Practical Application of a Branching Particle-based Nonlinear Filter

David J. Ballantyne[a], John R. Hoffman[b], and Michael A. Kouritzin[a]

[a]MITACS-PINTS,
Department of Mathematical Sciences, University of Alberta,
Edmonton, Canada

[b]Lockheed Martin Tactical Systems - Eagan
PO Box 64525; MS U2S25 St. Paul, MN 55164-0525

ABSTRACT

Particle-based nonlinear filters provide a mathematically optimal (in the limit) and sound method for solving a number of difficult filtering problems. However, there are a number of practical difficulties that can occur when applying particle-based filtering techniques to real world problems. These problems include

- highly directed signal dynamics
- highly definitive observations
- "clipped" observation data.

Current approaches to solving these problems generally require increasing the number of particles, but to obtain a given level of performance the number of particles required may be extremely large.

We propose a number of techniques to ameliorate these difficulties. We adopt the ideas of simulated annealing and add noise which is damped in time to the particle states when they are evolved or duplicated, and also add noise which is damped in time to the interpretation of the observations by the filter, to deal with signal dynamics and observation problems. We modify the method by which particles are duplicated to deal with different information flows into the system depending on the location of the particle and the information flow into the particle. We discuss the success we have had with these solutions on some of the problems of interest to Lockheed Martin and the MITACS-PINTS research center.

Keywords: tracking, image processing, nonlinear filtering, particle methods, branching particle system

1. INTRODUCTION

We consider a single target tracking problem modeled by the Itô equation

$$dX_t = A(X_t)dt + B(X_t)dW_t, \tag{1}$$

where X_t is the unobserved signal to track, and

$$Y_k = h_k(X_{t_k}, V_k), \tag{2}$$

where Y_k is a sequence of observations of the signal that are corrupted by an independent noise sequence given by $V_k, k = 1, 2, \ldots$. Tracking filters are useful in a variety of problem areas, such as surveillance, aeronautics, and

Further author information: (Send correspondence to D.J.B.)
D.J.B.: E-mail: dballant@math.ualberta.ca
J.R.H.: E-mail: john.r.hoffman@lmco.com
M.A.K.: E-mail: mkouritz@math.ualberta.ca
PINTS web page: http://www.math.ualberta.ca/pints

search-and-rescue. In the case in which A and h_k are linear, B is constant, and the V_k are Gaussian, the conditional distribution can be efficiently computed by the Kalman filter. For our case, it is a assumed that A, B, and h_k are nonlinear and that there are no exact finite or infinite-dimensional filters applicable to the problem.

Exact filtering loosely refers to those filtering problems that degenerate into the evolution of finite-dimensional sufficient statistics or into FFT-based convolution with a known kernel. That is, a difficult, infinite dimensional equation degenerates into a known, readily implementable computer algorithm without need for approximation. For nonlinear problems as in Equations 1 - 2 where there is no exact filter but there is still a density for the conditional distribution, the theoretical solution requires the use of Fokker-Plank density evolution combined with Bayes' rule,[1] which is computationally intractable when the dimension of the signal state space is large, and is often difficult even in small dimension spaces. In response, methods such as the extended Kalman filter and interacting multiple models[2] have been developed.

1.1. Branching Particle-based Filter

One type of solution to nonlinear filtering problems is the class of particle-based methods.[3–5] A particle filter approximates the conditional distribution of the signal, given the observations, by a finite sum of Dirac measures. Each particle X^j represents a Dirac measure in the space of the signal. For each new observation, all particles are evolved forward to account for the stochastic dynamics of the signal and then the set of particles is adjusted to account for the information from the observation. In this manner, the particles can function as an adaptive Monte-Carlo method for the filtering problem.

The set of particles then approximates the full data of the distribution of the signal conditioned on the set of all back observations. The approximated conditional probability that the signal lies within a given area is computed by dividing the number of particles in that area by the total number of particles.

Particle-based filters require an appropriate algorithm for the adjustment phase such that the filters provably converge to the conditional distribution as the number of particles approaches infinity. One such adjustment strategy is that of the branching particle-based filter.[6,7] In this type of filter, particles are branched (duplicated or removed) to form child particles at each observation, with each child particle having the same state in the problem domain as its parent. The number of child particles generated (zero, one, or two in the original formulation) is determined by an equation that incorporates the likelihood of the state of each particle given the current observation.

The method is initialized with N particles $\{X_0^j\}_{j=1}^N$ either randomly and uniformly distributed in the domain of X, or randomly sampled according to some *a priori* distribution. At each observation, the method progresses through the following stages: evolution of the particles, particle branching, and the approximation of the conditional distribution of the signal state.

1.1.1. Evolution

In the evolution stage, each of the particles is evolved independently for the time period between observations (i.e. $t_k - t_{k-1}$) according to the Itô equation of the signal, $X^j_{t_{k-1}} \to X^j_{t_k}$, as described in Equation (1).

1.1.2. Particle Adjustment

After evolving each particle, the particles are then branched to account for the information from the observation. A value labeled $\xi_k^j = \xi(X^j_{t_k})$ is calculated for a given particle $X^j_{t_k}$ in response to the observation at time t_k. The value of ξ is a function that depends on several different parameters, although for a fixed time t_k these parameters are fixed for all particles. ξ depends upon each particle, upon the observation Y_k at time t_k, it depends upon the distribution of the noise of the observation (i.e. V_k as defined in equation (2)), and it depends on the relationship between the the observation and the noise h_k from equation (2). The formula is constructed specifically so that the branching method will provably converge to the optimal filter as the number of particles is increased.

Once ξ is calculated, a uniform-(0,1) random variable U_k^j is generated for each particle and

- if $(\xi_k^j \geq U_k^j)$, a new particle $X^*_{t_k} = X^j_{t_k}$ is added,
- if $(\xi_k^j \leq -U_k^j)$, $X^j_{t_k}$ is removed,
- in the most frequent case where $|\xi_k^j| < U_k^j$, the particle is not branched and and is left in the state that it evolved to.

Intuitively we can interpret ξ_k^j as a measure of the "goodness of fit" of $X_{t_k}^j$ to the hypothesis, $X_{t_k}^j$ matches the observation Y_k. If this measure is near 1 we duplicate $X_{t_k}^j$ because it is a "good fit" to Y_k. If it is near -1 it is a poor fit, and we delete the particle. Otherwise, when X_{t_k} is near 0, it is unclear as to the "goodness" of the fit, and we let the particle evolve, hoping for more definitive information in the future.

Once this branching step is complete, a method is then used to control the number of particles to some number N_k, usually the original value N. This method maintains the convergence to optimality of the filter. Furthermore, the particles are renumbered from $1, \ldots, N_k$. (Because some of the particles have been deleted and others have been duplicated, the indices are no longer in sequence.)

1.1.3. Estimation

As discussed in the forthcoming paper by Kouritzin,[6] by conditional independence and the law of large numbers

$$\frac{1}{N_k}\sum_{j=1}^{N_k}\delta_{X_{t_k}^j}(A) \stackrel{N\to\infty}{\longrightarrow} \int_A \rho^k(x)dx, \tag{3}$$

for sets A within the signal domain, where $\rho^k(x)$ is the conditional probability density of the state of the signal at time t_k given observations $Y_1, \ldots, Y_k$.

1.1.4. Efficiency

The calculations performed by branching filter implementations can be simplified so that the computational complexity is directly proportional to the number of particles and is also directly proportional to the magnitude of the portion of the observation vectors that each particle references. For example, if the observations are rasters of pixels, the algorithm will execute at a speed proportional to the number of pixels that a given particle interacts with in the function that forms the observations. This might be the number of pixels that intersect with a figure associated with the state of the particle. The example problem that we introduce provides a model that illustrates such an association of particles to pixels.

As Equation 3 indicates, particle-based filters such as the branching filter converge to the optimal filter as the number of particles increases. However, practical implementations need to operate effectively with some finite amount of resources. In practical terms, the filter should perform adequately with some finite number of particles and under the assumption that either the observations do not have an excessive extent, or that each particle need only reference some small, fixed portion of the observations in the calculation of ξ.

1.2. Example Problem

We describe an example problem to motivate the discussion of practical difficulties in applying the branching filter. Suppose we are interested in observing a ship at sea from a helicopter using a digitized camera. The helicopter obtains a sequence of images of the ocean surface that are corrupted and distorted by spatial noise and sensor truncation effects. This noise is large enough that the position of the ship cannot be accurately estimated from a single image. However, knowledge of the stochastic law of the ship along with a sequence of observations over time enables filtering and tracking of the state of the target. While the difficulties discussed in the following sections are more universal than just this problem model, this model will be used as an illustration of practical difficulties that can be encountered.

1.2.1. Signal Description

The stochastic behavior of our ship is described by a nonlinear system whose five state components evolve with time according to a fixed law. Variables x_t, y_t, and θ_t indicate the current position and orientation of the ship on the ocean surface within some frame of reference. The ship moves forward with a randomly shifting speed s_t, and turns with a rate of change of orientation $\dot{\theta}_t$ that also shifts randomly. A stochastic differential equation to describe such motion is:

$$d\begin{bmatrix} s_t \\ \dot{\theta}_t \end{bmatrix} = \begin{bmatrix} \frac{1}{2}(\frac{a+b}{2} - s_t) \\ -\alpha\dot{\theta}_t \end{bmatrix} dt + \begin{bmatrix} \sqrt{(b-s_t)(s_t-a)} & 0 \\ 0 & c \end{bmatrix} dB_t, \tag{4}$$

where a and b, $0 < a < b$, are the minimum and maximum ship speeds, $c > 0$ models the extent to which the ship is likely to change course, and $\alpha > 0$ forces the rate of orientation change $\dot{\theta}_t$ back towards zero, thus tending to eventually level out turns to an extent proportional to α. Here, B is a standard Brownian motion in $\mathcal{R}^2$, and strong existence and uniqueness hold for Equation 4.

In this example, there is an explicit solution to Equation 4, given by

$$s_t^{s_0} = a + \frac{b-a}{2}[1 - cos(B_t^1 + cos^{-1}(\frac{\frac{a+b}{2} - s_0}{\frac{b-a}{2}}))] \tag{5}$$

and

$$\dot{\theta}_t^{\dot{\theta}_0} = \int_0^t ce^{-\alpha(t-s)} dB_s^2, \tag{6}$$

so that the equations of motion for the ship can be simulated exactly and no approximations are required in the evolution of the particle states, significantly increasing the speed of the filter calculations. While it is often possible to incorporate such serendipitous circumstances into the branching filter, the existence of such a solution is not required for this type of filter, or for any method that we describe in this paper. Note that as a tends to b and as c tends to zero, we approach a deterministic model with speed s_t constant at $a = b$ and, assuming that $\theta_0 = 0$, $\dot{\theta}_t$ and θ_t constant at zero, indicating a straight course.

We assume that the ship is initially positioned at (x_0, y_0) randomly in a uniform distribution over the observation domain, with a random orientation θ_0, random speed s_0 between a and b, and random initial rate of change of orientation $\dot{\theta}_0$ commensurate with the value of c, with all of these random variables independent of the Brownian motion B_t from Equations 4 - 6.

1.2.2. Observations

The observations consist of a discrete sequence Y_k of images, each of which is a two-dimensional raster of pixels (i.e. a lattice of points). These images are constructed by superimposing a figure based on a projection of the ship state, X_{t_k}, onto the raster $R = \{(\ell, m)\}$ and incorporating noise by the formula

$$Y_k^{(\ell,m)} = h^{(\ell,m)}(X_{t_k}, V_k^{(\ell,m)}), \tag{7}$$

where $V_k^{(\ell,m)}$ is pixel-by-pixel independent noise in the form of Equation (2). The value of $h^{(\ell,m)}$ will be a random, noisy function of whether or not the pixel (ℓ, m) is within the area of the observation domain (the ocean surface) that a ship with state X occupies.

We assume no preprocessing of the observation rasters before use by the filter.

1.2.3. Objective

The problem is to estimate the conditional distribution of the ship state based on the observations, that is,

$$P(X_{t_k} \in dx \,|\, Y_i, 0 \leq i \leq k). \tag{8}$$

Note that as the number of particles goes to infinity, the branching filter is asymptotically optimal in determining the conditional expectation of the ship state (and consequently, it is asymptotically optimal in determining conditional distributions of functions of the ship state). However, practical filters have a finite amount of computation available and will thus use a finite number of particles, thereby approximating the optimal distribution. Implementors need to ensure that there is no condition of the filtering solution that inhibits the approach to filter optimality with an increasing particle count. The following sections outline three potential conditions that cause such inhibition; these are

- **highly directed signals** - signals with very little randomness associated with them
- **definitive observations** - observations with very high signal to noise ratios
- **clipping** - observations that provide reduced information in some regions of the domain.

We will discuss each of these difficulties, and discuss approaches for mitigating them.

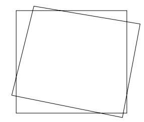

Figure 1. Example of rejected particle whose state is similar to the signal.

2. DIRECTED SIGNALS

Highly directed signal dynamics is a condition that can decrease the practical rate of convergence of the approximation. By directed signal dynamics, we mean any signal law in which the diffusion term, $B(X_t)$ in Equation (1), is small relative to the drift term $A(X_t)$ and the observation space. In the degenerate case, the signal dynamics are completely deterministic.

Any practical filter must operate with some finite particle count. In a scenario where the target is moving nearly deterministically, the particles used to model the target motion also move nearly deterministically. Thus, the particles have a reduced capacity to adapt themselves to the true motion, even if clustered in large numbers near the correct state. For example, a particle that is just behind the actual signal will have no possibility of diffusing forward over time to more closely match the signal state if both the signal and the particle move forward at a fixed rate and direction. Similarly, none of the child particles will have this opportunity even if this nearly accurate particle branches to form many children. Thus, assuming that none of the finite number of particles better matches the actual signal state, the estimate of the true position will at best be somewhat "behind" the true position.

To match the signal position within any given time period, the filter must initialize at time $t = 0$ with a particle that is near enough to the same state as the signal. In the degenerate case, all benefits of adaptive filtering are lost, and in any highly directed case an excessive number of particles is required.

A solution is, counter-intuitively, to add more noise by perturbing the particle states. At each time period, add an independent Gaussian random variable to each continuous state component of each particle, and implement an independent probability to switch to some other value for each discrete state component of each particle. Perform these changes independently for each particle. The relative scale of the perturbations or the chance to switch values for each state component can vary depending on the particular problem. If the orientation of the ship in our example varies widely but the speed is closer to constant, then it may be most useful to perturb the speed and location of each particle but not the orientation.

In order to more closely approximate the optimal filter in time, the intensity of the perturbations should be damped. The perturbations can be monotonically damped in time, or if tracking over a long time period is essential and the track may be lost, damped relative to the sample variance of the particles comprising the approximate filter at that time. The factor $e^{-(x+\gamma)}$ can be multiplied against each Gaussian random perturbation and state switching probability, where $\gamma = ln(\frac{1}{\kappa-1})$, κ is the initial factor, and x is either time or is inversely proportional to the filter variance. In the example problem, for each particle, we could perturb x_{t_k} and y_{t_k} independently by random variables $e^{-(t_k+\gamma)}\mathcal{N}(0,\sigma_{1,k}^2)$, and perturb s_{t_k} by $e^{-(t_k+\gamma)}\mathcal{N}(0,\sigma_{2,k}^2)$.

These perturbations are similar to those of simulated annealing,[8] but in the second case the analogue of temperature in simulated annealing is replaced by a dependency on the filter sample variance rather than being strictly decreasing. Additionally, a permanent level of perturbation can be used to account for imperfectly known signal laws. The factor then becomes $1 + e^{-(x+\gamma)}$.

3. DEFINITIVE OBSERVATIONS

The second area of potential degradation in performance occurs when observations are too definite. This counter-intuitive notion makes sense when we realize that with a clear picture, it is easy to determine whether or not a particle "matches" the signal. Since, with probability one, not one out of any of a finite set of particles in a continuous domain will match the signal perfectly, in some applications all particles are clearly wrong, and are thus deleted.

We illustrate this phenomena with figure 1. If the signal is the square which is "square" to the page, the particle is the tilted square, and we have no noise, then the algorithm will reject the tilted square, because it clearly does not match the observation. Thus in this degenerate case, in order for a particle to sufficiently match the observation to survive until the next round, the particle must have a state that is near enough to the signal state that it perfectly matches at the resolution of the observation.

A similar situation exists for targets that take up a small number of pixels. In such a situation, even particles that are relatively near to the signal have a high probability of being removed. This can happen in our scenario if the helicopter is at a high altitude.

In both of these situations, with a finite number of particles, there can be a significant probability that all of the particles that nearly match the signal will be removed. This eliminates all chance of the filter converging towards the true solution. Thus, although the mathematical formulations for particle filters indicate optimal performance for such a situation, in practice the filter tends to perform poorly in such situations. This is due to the fact that we approach optimality as the number of particles approaches infinity. In these cases, the number of particles needed for a good approximation can be beyond what is computationally feasible.

Having stated the problem, we turn our attention to the solution. There are a number of possible solutions that we have used with varying success depending upon the specific application being investigated.

The first solution technique is to introduce additional noise to the observation. For the case of additive Gaussian noise we accomplish this by multiplying Y_k by a constant. Despite the suspicious nature of the suggestion that the introduction of more noise will be helpful, we have found this technique to be effective in some scenarios in which the observations were too sharply determined for effective filter performance. This is particularly useful for imaging scenarios where the environment can change, for example, using an EO (electro-optical) sensor on a partly cloudy day. This solution technique is not universal in that depending upon the specific scenario, different means must be used to introduce the additional noise. Although initially we may seek to introduce a fair amount of noise, once the particles begin to coalesce around the target, we can reduce the noise. This is analogous to the techniques of simulated annealing. An obvious disadvantage of this approach is that by introducing additional noise, we are losing information in our signal, but with proper tuning of the simulated annealing parameters, it is possible to achieve reasonable performance.

A second solution technique is to extend the support of the particle object template. In this case we seek to blur the edges of our object template, to increase the hits we have with nearly accurate particles. This approach is a very general one, however it does involve a potentially significant increase in computations. The computational complexity of our filter is proportional to the number of pixels in the representation of the signal in the observations, thus when dealing with a small target, we can lose almost an order of magnitude in speed. We have not performed a detailed cost/benefit analysis to determine if the reduction in the number of particles necessary to accommodate the increase in the template size is compensated by increased accuracy. However, we have found that this modification improves the filter performance in situations where the signal to noise ratio is high and the ratio of the number of particles to the size of the observation domain is small.

A third solution to the problem is to use a different rule for branching particles. In particular, ξ as discussed in section 1.1.2 only depends upon a particle and the observation. If the definition of ξ is modified to incorporate the relative sizes of ξ for other particles, then it is possible for situations such as depicted in figure 1 to delete the particles that match poorly, but to keep the particles that match "well enough." This solution has minimal computational implications and is an extremely general solution. Unfortunately it does not fully address the problem of signatures of small extent.

Thus we have three approaches, each of which is applicable to a different subset of problems. The proper solution in a given case is problem dependent.

4. CLIPPING

Particles in some regions of the signal domain may, because of the form of the observation function, obtain less information from an observation than particles in other regions. For example, a particle that represents a ship state could indicate that the ship is in a location at the immediate edge of the observable region, so that the figure superimposed on the observation raster in Equation 7 is "clipped" and has a significant portion that is not within the observed region. In combination with definitive observations, the branching filter can now preferentially remove particles about which it has a greater amount of information while retaining those particles associated with little information that happen to match well, on this limited basis, with the observation.

This will result in approximate filter output that erroneously assigns high probability to the signal being in such a "clipped" region of the domain. The possibility of extreme results in some small set of observation components (e.g. raster pixels) near to the "clipped" region cause particles in some such states to be replicated with extremely high probability. At the same time, definitive observations can cause all particles not in a "clipped" region to have a much smaller probability of survival, since there is enough information in the observations regarding these particles to ascertain with more assurance the imperfect match between the particle and true signal states. Because the branching filter adapts to this information, it will no longer retain particles in a position to generate children near to the actual signal state, so that even in successive time frames, the filter can no longer adapt correctly to the observations.

A possible solution is to incorporate a probability, based on the extent to which survival of a particle was predicated on limited information availability from the observation, that the particle will not follow the rules of the branching particle-based filter for propagation. Instead of duplicating the state values of the parent particle, in some randomly determined cases generate uniform random state values for the particle. This random determination will take account of the information available regarding the particle in the present observation such that poorly determined particles will have a higher probability of generating randomly placed offspring. While this modification causes the filter to ignore, to some extent, the information from the observations, it assists in reducing the possibility of catastrophic incorrect adaptation of the filter which can be a significant risk in the case of definitive observations.

Another solution is to act only if the sample variance of the filter particles is low, but many particles are removed in the selection stage. Note that this is only likely to occur if the observation indicates that the filter has an approximate distribution that poorly matches the signal state. That is, the incorrect adaptation toward a region of the signal domain that has reduced impact on the observations has already occurred and later observations are confirming this. The filter can be modified to generate uniform random state data for some randomly determined particles such that the probability of random placement increases as the state of the parent particle is closer to the current filter sample mean. This probabilistically limits the extent to which the filter is ignoring observation data to those cases in which the approximate nature of the filter is likely to have caused incorrect adaptation at previous times.

A form of this last solution is to alter the algorithm by which the total particle count is re-balanced. If many particles are removed during one step, the algorithm can generate randomly placed particles instead of duplicating particles in regions that provide limited information.

5. CONCLUSION

While branching particle-based filters provide a mathematically rigorous and very general solution to a broad range of filtering problems, it can take extra effort to practically apply particle filters to some real-world problems. While the approximations of the particle-based filters provably converge to the optimal filter as the number of particles is increased, in some cases there is a difficulty in the problem that, absent remedy, necessitates a practically infeasible number of particles to obtain adequate filter performance. The resolutions to these difficulties can be specific to the filtering problem in question, but general categories of common problems can be described and resolutions to these difficulties can be provided that require only the selection of a parameter or two to obtain practicality of the filter.

This article has outlined three such scenarios: the law of the signal in the problem can be highly directed, or the observations can have little noise and great definition of a correct match, or there can be areas of the signal domain for which observations provide little data. In each of these cases, modifications can be made to the filter algorithms and formulae so that the filter provides a practical result without requiring as great an increase in the total number of particles as would otherwise be required.

By using the techniques outlined in this article, researchers at Lockheed Martin have improved the performance of branching filters and reduced the number of particles required for effective operation on a number of problems of interest. Particle-based filters, and the branching filter in particular, have proven to be adaptable to a wide variety of circumstances. Among the strengths are the ability to quickly implement a branching filter for a new problem and the capacity to reformulate and modify that filter to avoid performance issues.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support and sponsorship of Lockheed Martin Tactical Systems, Lockheed Martin Canada, Acoustic Positioning Research of Edmonton, the Pacific Institute for the Mathematical Sciences, the Natural Science and Engineering Research Council (NSERC) through the Prediction in Interacting Systems (PINTS) center of the Mathematics of Information Technology and Complex Systems (MITACS) network of centers of excellence, and the University of Alberta.

REFERENCES

1. A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.
2. H. A. P. Blom and Y. Bar-Shalom, “The interacting multiple model for systems with Markovian switching coefficients,” *IEEE Trans. Auto. Cont.* **33**, pp. 780–783, 1988.
3. N. Gordon, “Bayesian methods for tracking,” 1994.
4. P. Del Moral and G. Salut, “Non-linear filtering using Monte Carlo particle methods,” *C.R. Acad. Sci. Paris* **320, Série I**, pp. 1147–1152, 1995.
5. T. G. Kurtz and J. Xiong, “Particle representations for a class of nonlinear SPDEs,” *Stochastic Processes and their Applications* **83**, pp. 103–126, 1999.
6. M. A. Kouritzin, “A pathspace branching particle filter.” Preprint.
7. D. Ballantyne, H. Chan, and M. Kouritzin, “A novel branching particle method for tracking,” *Proceedings of SPIE, Signal and Data Processing of Small Targets 2000* **4048**, pp. 277–287, 2000.
8. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, “Optimization by simulated annealing,” *Science, Number 4598, 13 May 1983* **220, 4598**, pp. 671–680, 1983.