Published in final edited form as:

IEEE Trans Automat Contr. 2015 August ; 60(8): 2161–2165. doi:10.1109/TAC.2015.2409910.

A Nonlinear Stochastic Filter for Continuous-Time State Estimation

Atiyeh Ghoreyshi and

Masimo Corp, Irvine California

Terence D. Sanger

Departments of Biomedical Engineering, Neurology, and Biokinesiology, University of Southern California, Los Angeles, CA 90089

Terence D. Sanger: tsanger@usc.edu

Abstract

Nonlinear filters produce a nonparametric estimate of the probability density of state at each point in time. Currently-known nonlinear filters include Particle Filters and the Kushner equation (and its un-normalized version: the Zakai equation). However, these filters have limited measurement models: Particle Filters require measurement at discrete times, and the Kushner and Zakai equations only apply when the measurement can be represented as a function of the state. We present a new nonlinear filter for continuous-time measurements with a much more general stochastic measurement model. It integrates to Bayes' rule over short time intervals and provides Bayes-optimal estimates from quantized, intermittent, or ambiguous sensor measurements. The filter has a close link to Information Theory, and we show that the rate of change of entropy of the density estimate is equal to the mutual information between the measurement and the state and thus the maximum achievable. This is a fundamentally new class of filter that is widely applicable to nonlinear estimation for continuous-time control.

Index Terms

Nonlinear filter; particle filter; Kushner Equation; Zakai Equation; Fokker-Planck Equation

I. Introduction

Filtering techniques estimate the hidden states of a dynamical system from noisy measurements. They do so using a model of the underlying system dynamics and a measurement model that links observed variables to the hidden states. When sensor measurements used for feedback control are noisy, quantized, intermittent, ambiguous, or otherwise bandlimited, then filtering algorithms must be used by the observer process to extract the best possible estimate of the current state. The most commonly used algorithms are based on linear system and observation models and samples taken either in discrete time (e.g., Kalman filter [1], [2]) or continuous time (e.g., Kalman-Bucy filter [3]–[5]). The Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF) and their variations are approximate solutions that can be applied to nonlinear systems [6]–[8, for example], but these filters are not optimal and they may have poor performance when the noise distribution is non-Gaussian [6]. Although, unlike EKF, UKF does not require

differentiability of the dynamics and measurement models, it estimates only the first and second-order statistics of the state density. These are only sufficient statistics for processes that remain Gaussian distributed at all times (which is almost never the case when the dynamics are nonlinear). Thus for all Kalman filter variants there is a strong assumption that the probability distribution of the state remains unimodal (usually Gaussian). More complex distributions (such as those with hard boundaries, multiple possible values, or discrete values) cannot be represented. Recent extensions such as the use of Gaussian mixture models for the density estimate [9]–[11] permit more flexibility but still have significant limitations on the form of the density and the measurement models.

Improvements in processing speed have permitted a class of filtering algorithms to be developed that apply to nonlinear dynamic systems, nonparametric uncertainty models, and nonlinear observation models. This class is commonly referred to as “nonlinear filters”, although it would perhaps be more correct to describe it as “nonparametric filters” or “Bayesian filters” because the distinguishing feature is estimation of the nonparametric representation of the full density $p(x, t)$ at each time t , rather than just the first and second-order statistics.

Particle filters [12] are perhaps the most well known implementation of nonlinear filters, but they only apply when observations are taken at discrete time intervals. For continuously observed processes, the only currently known nonlinear (nonparametric) filters are the Kushner [13] and Zakai [14] equations which assume (similarly to EKF and UKF) that observations are given by

$$dz = h(x)dt + dB \quad (1)$$

where B is Brownian noise, and $h(x)$ is a deterministic and possibly nonlinear, nondifferentiable, or noninvertible function. This makes the technical assumption that the state is related to the derivative of the observation process; common observation models such as $z = x$ cannot be represented.

All existing filters for continuous-time measurement, including variants on the Kalman filter, Gaussian-mixture models, and the Kushner and Zakai equations require observations whose means are determined by algebraic functions of the state, and all other statistics are independent of state. Thus there are no filters that apply when observations are a Poisson process with rate x , a Gaussian process with variance x , or other stochastic models in which the uncertainty of the observation depends on the state. We here propose a new class of continuous-time nonlinear filter with a much more general observation model that requires only specification of the conditional density of the observation given the state $p(z|x)$. We show that the new filter achieves the upper bound on information extraction from the data, so that the rate of change of entropy of the state density estimate is equal to the mutual information between the observation and the state.

II. Existing Nonlinear Filters

Consider a process with the following stochastic differential equation for the state:

$$dx = f(x)dt + g(x)dW \quad (2)$$

where W is Brownian noise, x is state, and $f(x)$ and $g(x)$ are nonlinear differentiable functions that model the deterministic dynamics of state and the state-dependent noise, respectively. For this class of smoothly-evolving Markovian system dynamics, the probability density of state evolves approximately according to the linear partial differential Fokker-Planck equation:

$$\frac{\partial p}{\partial t} = -\frac{\partial(fp)}{\partial x} + \frac{1}{2} \frac{\partial^2(g^2 p)}{\partial x^2} \quad (3)$$

where $p = p(x, t)$ is the pdf of state. This can be written as $\dot{p} = Fp$, where F is the linear Fokker-Planck operator on p , and $\dot{p}$ indicates the partial differential with respect to t .

The Extended Kalman Filter (EKF) (with observation model $z = h(x)$) requires calculation of the Jacobians f'/x and h'/x , and it therefore assumes that f and h are both differentiable. Both the EKF and the Unscented Kalman Filter (UKF) assume that $p(x, t)$ can be approximated by a Gaussian, so that first and second-order statistics are sufficient. But note that for almost all nonlinear functions $f(x)$, even when $p(x, t_0)$ is Gaussian, $p(x, t)$ is unlikely to be Gaussian when $t > t_0$. Common examples include non-uniform distributions of flow, container or obstacle boundaries, or repulsive forces.

The Zakai filter (with observation model (1)) is given by the stochastic partial differential equation:

$$dp = Fpdt + h^T R^{-1} dz p \quad (4)$$

and the Kushner filter is:

$$dp = Fpdt + (h - \bar{h}_t)^T R^{-1} (dz - \bar{h}_t dt) p \quad (5)$$

where $\bar{h}_t$ is the expected value over x of $h(x)$ at time t , dz is the differential of the observation process (1), R is the autocorrelation of the observation noise, and h is to be interpreted as the function $h(\cdot)$. Unlike EKF, h does not have to be differentiable. But note that the observation process z is the integral of $h(x)$ and thus often unbounded.

Although the Zakai equation is mathematically equivalent to the Kushner equation up to normalization, in practice it seems to diverge from solutions of the Kushner equation due to numerical sensitivity [15].

III. Method

We propose the following general filter equations for unknown state $x(t)$ and observations $z(t)$ with a known memoryless observation model $p(z(t)|x(t))$:

$$\frac{\partial p(x, t)}{\partial t} = Fp(x, t) + [\log p(z|x) - \log p(z, t)] p(x, t) \quad (6)$$

where z is the current observation, $p(z|x)$ is the observation model, $p(z, t) = \int p(z|x)p(x, t)dx$ is the scalar probability of observing z given the current pdf estimate, and F is the Fokker-Planck operator:

$$Fp = -\frac{\partial(fp)}{\partial x} + \frac{1}{2} \frac{\partial^2(g^2 p)}{\partial x^2} \quad (7)$$

Equation 6 can also be written

$$\frac{\partial p(x, t)}{\partial t} = Fp(x, t) + \log \left(\frac{p(z, x)}{p(z)p(x)} \right) p(x, t) \quad (8)$$

which emphasizes that observations z update the state x only when z and x are not independent.

The observation model requires only the specification of the conditional density $p(z|x)$. This describes memoryless observations that can be deterministic, nondeterministic, nonlinear, or noninvertible. The initial condition $p(x, 0)$ represents the prior information available before making any observations, and it is often taken as uniform (for bounded domains), Gaussian (for unbounded domains), or a delta-function (when initial conditions are precisely known).

We can show that equation 6 is the continuous-time limit of the discrete Bayesian update equation:

$$p(x, t + \Delta t) = \frac{p(z(t + \Delta t)|x)}{p(z(t + \Delta t))} Mp(x, t) \quad (9)$$

$$= B_z Mp(x, t) \quad (10)$$

where M is the Markov update operator that is the short-time solution to the Fokker-Planck equation over interval t , and B_z is a diagonal operator that implements the Bayesian update of x at time $t + t$. B is defined so that for any prior density $p(x)$, $B_z p(x) = p(z|x)p(x)/p(z)$. Thus $Mp(x, t)$ is the estimate of $p(x, t + t)$ immediately prior to making a (discrete) observation $z(t + t)$, and $B_z Mp(x, t)$ is the posterior estimate of $p(x, t + t)$ immediately after observing $z(t + t)$.

To see this, note that iteration of equation 10 gives $p(x, T) = (\prod B_{z(t_i)} M)p(x, 0)$ and if we assume a countable basis for the space of realizable probability densities, then we can write this as:

$$p(x, T) = e^{\sum (\log B_{z(t_i)} + \log M)} p(x, 0) \quad (11)$$

where $\log B$ is defined by the log of its (diagonal) elements, and $\log M$ is defined such that $\exp(\log M)p = Mp$ for all realizable p . Since M is the solution of the Fokker-Planck equation $\dot{p} = Fp$ which for short intervals has quasi-static solution $\exp(F t)$, we see that we can identify $\log M$ with F . Letting $t \rightarrow 0$, we write:

$$p(x, T) = e^{\int (\log B_{z(t)} + F) dt} p(x, 0) \quad (12)$$

which is the solution of the differential equation

$$\dot{p}(x, t) = (\log B_{z(t)} + F) p(x, t) \quad (13)$$

This is equivalent to equation 6.

IV. Observation Models

Since $p(z, t)$ is a scalar, its effect is to normalize the probability density, and if this term is eliminated the effect is to produce a non-normalized estimate of the density:

$$\frac{\partial p(x, t)}{\partial t} = Fp + \log p(z|x) p(x, t) \quad (14)$$

For exponential-family observation models with unknown dynamics, this leads to a particularly simple form of the estimation equation. For example, unit variance additive Gaussian noise leads to:

$$\frac{\partial p(x, t)}{\partial t} = Fp - \frac{1}{2}(z-x)^2 p(x, t) \quad (15)$$

and unit width Laplace-distributed noise leads to

$$\frac{\partial p(x, t)}{\partial t} = Fp - |z-x| p(x, t) \quad (16)$$

For an observation model that is a non-homogeneous Poisson process with rate $x(t)$, the observations are a series of event times t_i , and the non-normalized estimation equation is given by

$$\frac{\partial p(x, t)}{\partial t} = Fp + \left(-x + \log(x) \sum_i \delta(t-t_i) \right) p(x, t) \quad (17)$$

where $\delta(t-t_i)$ is a Dirac delta-function centered at the time of an event. (This can be implemented by having the state evolve according to $\dot{p} = Fp - xp$ between events, and assigning $p \leftarrow xp$ at the time of each event.) Integration of this equation between t_i and t_{i+1} for constant x (and leaving out the term Fp for simplicity) yields

$$p(x, t_{i+1}) = x e^{-x(t_{i+1}-t_i)} p(x, t_i) \quad (18)$$

which is the standard equation for the probability of rate x given an inter-event interval of $t_{i+1} - t_i$.

V. Information Transmission

To quantify the effect of measurement by this algorithm, consider the case when x is varying slowly so that dx/dt is small and the intrinsic dynamics Fp are small. If $p(x, t)$ is differentiable with respect to both its arguments, then the average change in entropy of the estimated density of x due to equation 6 is

$$\frac{d}{dt} E_x E_z [\log p(x)] = E_x E_z \left[\frac{1}{p(x)} \frac{\partial p(x, t)}{\partial t} \right] \quad (19)$$

$$= E_x E_z [\log p(z|x) - \log p(z, t)] \quad (20)$$

$$= I(Z, X) \quad (21)$$

where E_x and E_z are the expectation operators with respect to $p(x, t)$ and $p(z, t)$ respectively, and $I(Z, X)$ is the mutual information between the observation and the state. From this we see that the expected change in entropy per unit time due to the measurement is given by the mutual information between the observation and the state, which is the maximum that could be achieved by any process.

VI. Results

Figure 1 shows a comparison between Kalman-Bucy, the Kushner equation, and our method for state estimation. In this example, the true state dynamics are one-dimensional smoothed white noise, and observations are continuously available quantized measurements of state. For the Kalman-Bucy filter, quantization error must be approximated as Gaussian noise with a standard deviation equal to half of the bin width. For the Kushner equation, the observed variable z is required to be the integral of the quantized state and therefore its variance grows without bound, so we work only with its differential dz . Note that quantization causes uncertainty even with a deterministic (noise-free) observation process. This cannot be described accurately by linear filters such as Kalman-Bucy, and the extended Kalman filter fails because the observation model is not differentiable and thus cannot be approximated as a locally linear process.

The correct behavior (figure 1a) is interesting, because at the moment of a transition between two different quantization levels we have precise information as to the underlying state (it must lie exactly on the transition threshold). Uncertainty in the state estimate grows between the time of transitions, but the set of possible values is bounded by the nearest transition thresholds. Therefore, between jumps of the observed variable, the state should gradually approach a uniform distribution bounded by the two nearest transition thresholds. Both the Kushner equation and our method show the expected behavior at the time of

transitions, but the Kushner equation incorrectly predicts that there is nonzero probability of crossing a bit-transition boundary in the absence of a change in the quantized observation.

To compare performance, we created 100 different 0.5sec time-series of filtered white noise (500 time points at 1kHz sample rate). The mean-squared error in state estimates (averaged over the full 0.5sec) for the new method and the Kushner equation are not statistically different (new method: 0.345 (SD 0.06), Kushner equation: 0.341 (SD 0.06), $p > 0.5$, t-test), and both are significantly better than Kalman-Bucy (0.922 (SD 0.40), $p < 0.0001$). The average width (standard deviation) of the estimated conditional density $p(x, t)$ was significantly less for the new method than for the Kushner equation (0.548 vs. 1.768 for Kushner, $p < 0.0001$), so the Kushner equation overestimated the uncertainty (probably due to sensitivity to time discretization in the simulation; in the theoretical continuous-time case the Kushner equation is expected to produce the correct variance).

Figure 2 shows an example where the measurement model is the absolute value function $h(x) = \text{abs}(x)$. For this type of ambiguous observation, it is not possible to distinguish positive from negative values of the state, so the correct estimate of $p(x, t)$ is bilobed. Methods based on estimation of second-order statistics, including the Kalman-Bucy filter, EKF, or UKF are unable to even represent such densities, let alone estimate them. One of the advantages of nonlinear filters is that state estimates can still be made in the context of observation ambiguity, and the correct value of the state can be disambiguated at a later time (or using additional sensor data) by Bayesian combination with other measurements.

Figure 2 provides an example in which our method is superior to all other existing methods. As noted above, parametric methods such as Kalman-Bucy, EKF, or UKF cannot even represent this class of problem. The Kushner (and Zakai) equation can represent this type of problem, but because of the overestimation of the variance, Kushner incorrectly estimates a unimodal density (right side of the 3D figure) where the actual density is bimodal. Thus even in situations in which the new method and the Kushner equation produce similar estimates of the mean, more accurate estimation of the probability density is a significant superiority to our method. At lower sample rates, the Kushner equation will perform even worse, since its mechanism of estimation depends on near-continuous sampling. Accurate density estimation is critically important for estimation of confidence, Bayesian combination of estimates from multiple sensors, and for estimation of risk of error.

VII. Complexity

For the above one-dimensional problems, both nonlinear filters required considerably more time than the Kalman-Bucy filter on a single-processor computer (57 times longer for the new filter, 54 times longer for the Kushner equation). If the state has multiple dimensions $x_1 \cdots x_N$, then we can estimate the marginal densities $p(x_i, t)$ using N copies of the algorithm with measurement models $p(z|x_i)$. In this case, the complexity scales linearly with N . However, to estimate the joint density $p(x, t) = p(x_1 \cdots x_N, t)$ we require use of the full conditional measurement model $p(z|x_1 \cdots x_N)$ and storage of the full joint density, and the complexity will scale exponentially with N . An approximation with linear scaling estimates the marginal densities using $p(z|x_1 \hat{x}_1 \cdots x_{i-1} \hat{x}_{i-1}, x_i, x_{i+1} \hat{x}_{i+1} \cdots x_N \hat{x}_N)$ where the $x_j \hat{x}_j$'s are the estimated

values from the other marginal filter equations (similar to Gibbs sampling). These scaling issues are common to all nonlinear (nonparametric) filters due to the need to store and update the full density estimate.

VIII. Applications

Our goal here is to describe the state estimator. When the resulting density is used for control, a single estimate of state must be extracted. The minimum mean-squared error estimate will be the mean of the probability density at each point in time, and this can be used in a standard feedback control loop. Other estimates such as the maximum likelihood value could be used and might be more appropriate for non-quadratic cost functions.

A. Sensor Fusion

Suppose we have two conditionally-independent sensors $z_1(t)$ and $z_2(t)$ (by conditional independence we mean that $p(z_1, z_2|x) = p(z_1|x)p(z_2|x)$). We can then update the Bayes-optimal combined probability density using:

$$\frac{\partial p(x, t)}{\partial t} = Fp(x, t) \quad (22)$$

$$+ [\log p(z_1|x) - \log p(z_1, t)] p(x, t) \quad (23)$$

$$+ [\log p(z_2|x) - \log p(z_2, t)] p(x, t) \quad (24)$$

(If the sensors are not conditionally independent, then the last line should be replaced by $[\log p(z_2|z_1, x) - \log p(z_2|z_1, t)] p(x, t)$.) This process can be generalized to an arbitrary number of sensors. Furthermore, if we create two independent estimates of the state density so that $p_1(x, T)$ is based on data from $z_1(t)$ up to time T , and $p_2(x, T)$ is based on data from $z_2(t)$ up to time T , then we can obtain the un-normalized density at time T from the combined sensor estimates as the product:

$$p(x, T) = p_1(x, T)p_2(x, T) \quad (25)$$

Combination of estimates from multiple sensors provides a considerable advantage over standard second-order estimators, for which a combination of the mean estimates from different sensors (weighted by the inverse of the variances) is accurate only for unimodal Gaussian distributions. In particular, note that sensors can be combined even if they have different accuracy, different sampling intervals, time-varying noise, or ambiguity leading to multi-lobed density estimates.

B. Delayed Measurements

Suppose that we have a delayed measurement, so that sensor data is only available for $z(t - \tau)$. There are two ways to manage such delayed measurements within this framework. The most straightforward is to estimate the delayed density $p(x, t - \tau)$ and then to propagate

forward for time open-loop using the Fokker-Planck equation $\dot{p} = Fp$ (equation 7) with initial condition $p(x, t - \tau)$. This provides an estimate of the current density, but requires forward propagation of the Fokker-Planck equation at every time, and it is difficult to use if there are multiple sensors with different time delays.

The second method is mathematically equivalent, but it depends upon calculation of the observation model $p(z(t - \tau)|x(t))$. This represents the information about the current value of state that is provided by the delayed observation (or, equivalently, the probability of the delayed observation given the current state). It will necessarily be less certain because time has passed since the observation. For example, if $z(t) = x(t)$, then the observation model is $p(x(t - \tau)|x(t))$ which will spread the density (according to the backwards-time Fokker-Planck equation). This delayed observation model can be directly used in equations 6 or 24, and it allows straightforward combination of information from sensors with different delays.

We will not here examine the stability properties of closed-loop systems with delay based on this type of measurement model, but we do point out that such systems may be stable under circumstances that would lead to instability in standard closed-loop control. This is because the delayed measurement itself is not used for feedback, but instead it contributes to an estimate of the density of future state that may include values that are not out of phase with the true state.

IX. Conclusion

We have introduced a new continuous-time nonlinear filtering algorithm for state estimation with a general observation model. As with other nonlinear filters, the statistical models of the underlying process and time-varying uncertainty are nonparametric and therefore allow non-Gaussian behavior including skewed or multi-lobed densities of state. Our innovation is the incorporation of a new continuous measurement model that allows for implicit and non-algebraic dependence of the observed variable on the state; the only knowledge needed about the observed variable is the likelihood function $p(\text{observation}|\text{state})$. This allows for sensor measurements to be in a completely different space from the state, such as discrete or categorical measurements of continuous variables, quantized and non-monotonic measurements of state, or other flexible situations to which previous methods cannot be applied. For instance, our method applies when the observation is a Poisson process with rate determined by the state, when there is independent noise on individual bits of a quantized measurement, or when the noise amplitude is state-dependent. Neither parametric methods nor the Kushner/Zakai equation are applicable in such situations. In situations when the Kushner equation does apply, our method produces equivalent estimates of the mean, but the estimated density (and variance) is significantly improved. Our method has comparable computational complexity and performance to other nonlinear filters, but it is applicable to a much more general set of stochastic observation models. When the assumptions of standard filters are satisfied, the new filter is expected to produce identical results. The primary significance of the new filter is that it permits the use of Bayes-optimal continuous-time nonlinear filtering in a wide variety of state observers for which there was previously no mathematically valid algorithm.

Acknowledgments

Support for this project was provided by the National Institutes of Neurological Disorders and Stroke (NS069214), and the James McDonnell Foundation.

References

1. Kalman R. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*. 1960; 82:35–45. no. Series D.
2. Welch G, Bishop G. An introduction to the kalman filter. *Design*. 2001; 7(1):1–16.
3. Kalman R, Bucy R. New results in linear prediction and filtering theory. *Trans AMSE J Basic Eng*. 1961:95–108.
4. Bucy, R.; Joseph, P. *Filtering for stochastic processes with applications to guidance*. American Mathematical Society; 1987.
5. Bucy R. Linear and nonlinear filtering. *Proceedings of the IEEE*. 1970; 58(6):854–864.
6. Julier S, Uhlmann J. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*. 2004; 92(3):401–422.
7. Julier S, Uhlmann J, Durrant-Whyte H. A new method for the nonlinear transformation of means and covariances in filters and estimators. *Automatic Control, IEEE Transactions on*. 2000; 45(3): 477–482.
8. Einicke G, White L. Robust extended kalman filtering. *Signal Processing, IEEE Transactions on*. 1999; 47(9):2596–2599.
9. Ito K, Xiong K. Gaussian filters for nonlinear filtering problems. *Automatic Control, IEEE Transactions on*. 2000; 45(5):910–927.
10. Terejanu G, Singla P, Singh T, Scott P. Adaptive gaussian sum filter for nonlinear bayesian estimation. *Automatic Control, IEEE Transactions on*. 2011; 56(9):2151–2156.
11. Alspach D, Sorenson H. Nonlinear bayesian estimation using gaussian sum approximations. *Automatic Control, IEEE Transactions on*. 1972; 17(4):439–448.
12. Gordon N, Salmond D, Smith A. Novel approach to nonlinear/non-gaussian bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*. 1993; 140(2):107–113.
13. Kushner H. Dynamical equations for optimal nonlinear filtering. *Journal of Differential Equations*. 1967; 3:179–190.
14. Zakai M. On the optimal filtering of diffusion processes. *Probability Theory and Related Fields*. 1969; 11(3):230–243.
15. Ito K, Rozovskii B. Approximation of the kushner equation for nonlinear filtering. *SIAM Journal on Control and Optimization*. 2000; 38:893.

Fig. 1.

Comparison of filtering methods for a quantized observation process. State dynamics are governed by smoothed white noise. The measurement model for Kushner's equation is $dz = h(x)dt + gdB$, and for the Kalman filter and the new method is: $z = h(x)dt + gn$ where $h(x)$ is a 10-level quantizer and $g = 1/2$. B is standard Brownian noise, and n is standard white noise. Vertical axis is state x , horizontal axis is time t , and color indicates $p(x, t)$. Black line is the true state x , pink line is the quantized observation, and the red line is the maximum likelihood estimate returned by each filter. Both the new method and the Kushner equation correctly estimate the state, but the Kushner method overestimates the state uncertainty.

(a) New method

(b) Kushner equation

Fig. 2.

Comparison of (a) our method and (b) Kushner equation when the observation is the absolute value of the state $h(x) = \text{abs}(x)$. State dynamics are governed by smoothed white noise. Height of the surface indicates $p(x, t)$. The true value of the state is ambiguous, so both filters correctly estimate a bilobed distribution, although the variance of the new method is smaller and it therefore correctly separates the bilobed distribution near zero. Note that linear methods such as Kalman-Bucy cannot be applied to this observation model.