observation $Y_1 = y_1$ (using again the Bayes' formula), and we predict the next value X_2 using the linear transport equation (dictated by M_2), and so on.

For linear models (with a Gaussian initial distribution η_0), these distributions are Gaussian and these recursive formulae can be solved explicitly by the Kalman filter discussed above. Also notice that for finite state spaces, the above updating-prediction transitions reduce to elementary finite sums and multiplication operators on finite-dimensional vectors.

Finally, it is worth mentioning that the conditional distributions η_n and $\widehat{\eta}_n$ have a natural functional Feynman–Kac representation. For instance, the one step optimal predictors are given by the formula

$$\eta_n(f) = \frac{\gamma_n(f)}{\gamma_n(1)} \quad \text{with} \quad \gamma_n(f) := \int f(x)\gamma_n(dx) = \mathbb{E}_y \left(f_n(X_n) \prod_{0 \leq p < n} g_p(X_p, y_p) \right). \tag{4.3}$$

In the above displayed formula, $\mathbb{E}_y(.)$ stands for the expectation w.r.t. the law of X_t, with a fixed observation process $Y = y$. That is, we have that

$$\gamma_n(f) = \int f_n(x_n) \left\{ \prod_{0 \leq p < n} g_p(x_p, y_p) \right\} \eta_0(dx_0) M_1(x_0, dx_1) \cdots M_n(x_{n-1}, dx_n).$$

On the other hand, the unnormalized measures γ_n can be expressed in terms of the flow of measures $(\eta_p)_{0 \leq p < n}$, with the multiplicative formula

$$\gamma_{n+1}(1) = \gamma_n(g_n(., y_n)) = \eta_n(g_n(., y_n)) \times \gamma_n(1) = \prod_{0 \leq p \leq n} \eta_p(g_p(., y_p)). \tag{4.4}$$

These abstract formulations have a natural interpretation in terms of Bayesian posterior densities. More precisely, in a formal way we have

$$p_n(y_0, \ldots, y_n | x_0, \ldots, x_n) = \prod_{0 \leq p \leq n} g_p(x_p, y_p) \quad \text{and} \quad p_n(y_0, \ldots, y_n) = \gamma_{n+1}(1).$$

The multiplicative formula (4.4) can be used to approximate the unnormalized measures in terms of approximations of the normalized probability measures.

5. Continuous time filters

The rigorous mathematical derivation of continuous time filtering equations is technically much more involved. As shown in the book, two different approaches can be used: the change of measure method and the innovation process technique.

Roughly speaking, the first one can be thought of as a sophisticated extension of the Bayes rule for discrete stochastic processes to continuous time signal-observation models. The corresponding conditioning principle is known as the Kallianpur–Striebel formula. In theory, at least, this formula provides a method for solving the filtering problem. As in the Bayes' identity, it represents the optimal filter as a functional normalized ratio of a weighted unnormalised distribution.

The second approach isolates the diffusion term of the signal and the corresponding linear *heat* type transport term, and then identifies the nonlinear observation mass variations in the Doob–Meyer decomposition of the optimal filter. This approach shows that the conditional distribution of the signal is a solution of a nonlinear (quadratic type) stochastic and parabolic type partial differential equation often called the Kushner–Stratonovich equation. Its unnormalised weighted version discussed above satisfies a linear equation often called the Zakai equation. The developments around these two probabilistic approaches rely on sophisticated and advanced stochastic calculus.

As in the discrete time case, the prototype continuous filtering model is given by the $\mathbb{R}^2$-valued diffusion

$$\left\{ \begin{array}{rcl} dX_t & = & a_t(X_t)dt + dW_t, \\ dY_t & = & b_t(X_t)dt + dV_t, \end{array} \right.$$

with some regular functions (a_t, b_t) and some independent Brownian motions (W_t, V_t), with unit variance. To simplify the presentation, we assume that $Y_0 = 0$. In further development of this review, we briefly introduce the optimal filter evolution equations associated with this elementary filtering model in a rather formal way.

5.1. **A brief introduction to stochastic calculus.** We recall that the infinitesimal generator L_t associated with the signal diffusion X_t defined above is the second-order differential operator given by

$$L_t(f) := a_t \ \frac{\partial f}{\partial x} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}.$$

Roughly speaking, using a second-order Taylor type expansion, for any infinitely differentiable test function f, one can write in evocative shorthand the formula

$$df(X_t) \quad = \quad \frac{\partial f}{\partial x}(X_t)dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}dX_t dX_t + \cdots = L_t(f)(X_t)dt + dM_t(f)$$

with the martingale term $dM_t(f) := \frac{\partial f}{\partial x}(X_t)\ dW_t$. The above formula can be obtained in an formal way using the rule $dX_t dX_t \simeq dW_t dW_t \simeq dt$. The magnitude of the stochastic *remainder* martingale term $M_t(f)$ given above is often measured in terms of its predictable angle bracket $\langle M(f)\rangle_t$. We recall that this predictable process is such that $M_t(f)^2 - \langle M(f)\rangle_t$ is again a martingale, and it is defined by

$$\langle M(f)\rangle_t = \int_0^t \Gamma_{L_t}(f,f)(X_s)\ ds$$

in terms of the *carré du champ* operator associated with L and defined as

$$\Gamma_{L_t}(f,f) = L_t(f^2) - 2fL_t(f) = \left(\frac{\partial f}{\partial x}\right)^2.$$

This stochastic analysis technique works for a rather general class of infinitesimal generators L, including jump type models. This Taylor type second-order expansion is known as the Itô formula. For further details on this subject, we refer the reader to the seminal books by Daniel Revuz and Marc Yor [13], Ioannis Karatzas and Steven Shreve [10], Philip Protter [12], and Stewart Ethier and Thomas Kurtz [8]. The infinitesimal generator L plays a fundamental role in the stochastic analysis of Markov processes and parabolic type partial differential equations. Ignoring, as we will do throughout most of this review, that theorems need specific assumptions to make them true, if we let μ_t be the law of the random state X_t and set

$$\mu_t(g) := \mathbb{E}(g(X_t)) = \int_{\mathbb{R}^d} \mu_t(dx)\ g(x)$$

for any sufficiently regular function g from $\mathbb{R}^d$ into $\mathbb{R}$, then from (5.1) below we readily find that μ_t solves the second-order partial differential equation

$$\frac{d}{dt}\ \mu_t(f) = \mu_t(L_t(f)).$$

5.2. **The Zakai and the Kushner–Stratonovich equations.** As in the Bayes' formula, the conditional distribution π_t of the random states X_t given the observations from the origin up to the current time t can be represented as the normalized ratio of a weighted positive measure ρ_t; that is, we have that

$$(5.1) \qquad \pi_t(f) = \mathbb{E}(f(X_t)|Y_s, s \le t) = \rho_t(f)/\rho_t(1) \quad \text{with} \quad \rho_t(f) = \mathbb{E}_Y\left(f(X_t)Z_t\right)$$

and the likelihood function Z_t is given as in (4.1) by the formula

$$\log Z_t = \int_0^t b_s(X_s)dY_s - \frac{1}{2}\int_0^t b_s^2(X_s)ds \ \ (\Longleftrightarrow \ dZ_t = Z_t b_t(X_t)dY_t)\,.$$

In the above displayed formula $\mathbb{E}_Y(.)$ stands for the expectation w.r.t. the law of X_t, with a fixed observation process Y.

Roughly speaking, using the fact that

$$\begin{aligned} d\rho_t(f) &= \mathbb{E}_Y\left(Z_t\ df(X_t)\right) + \mathbb{E}_Y\left(f(X_t)\ dZ_t\right) \\ &= \mathbb{E}_Y\left(Z_t L_t(f)(X_t)\right) dt + \mathbb{E}_Y\left(f(X_t) Z_t b_t(X_t)\right)\ dY_t, \end{aligned}$$

we find that the measure valued process ρ_t satisfies the following evolution equation, called the Zakai equation,

$$d\rho_t(f) := \rho_t(L_t(f))dt + \rho_t(fb_t)\ dY_t.$$

Using the fact that $L_t(1) = 0 \Rightarrow d\rho_t(1) := \rho_t(b_t)\ dY_t = \rho_t(1)\pi_t(b_t)\ dY_t$, we also prove that

$$\log \rho_t(1) = \int_0^t \pi_s(b_s) dY_s - \frac{1}{2}\int_0^t \pi_s(b_s)^2 ds,$$

from which we deduce that

$$\log\left(Z_t/\rho_t(1)\right) = \int_0^t [b_s(X_s) - \pi_s(b_s)][dY_s - \pi_s(b_s)ds] - \frac{1}{2}\int_0^t [b_s(X_s) - \pi_s(b_s)]^2 ds.$$

Arguing as above, replacing b_s by the centered function $[b_s - \pi_s(b_s)]$ and dY_s by the so-called innovation process $[dY_s - \pi_s(b_s)ds]$, we prove that the optimal filter π_t satisfies the following stochastic partial differential equation, called the Kushner–Stratonovich equation

$$d\pi_t(f) := \pi_t(L_t(f))dt + \pi_t(f(b_t - \pi_t(b_t)))\ (dY_t - \pi_t(b_t)dt).$$

The rigorous derivations of the above probabilistic models and their multidimensional versions are discussed in full detail in the book under review. The authors also provide a very nice discussion of the existence and the smoothness properties of a density w.r.t. the Lebesgue measure of the conditional distributions.

Of course, real life observation processes are clearly far from being the perfect solution to a diffusion equation, in the sense that we never observe the complete continuous time observation trajectories. From this viewpoint, one might think that these evolution equations are useless! Nevertheless, the continuous time evolution equations derived above enable the use of traditional numerical approximation tools for partial differential equations, such as spectral methods, deterministic grid type models and splitting-up techniques. As underlined in the book, these techniques can be very successful in low-dimensional and regular filtering problems, but they cannot be applied in high-dimensional models or in nonlinear filtering models involving high probability mass variations.

5.3. **The robust filtering equation.** In practice, the observations delivered by any type of measurement sensors always arrive at discrete moments in time. Of course, using a simple linear interpolation, we almost obtain the solution of the sensor equation, but still real life observation data are never the output of a formalized mathematical model! It is therefore essential to design a continuous time filter representation that enjoys continuity properties w.r.t. the observation process. To solve this important question, a more sensible mathematical model is provided by the so-called robust version of the conditional distributions flows discussed in the previous section. The term *robust* refers to the fact that the dependence on the observation data is continuous. As mentioned in [1], such robust formulation was stated in the late-1970s by J.M.C. Clark in [2] without proof. More recent

treatments with rigorous mathematical developments can be found in the pair of articles [1, 4].

At a formal level, the robust filtering equation is a Feynman–Kac–Schrödinger type evolution that can be derived using an elementary integration-by-parts formula combined with an exponential change of reference probability measure. Because of its importance in practice and its interesting connections with Schrödinger semigroups and their mean field particle interpretations, we provide a direct informal proof of these models.

Firstly, using Ito's integration-by-parts formula, we find that

$$\begin{aligned} b_t(X_t)dY_t &= d(b_t(X_t)Y_t) - Y_t db_t(X_t) \\ &= db'_t(X_t) + \left[e^{b'_t} L_t(e^{-b'_t})\right](X_t)dt + d\log Z'_t \end{aligned}$$

with the random function $b'_t(x) = Y_t b_t(x)$ and the exponential martingale Z'_t defined in terms of $dM_t(b'_t) := -Y_t dM_t(b_t)$ by the formula

$$\begin{aligned} d\log Z'_t &:= dM_t(b'_t) - \left[e^{b'_t} L_t(e^{-b'_t}) + L_t(b'_t)\right](X_t)dt \\ &= dM_t(b'_t) - \frac{1}{2}\Gamma_{L_t}(b'_t, b'_t)(X_t)dt = dM_t(b'_t) - \frac{1}{2}d\langle M(b')\rangle_t. \end{aligned}$$

This implies that

$$\mathbb{E}_Y\left(f(X_t)Z_t\right) = \mathbb{E}_Y\left(f(X_t)Z'_t \exp\left(b'(X_t) + \int_0^t V_s(X_s)ds\right)\right)$$

with the random potential function V_t defined by

$$V_t = e^{b'_t} L_t(e^{-b'_t}) - b_t^2/2.$$

By classical probabilistic techniques of exponential change of measures, the martingale Z'_t induces a new reference measure, under which the signal diffusion has an infinitesimal generator given by

$$L'_t(f) = L_t(f) + e^{b'_t}\Gamma_{L_t}\left(e^{-b'_t}, f\right) = L_t(f) - \frac{\partial b'_t}{\partial x}\,\frac{\partial f}{\partial x}.$$

In other words, L'_t is the infinitesimal generator of the signal diffusion process X'_t defined as X_t by replacing the drift term a_t by the function $a'_t := a_t - \frac{\partial b'_t}{\partial x}$. By (5.1), we conclude that the optimal filter can be rewritten as

$$\pi_t(f) = \gamma_t\left(e^{b'_t} f\right)/\gamma_t\left(e^{b'_t}\right) \text{ with } \gamma_t(f) = \mathbb{E}_Y\left(f(X'_t)\exp\left(\int_0^t V_s(X'_s)ds\right)\right). \tag{5.2}$$

It is not difficult to check that the flow of unnormalized measures γ_t satisfies the Schrödinger equation

$$\frac{d}{dt}\gamma_t(f) = \gamma_t(L_t^V(f)) \quad \text{with} \quad L_t^V(f)(x) = L'_t(f)(x) + V_t(x)f(x).$$

and the continuous time version of the multiplicative formula (4.4) is given by

$$\frac{d}{dt}\log\gamma_t(1) = \eta_t(V_t) \Rightarrow \gamma_t(1) = \exp\left\{\int_0^t \eta_s(V_s)ds\right\}. \tag{5.3}$$

6. Stochastic particle approximation algorithms

The final part of the book is centered around the theoretical and numerical analysis of modern stochastic particle filtering models.

Branching and interacting particle methods are one of the most active contact points between applied probability, Bayesian statistics, theoretical chemistry, quantum physics, and engineering sciences, including rare event analysis and advanced signal processing. The origins of particle methods can be traced back in physics and molecular chemistry to the 1950s with the pioneering works of T.E. Harris and H. Kahn [9] and that of M.N. Rosenbluth and A.W. Rosenbluth [14]. Further details on the origins and the application domains can be found in the studies [5, 7, 11, 15, 16].

These stochastic particle algorithms can be thought of in various ways.

From the physical point of view, they can be seen as microscopic particle interpretations of physical nonlinear measure-valued equations. From the probabilistic point of view, they can be interpreted as interacting recycling acceptance-rejection sampling techniques. In this case, they can be seen as a sequential and interacting importance sampling technique. From the pure mathematical point of view, they can also be interpreted as natural stochastic linearizations of nonlinear evolution semigroups.

Next, we provide a synthetic and universal picture of these mathematical models. This presentation differs slightly from the one presented in the book, but it is closer to more traditional mean field particle models arising in physics and fluid mechanics, and it applies to more general situations.

6.1. **Discrete time models.** In the discrete time situation, the central idea is to rewrite the desired flow of measures η_n, such as the one introduced in (4.2), in terms of a nonlinear integral evolution equation of the form

$$\eta_n(dx') = (\eta_{n-1}K_{n,\eta_{n-1}})(dx') := \int \eta_{n-1}(dx)\; K_{n,\eta_{n-1}}(x,dx') \tag{6.1}$$

with some Markov transition $K_{n,\eta_{n-1}}$. In this case, η_n can be interpreted as the distribution of the random states $\overline{X}_n$ of the time-inhomogeneous Markov process defined as

$$\mathbb{P}\left(\overline{X}_n \in dx | \overline{X}_{n-1}\right) = K_{n,\eta_{n-1}}(\overline{X}_{n-1}, dx) \quad \text{with} \quad \eta_{n-1} = \mathrm{Law}(\overline{X}_{n-1}),$$

starting with some initial random variable with distribution $\eta_0 = \mathrm{Law}(\overline{X}_0)$. The Markov chain $\overline{X}_n$ can be thought of as a perfect sampling algorithm. Unfortunately, the sampling of its random transitions requires computation of the current distribution of the chain.

The N-mean field particle interpretation of the equation (6.1) is a Markov chain $\xi_n := (\xi_n^i)_{1\leq i\leq N}$ with elementary transitions given by

$$\mathbb{P}\left(\xi_n \in d(x^1,\ldots,x^N) | \xi_{n-1}\right) = \prod_{1\leq i\leq N} K_{n,\eta_{n-1}^N}(\xi_{n-1}^i, dx^i) \tag{6.2}$$

with the current occupation measures $\eta_{n-1}^N := \frac{1}{N}\sum_{i=1}^N \delta_{\xi_{n-1}^i}$. The initial system consists of N independent and identically distributed random variables with common law η_0.

The stochastic perturbation model associated with these discrete generation particle models is now defined in terms of a sequence of centered random fields V_n^N

defined by the equation

$$\eta_n^N = \eta_{n-1}^N K_{n,\eta_{n-1}^N} + \frac{1}{\sqrt{N}}\, V_n^N. \tag{6.3}$$

We notice that the above equation only differs from (6.1) by the Monte Carlo precision parameter $\frac{1}{\sqrt{N}}$ and the centered random fields V_n^N. Under rather weak regularity conditions, we can prove that V_n^N converge in law, as $N \to \infty$, to a sequence of independent centered Gaussian fields V_n with a variance function that can be explicitly expressed in terms of the transitions $K_{n,\eta_{n-1}}$. A detailed proof of this functional fluctuation theorem, including uniform convergence estimates, can be found in [4, 5] in the context of filtering models and in [6] for general and abstract distribution flows.

The mean field particle model associated with the flow of optimal one-step predictors η_n defined in (4.2) is a simple genetic type stochastic algorithm: The mutation and the selection transitions are dictated by the prediction and the updating transitions defined in Section 4. During the selection transition, one updates the positions of the particles in accordance with the fitness likelihood functions $g_n(., y_n)$. This mechanism is called the selection-updating transition as the adapted particles are selected for reproduction. In other words, this transition allows particles to give birth to some particles at the expense of light particles which die.

The second mechanism is called the mutation-prediction transition since at this step each particle evolves randomly according to the transition kernels M_n.

Another important feature of genetic type particle models is that their ancestral or their complete genealogical tree structure can be used to approximate the smoothing problem, including the computation of the distribution of the signal trajectories given the observations. Further details on this subject can be found in [3, 5].

Mimicking the multiplicative formula (4.4), an unbiased particle approximation of the unnormalized measures γ_n defined in (4.3) is given by

$$\gamma_n^N(f) = \eta_n^N(f) \times \prod_{0 \le p < n} \eta_p^N(g_p(., y_p)).$$

These genetic type particle algorithms can be extended to branching approximation models with random population sizes. The book provides a series of judicious branching strategies that can be used in the context of filtering, including minimal variance branching number of offsprings allocated to individual sites and the stochastic universal sampling technique introduced by Baker at the end of the 1980s.

6.2. **Continuous time models.** Continuous time filtering problems can be approximated by discrete time generation models using elementary Euler type approximation schemes. Besides this fact, the robust evolution equations discussed in section 5.3 have a natural mean particle interpretation in term of the normalized probability measures $\eta_t(f) = \gamma_t(f)/\gamma_t(1)$, with the measures γ_t defined in (5.2). After some elementary computations, we find that the flow of probability measures η_t satisfies the evolution equation

$$\frac{d}{dt}\eta_t(f) = \eta_t(L_{t,\eta_t}(f))$$

with the integro-differential interacting jump type generator L_{t,η_t} defined by

$$L_{t,\eta_t}(f)(x) = L'_t(f)(x) + V_t^-(x)\int (f(x')-f(x))\ \eta_t(dx') + \int (f(x')-f(x))\ V_t^+(x')\ \eta_t(dx')$$

for any pair of nonnegative potential functions (V_t^+, V_t^-) s.t. $V_t = V_t^+ - V_t^-$. As in the discrete time case, L_{t,η_t} can be interpreted as the infinitesimal generator of a nonhomogeneous Markov process $\overline{X}_t$ with jump type transitions that depend on the current distribution $\eta_t = \mathrm{Law}(\overline{X}_t)$. Between jumps, $\overline{X}_t$ explores the state space using a free L'_t-motion. At rate $V_t^-(\overline{X}_t)$, the random state $\overline{X}_t$ jumps to a new location randomly chosen with the current distribution η_t, and at rate $\eta_t(V_t^+)$, the random state $\overline{X}_t$ jumps to a new location randomly chosen with the updated distribution $V_t^+(x)\eta_t(dx)/\eta_t(V_t^+)$.

As in the discrete time case, the Markov process $\overline{X}_t$ can be thought as a perfect sampling algorithm, but its effective sampling again requires computation of the current distribution of the process at any time step.

The N-mean field particle interpretation of the evolution equation

$$\frac{d}{dt}\eta_t(f) = \eta_t(L_{t,\eta_t}(f))$$

is the Markov process $\xi_t = (\xi_t^i)_{1\leq i\leq N}$ with the infinitesimal generator $G_t^{(N)}$ defined on sufficiently regular functions F as

$$G_t^{(N)}(F)(x^1,\ldots,x^N) := \sum_{i=1}^N L^{(i)}_{t,\frac{1}{N}\sum_{i=1}^N \delta_{x^i}} F(x^1,\ldots,x^i,\ldots,x^N).$$

In the above formula, we have used the upper index $L^{(i)}_{t,\eta}$ to indicate that the operator $L_{t,\eta}$ acts on the ith coordinate. As in the discrete time case, the initial system consists of N independent and identically distributed random variables with common law η_0.

From the above discussion, ξ_t is a Moran genetic type interacting particle model: Every particle ξ_t^i explores the state space using a free L'_t-motion. At rate $V_t^-(\xi_t^i)$, it jumps to a new location randomly chosen with the current distribution $\eta_t^N = \frac{1}{N}\sum_{j=1}^N \delta_{\xi_t^j}$, and at rate $\eta_t^N(V_t^+)$, it jumps to a new location randomly chosen with the updated discrete measures $V_t^+(x)\eta_t^N(dx)/\eta_t^N(V_t^+)$.

Mimicking the multiplicative formula (5.3), an unbiased particle approximation of the unnormalized measures γ_t defined in (5.2) is given by

$$\gamma_t^N(f) = \eta_t^N(f) \times \exp\left\{\int_0^t \eta_s^N(V_s)ds\right\}.$$

Further details on the origins and the mathematical analysis of these interacting particle models can be found in [4, 15] and the references therein.

7. Conclusions

Stochastic filtering is a natural and universal probabilistic learning technique. It is currently used in a wide variety of scientific disciplines, including Bayesian inference, information theory, speech recognition, communication systems, signal processing, econometrics, financial mathematics, numerical physics, biology, forecasting data assimilation, and many others.

Understanding the full theoretical and numerical picture of nonlinear filtering theory demands significant mathematical and statistical sophistication far beyond that required to understand and appreciate the derivation of the optimal filter equations and its numerical approximations in a given application domain.

Nowadays, in diverse scientific communities, it seems that most of the practitioners are avoiding the mathematical foundations of filtering, directly applying some filtering techniques to a particular estimation problem. In principle there is nothing really wrong with this approach, but the danger is in having a new generation of young researchers who have no idea of the mathematical foundation nor the numerical performance, the robustness, the stability properties or even the limitation of the nonlinear filtering model they employ.

In this respect, the book of Bain and Crisan provides an excellent expository breakthrough, making the fundamentals of stochastic filtering accessible and coherent to any researcher having some background in probability theory and stochastic analysis. In an effort to make the book accessible and self-contained, two appendices contain a selected assortment of measure theory, probability theory, and stochastic analysis calculus. The writing and the edition is precise and enjoyable to read, with ample and detailed references.

The book is divided into two clearly separated parts, each written as a self-contained essay.

The first part provides a clean probabilistic treatment of the theoretical aspects of the filtering problem, with rigorous derivations of optimal filter evolution equations. By a lovely conceptual argument, the authors show that the conditional distributions of the signal is a stochastic process with values in the space of probability measures. From there, in the span of just three sections, weighing in at a not-very-dense ninety pages, the authors provide a rigorous derivation of the Zakai equation, the Kushner–Stratonovich evolution equation, as well as the robust representation formula using a partial differential equations approach or an alternative functional analytic approach. The remaining two sections of this first chapter are devoted to regularity properties of optimal filters and finite-dimensional filters, including the Kalman–Bucy filter and the Beneš filters, using the robust formulation discussed above.

The second part of the book is dedicated to numerical methods for solving the nonlinear filtering problem, with an emphasis on particle approximation models. Six classes of algorithms are presented, including linearization techniques, deterministic grid approximations, or spectral methods. The final two sections contain a detailed study of continuous and discrete generation particle algorithms, also known as sequential Monte Carlo methods in Bayesian inference literature.

I highly recommend this book to any researcher in applied mathematics, as well as to any researchers in engineering and computer sciences with some background in statistics and probability. Alan Bain and Dan Crisan masterfully guide the reader through a fascinating area of probability and Bayesian inference which has been very active in the last two decades. By the end of the book the serious reader will be well equipped with most of the theoretical foundations of filtering and with numerical methods, including advanced cutting-edge particle techniques, to study and solve a large class of nonlinear filtering problems. The book can also serve as a useful text for an informal seminar or a second year graduate course on stochastic filtering.

References

1. J. M. C. Clark and D. Crisan. *On a robust version of the integral representation formula of nonlinear filtering*. Probability theory and related fields, vol. 133, no. 1, pp. 43-56, 2005. MR2197136 (2008g:60117)
2. J. M. C. Clark. The design of robust approximations to the stochastic differential equations of nonlinear filtering. Communication systems and random process theory. Proc. 2nd NATO Advanced Study Inst., Darlington 1977. pp. 721-734, Advanced Study Inst. Ser. E: Appl. Sci., no. 25, Sijthoff & Noordhoff, Alphen aa den Rikn, 1978. MR0529130 (58:26431)
3. P. Del Moral, A. Doucet and S. S. Singh *A Backward Particle Interpretation of Feynman-Kac Formulae*, HAL-INRIA RR-7019 (07-2009); M2AN, vol 44, no. 5, 947–976 (Sept. 2010).
4. P. Del Moral and L. Miclo. *Branching and Interacting Particle Systems Approximations of Feynman-Kac Formulae with Applications to Non-Linear Filtering*. Séminaire de Probabilités XXXIV, Ed. J. Azéma and M. Emery and M. Ledoux and M. Yor, Lecture Notes in Mathematics, Springer-Verlag Berlin, Vol. 1729, 1-145. 2000. MR1768060 (2001g:60091)
5. P. Del Moral. *Feynman-Kac formulae. Genealogical and interacting particle approximations*, Springer New York, Series: Probability and Applications, 2004. MR2044973 (2005f:60003)
6. P. Del Moral and E. Rio, "Concentration Inequalities for Mean Field Particle Models", HAL-INRIA RR-6901, 2009; to appear in Annals of Applied Probability (2010–2011).
7. A. Doucet, N. De Freitas and N. Gordon, editors. *Sequential Monte Carlo Methods in Pratice. Statistics for engineering and Information Science*. Springer, New York, 2001. MR1847783 (2003h:65007)
8. S. N. Ethier and T. G. Kurtz. *Markov processes: characterization and convergence*, Wiley Series Probability & Statistics, 1986. MR838085 (88a:60130)
9. T. E. Harris and H. Kahn. *Estimation of particle transmission by random sampling*. Natl. Bur. Stand. Appl. Math. Ser., 12:27-30, 1951.
10. I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Graduate Texts in Mathematics, Springer, 2004. MR1121940 (92h:60127)
11. S. Méléard, "Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models", Lecture Notes in Mathematics, Springer, Berlin Heidelberg, Vol. 1627, pp. 44-95, Probabilistic Models for Nonlinear Partial Differential Equations, 1996. MR1431299 (98f:60211)
12. Ph. Protter. *Stochastic integration and differential equations*, volume 21 of Stochastic Modelling and Applied Probability. Springer-Verlag, Berlin, 2005. MR2273672 (2008e:60001)
13. D. Revuz and Marc Yor, *Continuous martingales and Brownian motion*, edited by Springer-Verlag, New York, 1991. MR1083357 (92d:60053)
14. M. N. Rosenbluth and A. W. Rosenbluth. *Monte Carlo calculations of the average extension of macromolecular chains*. J. Chem. Phys., 23:356-359, 1955.
15. M. Rousset, "On the control of an interacting particle approximation of Schrödinger ground states", SIAM J. Math. Anal., 38 (3), 824-844, 2006. MR2262944 (2007m:60300)
16. A. S. Sznitman, *Topics in propagation of chaos*, Lecture Notes in Math. 1464, Springer, Berlin, 164-251, 1991. MR1108185 (93b:60179)

Pierre Del Moral

INRIA Bordeaux-Sud Ouest Center; The Mathematical Institute of Bordeaux

E-mail address: pierre.del-moral@inria.fr