Numerical Methods for Nonlinear Filtering **UNFINISHED**

Bill Triggs
INRIA Rhône-Alpes,
655 avenue de l'Europe, 38330 Montbonnot Saint-Martin, France.
Bill.Triggs@inrialpes.fr

Abstract

This note briefly summarizes numerical approaches to state estimation (filtering) for nonlinear discrete time continuous state Markov systems…

1 Notation

Conditional probability densities are denoted $\mathbf{dp}(\mathbf{x}|\ldots)$. These are densities in $\mathbf{x}$ — measures rather than numerically valued functions. Only ratios of like densities are numbers. $\langle\,\cdot\,|\ldots\rangle$ denotes **conditional expectation value**: $\langle \mathbf{f}(\mathbf{x})|\ldots\rangle \equiv \int_{\mathbf{x}} \mathbf{f}(\mathbf{x}) \cdot \mathbf{dp}(\mathbf{x}|\ldots)$.

2 Hidden Markov Systems

A discrete time **Markov system** is a sequence of **state spaces** $\{\ldots,\mathcal{X}_{t-1},\mathcal{X}_t,\ldots\}$ together with a sequence of **probabilistic state transition** laws $\mathcal{X}_{t-1} \to \mathcal{X}_t$. **Probabilistic** means that even if the input state is exactly known, the output state is intrinsically uncertain. The uncertainty is encapsulated in a **state transition kernel** $\mathbf{dp}(\mathbf{x}_t|\mathbf{x}_{t-1})$. This gives the probability density of ending up near state $\mathbf{x}_t \in \mathcal{X}_t$ given that you started from state $\mathbf{x}_{t-1} \in \mathcal{X}_{t-1}$. If the initial state is uncertain, as described by a probability density $\mathbf{dp}(\mathbf{x}_{t-1}|\ldots)$, the uncertainty in the final state is described by the convolution

$$\mathbf{dp}(\mathbf{x}_t|\ldots) = \int_{\mathbf{x}_{t-1}} \mathbf{dp}(\mathbf{x}_t|\mathbf{x}_{t-1}) \cdot \mathbf{dp}(\mathbf{x}_{t-1}|\ldots)$$

This just says that the total probability of ending up in state $\mathbf{x}_t$ is the sum of the probabilities of getting there via any of the exhaustive, mutually exclusive states $\mathbf{x}_{t-1}$.

Markov means that the transition law has no hidden dependencies. The transition probability depends *only* on the input state, not on how the system got into it: given the state itself, conditioning on past history has no effect. This decoupling makes Markov systems relatively tractable. With a sufficiently rich state descriptor, many physical systems become Markov.

A state is **hidden** if it is not directly observable, but some of its properties can be inferred indirectly by observing a set of correlated **output variables**. We model this as a series of **output spaces** $\{\ldots,\mathcal{Z}_{t-1},\mathcal{Z}_t,\ldots\}$ together with **probabilistic output laws** $\mathcal{X}_t \to \mathcal{Z}_t$ described by **output kernels** $\mathbf{dp}(\mathbf{z}_t|\mathbf{x}_t)$. $\mathbf{dp}(\mathbf{z}_t|\mathbf{x}_t)$ is a density in $\mathbf{z}_t \in \mathcal{Z}_t$, giving the probability of observing an output near $\mathbf{z}_t$ when the state is exactly $\mathbf{x}_t$. For an uncertain state $\mathbf{x}_t$, the output density is the convolution

$$\mathbf{dp}(\mathbf{z}_t|\ldots) = \int_{\mathbf{x}_t} \mathbf{dp}(\mathbf{z}_t|\mathbf{x}_t) \cdot \mathbf{dp}(\mathbf{x}_t|\ldots)$$

This says that the total probability of output $\mathbf{z}_t$ is the sum of the probabilities of seeing $\mathbf{z}_t$ via any of the exhaustive, mutually exclusive states $\mathbf{x}_t$. The distribution of observations depends only on the state itself, not on how the system got into it.

We will imagine the state and output spaces $\mathcal{X}_t$ and $\mathcal{Z}_t$ as continuous nonlinear manifolds. For discrete spaces, integrals can be replaced with summations. The state and output spaces may change structure and dimension as t progresses, *e.g.* to allow for model order transitions or variable types of measurement. In applications, the abstract label sequence $\{\ldots,t-1,t\ldots\}$ usually stands for a non-decreasing but perhaps unevenly spaced sequence of physical times.

Although a hidden Markov system does always have a definite state, it is not directly accessible and our imperfect knowledge only allows us to derive a 'subjective' conditional probability density $\mathbf{dp}(\mathbf{x}_t|\mathbf{z}_t, \mathbf{z}_{t-1}, \ldots)$ for the true state given all of our observations up to the present. This is called **filtering**. Given a subjective probability density $\mathbf{dp}(\mathbf{x}_t|\ldots)$ for the system state at time t, and an uncertain observation $\mathbf{z}_t$ of the output at t, Bayes law allows the estimated state density to be refined as follows:

$$\mathbf{dp}(\mathbf{x}_t|\mathbf{z}_t, \ldots) = \frac{\mathbf{dp}(\mathbf{z}_t|\mathbf{x}_t)}{\mathbf{dp}(\mathbf{z}_t)} \cdot \mathbf{dp}(\mathbf{x}_t|\ldots)$$

The probability of each state is enhanced proportionally to the relative likelihood $\mathbf{dp}(\mathbf{z}_t|\mathbf{x}_t)$ of its producing the observed output $\mathbf{z}_t$. The state-independent observation prior $\mathbf{dp}(\mathbf{z}_t)$ is difficult to estimate, but for a fixed observation $\mathbf{z}_t$ it simply amounts to an overall normalization constant that can be evaluated by requiring the posterior state density to be normalized: $\int_{\mathbf{x}_t} \mathbf{dp}(\mathbf{x}_t|\mathbf{z}_t, \ldots) = 1$.

Combining the state transition and observation-update laws, the net result of an evolve-observe-update cycle on the state density is given by the **posterior state transition** rule:

$$\mathbf{dp}(\mathbf{x}_t|\mathbf{z}_t, \ldots) = \int_{\mathbf{x}_{t-1}} \mathbf{dp}(\mathbf{x}_t|\mathbf{z}_t, \mathbf{x}_{t-1}) \cdot \mathbf{dp}(\mathbf{x}_{t-1}|\ldots)$$

Here, the **posterior kernel**

$$\mathbf{dp}(\mathbf{x}_t|\mathbf{z}_t, \mathbf{x}_{t-1}) = \frac{\mathbf{dp}(\mathbf{z}_t|\mathbf{x}_t)}{\mathbf{dp}(\mathbf{z}_t)} \cdot \mathbf{dp}(\mathbf{x}_t|\mathbf{x}_{t-1})$$

combines the uncertain nonlinear system dynamics, the output law, and the Bayesian observation-update law into a single convolution operator on state densities.

…relation to SDE's [3, 4, 2]

…continuous-time SDE $\leftrightarrow$ discrete-time kernel

…kernel as elementary solution of SDE

3 Nonlinear Filtering

Filtering is the science of estimating current properties of a hidden state from current and past observations $\{\mathbf{z}_t, \mathbf{z}_{t-1}, \ldots\}$. This basically involves estimating the fully conditioned state density $\mathbf{dp}(\mathbf{x}_t|\mathbf{z}_t, \mathbf{z}_{t-1}, \ldots)$ and its associated statistics. Here we are interested in numerically useful techniques for filtering general nonlinear hidden Markov systems with continuous state spaces. Exact solutions are seldom available, so approximations must be made.

3.1 Parametric State Densities

Numerically, it is impossible to represent the entire infinite dimensional space of all state probability distributions. We must truncate the problem by restricting attention to some sequence of finite-dimensional parametric families of densities $\{\ldots, \mathcal{M}_{t-1}, \mathcal{M}_t, \ldots\}$. Each $\mathcal{M}_t$ denotes a finite dimensional parameter space, and each point $\boldsymbol{\mu}_t \in \mathcal{M}_t$ determines a (possibly singular or unnormalized) state probability density $\mathbf{dp}(\mathbf{x}_t|\boldsymbol{\mu}_t)$ on $\mathcal{X}_t$. For example, $\boldsymbol{\mu}_t$ might be the mean and variance of a Gaussian, a set of moments in a central moment expansion, the weights of a spline approximation to the distribution, *etc*.

The exact state density evolution laws must be approximated by parametric evolution laws $\mathcal{M}_{t-1} \to \mathcal{M}_t$. Assuming that the input density $\mathbf{dp}(\mathbf{x}_{t-1}|\ldots)$ has already been approximated by a distribution $\boldsymbol{\mu}_{t-1} \in \mathcal{M}_{t-1}$, the exact evolution-observation-update law gives an output density $\mathbf{dp}(\mathbf{x}_t|\mathbf{z}_t, \boldsymbol{\mu}_{t-1})$. This does not usually lie in $\mathcal{M}_t$, so it is necessary to **reproject** it into $\mathcal{M}_t$ at each time step. The net result will be some nonlinear recursive estimation law $\boldsymbol{\mu}_{t-1} \to \boldsymbol{\mu}_t(\mathbf{z}_t, \boldsymbol{\mu}_{t-1})$ for the parameters $\boldsymbol{\mu}_t(\mathbf{z}_t, \ldots)$ of the subjective state distribution.

For this approach to make sense, the parametric families $\mathcal{M}_t$ must satisfy several criteria:

- They must be approximately closed under the exact evolution-observation-update law, or the reprojection will be meaningless.
- They must be stable in the sense that the exact evolution law tends to push other densities (especially the initial one!) towards them.
- They must be sufficiently rich to model any aspects of the true uncertainty relevant to the application.

For example, for affine state spaces and transition laws with Gaussian uncertainties, the subspace of Gaussian densities is closed (*i.e.* state updates preserve the Gaussian form). If the system is also observable and dynamically stable, the central limit theorem suggests that an arbitrary initial distribution will asymptotically tend to be pushed onto the Gaussian subspace by repeated averaging and dilution, so the Gaussian form is also stable.

For more general nonlinear systems, closedness and stability are much harder to achieve, although a handful of special ("almost linear") examples are known (Maybank,…). Exact solutions to the density evolution are rarely available, and even when they are their parametric complexity tends to increase exponentially as they are iterated. More often than not, the model spaces $\mathcal{M}_t$ must be chosen heuristically, with little formal justification. Gaussians, Gaussian mixtures, moment expansions, and spline-based decompositions are all popular choices.

3.2 Describing Functions

Once a suitable parametric family $\{\mathbf{dp}(\mathbf{x}_t|\boldsymbol{\mu}_t)\;|\boldsymbol{\mu}_t \in \mathcal{M}_t\}$ has been chosen, we must define some means of projecting an arbitrary state density $\mathbf{dp}(\mathbf{x}_t|\ldots)$ onto it. One approach is to use a vector of *expectation values* to summarize the information in the distribution. This tends to be relatively robust because it makes very few assumptions about the structure of the input distribution, and expectation values (being integral operators) are numerically rather stable. Of course, the results will only be useful if the expectation values and the family $\mathcal{M}_t$ are sufficiently rich to capture the relevant features of the input distribution.

For any parametric family $\mathcal{M}$, the expectation value of any function $\mathbf{m}(\mathbf{x})$ is some (complex, nonlinear) function of the parameters

$$\langle \mathbf{m}|\boldsymbol{\mu}\rangle \equiv \int_{\mathbf{x}} \mathbf{m}(\mathbf{x}) \cdot \mathbf{dp}(\mathbf{x}|\boldsymbol{\mu})$$

With a suitable choice of $\mathbf{m}(\cdot)$, this relation can often be (approximately) inverted to give the parameters in terms of the expectation values

$$\boldsymbol{\mu} \approx \boldsymbol{\mu}\left(\,\langle \mathbf{m}\rangle\,\right)$$

where $\boldsymbol{\mu}(\cdot)$ is some known function depending on the parametric family $\mathcal{M}$. For example the μ and σ parameters of a Gaussian are functions of its first two moments $\langle x\rangle$ and $\langle x^2\rangle$. This equation defines a value of $\boldsymbol{\mu}$ for *any* input distribution, so it gives a heuristic method of "projecting" an arbitrary distribution onto $\mathcal{M}$. The projection is self-consistent provided it fixes $\mathcal{M}$ itself: $\boldsymbol{\mu} = \boldsymbol{\mu}(\,\langle \mathbf{m}|\boldsymbol{\mu}\rangle\,)$. The functions $\langle \mathbf{m}\rangle$ or $\boldsymbol{\mu}(\,\langle \mathbf{m}\rangle\,)$ are called **describing functions** because they summarize a potentially complicated situation by abstracting a few overall statistical parameters.

This process allows us to express the subjective state density update as a recursion relation involving: *(i)* the evaluation of the input density $\mathbf{dp}(\mathbf{x}_{t-1}|\boldsymbol{\mu}_{t-1})$ from $\boldsymbol{\mu}_{t-1}$; *(ii)* an evolution-observation-update integral over $\mathbf{x}_{t-1}$; *(iii)* a summarization integral over $\mathbf{x}_t$; and *(iv)* and a nonlinear 'output' mapping $\boldsymbol{\mu}_t(\cdot)$

$$\begin{aligned}
\boldsymbol{\mu}_t &\approx \boldsymbol{\mu}_t\left(\,\langle \mathbf{m}_t|\mathbf{z}_t,\boldsymbol{\mu}_{t-1}\rangle\,\right)\\
\langle \mathbf{m}_t|\mathbf{z}_t,\boldsymbol{\mu}_{t-1}\rangle &\equiv \int_{\mathbf{x}_t} \mathbf{m}(\mathbf{x}_t)\cdot \mathbf{dp}(\mathbf{x}_t|\mathbf{z}_t,\boldsymbol{\mu}_{t-1})\\
\mathbf{dp}(\mathbf{x}_t|\mathbf{z}_t,\boldsymbol{\mu}_{t-1}) &\equiv \int_{\mathbf{x}_{t-1}} \mathbf{dp}(\mathbf{x}_t|\mathbf{z}_t,\mathbf{x}_{t-1})\cdot \mathbf{dp}(\mathbf{x}_{t-1}|\boldsymbol{\mu}_{t-1})
\end{aligned}$$

In practice, the equivalent kernel based formulation is often more useful

$$\begin{aligned}
\langle \mathbf{m}_t|\mathbf{z}_t,\boldsymbol{\mu}_{t-1}\rangle &\equiv \int_{\mathbf{x}_{t-1}} \langle \mathbf{m}_t|\mathbf{z}_t,\mathbf{x}_{t-1}\rangle\cdot \mathbf{dp}(\mathbf{x}_{t-1}|\boldsymbol{\mu}_{t-1})\\
\langle \mathbf{m}_t|\mathbf{z}_t,\mathbf{x}_{t-1}\rangle &\equiv \int_{\mathbf{x}_t} \mathbf{m}(\mathbf{x}_t)\cdot \mathbf{dp}(\mathbf{x}_t|\mathbf{z}_t,\mathbf{x}_{t-1})
\end{aligned}$$

Here $\mathbf{m}_t(\cdot)$ and $\boldsymbol{\mu}_t(\cdot)$ are known nonlinear functions depending only on the parametric form of the distribution and chosen to give accurate, stable estimation behaviour. The key component is the **describing kernel** $\langle \mathbf{m}_t|\mathbf{z}_t,\mathbf{x}_{t-1}\rangle$, which combines the uncertain nonlinear dynamics, the observation-update law, and the body of the summarization process in a single density convolution operator. The 'output function' $\boldsymbol{\mu}_t(\cdot)$ simply represents a nonlinear change of variables that could have been absorbed into the parametric form of the distributions. Unnormalized distributions can be handled by including the constant function 1 in the describing vector and working with $\langle \mathbf{m}\rangle/\langle 1\rangle$.

4 Multidimensional Quadrature

The above state space integrals must often be evaluated numerically. This section considers multidimensional quadrature techniques suitable for this sort of problem [5]. Multidimensional integration is not difficult but it becomes very expensive as the dimension increases, to the extent that simple grid-based techniques are almost never practical in dimensions greater than 3 or 4. In many practical applications of filtering, the state dimension is high and speed is more important than high precision, so we will emphasize fast approximate high dimensional techniques.

We will concentrate on the $\mathbf{x}_{t-1}$ integral. The $\mathbf{x}_t$ (describing kernel) one is similar and potentially more complicated as it involves the full nonlinearity of the system, the effects of the observation-update law and the describing functions $\mathbf{m}_t(\cdot)$. However in many applications the observations are relatively weak[1], so that the uncertainty in the input state is usually significantly greater than the spread of the posterior kernel $\mathbf{dp}(\mathbf{x}_t|\mathbf{z}_t,\mathbf{x}_{t-1})$. In such cases it is often possible to find an adequate analytic approximation for the describing kernel integral, given that it will in any case be smoothed against the input state density.

Ultimately, most numerical quadrature rules reduce to a weighted sum of function samples

$$\begin{aligned}\langle \mathbf{m}_t|\mathbf{z}_t,\boldsymbol{\mu}_{t-1}\rangle &\equiv \int_{\mathbf{x}_{t-1}} \langle \mathbf{m}_t|\mathbf{z}_t,\mathbf{x}_{t-1}\rangle \cdot \mathbf{dp}(\mathbf{x}_{t-1}|\boldsymbol{\mu}_{t-1}) \\ &\approx \sum_i w^i_{t-1} \cdot \langle \mathbf{m}_t|\mathbf{z}_t,\mathbf{x}^i_{t-1}\rangle \cdot \mathbf{dp}(\mathbf{x}^i_{t-1}|\boldsymbol{\mu}_{t-1})\end{aligned}$$

The sample points $\mathbf{x}^i_{t-1}$ and weights w^i_{t-1} may depend on the known parameters $\boldsymbol{\mu}_{t-1}$ and $\mathbf{z}_t$, the parametric form of the input distribution $\mathbf{dp}(\mathbf{x}_{t-1}|\boldsymbol{\mu}_{t-1})$, and the describing kernel $\langle \mathbf{m}_t|\mathbf{x}_{t-1},\mathbf{z}_t\rangle$. In effect, the prior distribution is approximated by a finite sum of point distributions (delta functions). Each point distribution is propagated independently through the forward dynamics and measurement processes, and the combined results are 'summarized' or regularized by projection onto a parametric space of well-behaved distributions. This works reasonably well provided the dynamical noise, the summarization process, and our tendency to use low-order statistics to characterize the output are sufficient to remove the high frequency perturbations introduced by the sampling process.

4.1 Gaussian Quadrature

When available, Gaussian quadrature rules [5] are potentially well adapted to this sort of problem. In these, the weights and sample points are carefully chosen to give accurate results for the integration of some set of elementary test functions $g(\mathbf{x})$ against a prespecified density $\mathbf{df}(\mathbf{x})$. The resulting rules tend to be quite model sensitive, but within their domain of applicability they can be orders of magnitude more efficient than the best grid-based methods. The most obvious approach is to optimize the weights $w^i_{t-1}(\boldsymbol{\mu}_{t-1})$ and sample points $\mathbf{x}^i_{t-1}(\boldsymbol{\mu}_{t-1})$ for integration against the given input density $\mathbf{dp}(\mathbf{x}_{t-1}|\boldsymbol{\mu}_{t-1})$, so that the describing function estimate becomes simply

$$\langle \mathbf{m}_t|\mathbf{z}_t,\boldsymbol{\mu}_{t-1}\rangle \approx \sum_i w^i_{t-1} \cdot \langle \mathbf{m}_t|\mathbf{z}_t,\mathbf{x}^i_{t-1}\rangle$$

…problem of choosing weights
…trade stability (positive weights) vs. higher order
…avoid Cartesian product grids – too big!
…newer refs??

4.2 (Quasi) Monte Carlo Methods

It is difficult to find efficient explicit quadrature rules in high dimensions, especially when the dynamics, prior or observation distributions are multimodal or highly nonlinear. In these cases, the Monte Carlo approach may be the only practicable method of estimating the integral. The basic idea is to use a quadrature rule as above, but to choose the sample point positions (quasi-) randomly rather than systematically.

…formulae …

The sample points must be generated according to some underlying distribution. Although the quadrature weights are adjusted to cancel out the bias introduced by this, the overall efficiency and precision of the technique depend strongly on the distribution chosen. In practice the samples must usually be distributed rather sparsely over the available volume, so it is important to make the most of each one. To avoid samples being 'wasted' in areas of low probability, the sample distribution should be chosen to closely mirror the relative contribution (posterior probability) of each region to the integral. This is called **importance weighting**. Moreover, small randomly chosen sets of samples tend to be distributed rather unevenly. **Quasi-random** methods try to improve this by using systematic pseudo-random sample placement heuristics. Finally, **resampling** schemes use the results of a trial integration to generate a better sample distribution, so that a subsequent integration produces more accurate results.

All of these techniques trade implementation and computational complexity against 'wasted' samples. Completely random placement is simple and robust but tends to be very inefficient, while carefully importance-weighted quasi-random methods are analogous to Gaussian quadrature rules: they can be efficient, but they tend to be quite complex and model-sensitive.

[1] Filtering augments uncertain or incomplete observations with dynamical predictions. It is not usually worthwhile if the observational uncertainty is much smaller than the dynamical uncertainty.

4.2.1 Importance Weighting

…*c.f.* blurb in numerical recipes…

4.2.2 Resampling

…Is quadratic resampling [1] useful in general case??
…Need to importance weight on posterior not prior, *i.e.* approximately push observation density back through to input space.
…How to handle multimodality??

4.3 Particle-based Methods

The particle-based approach is a simple, intuitively appealing variant of the Monte Carlo method. It was originally developed for physical simulations of diffusion processes. The main difference from Monte Carlo quadrature is that the state transition uncertainty is sampled randomly rather than being represented as an explicit probability density.

The basic idea is to model probability distributions as clouds of noninteracting randomly diffusing 'particles' (samples) whose local density is proportional to the probability density. Each particle follows the dynamical law independently and is randomly displaced ('scattered') by the dynamical noise. Doing this for all of the particles and estimating the resulting particle density gives an estimate of the state probability density prior to the observation. This is then modulated by the observation likelihood to estimate the posterior state density, and the cycle continues.

An easy way to do the modulation is to give each particle a probability weight: the observation likelihood scales these weights, and then the local weighted particle density gives an estimate of the posterior state density. In theory, weighted particles can be tracked through several time-steps, since the weight modulation (observation-update) is independent of the dynamics (particle position). However in practice resampling of the distribution is advisable to suppress particles whose weights have become negligible and split ones whose weights have become disproportionately large.

Particle-based methods are simple to implement and parallelize. Statistics of the resulting particle distributions are easily estimated by simple weighted sums. However importance weighting, quasi-random sampling and resampling are probably even more important for them than they are for Monte Carlo techniques.

…refs??

4.4 Sums of Gaussians

…useful in highly multimodal cases?? — refs??

4.5 Hybrid Techniques

…analytic evaluation of some dimensions of integral
…fast and slow dynamics, tight and loose observations…

5 Examples

In this section we consider a few special cases for which particular simplifications occur.

5.1 Noiseless Dynamics

If there is no uncertainty in the underlying system dynamics, the state evolution can be described by a nonlinear time evolution *function* $\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_{t-1})$. The prior and posterior kernels are singular densities with all their weight at a single point[2]:

$$\begin{aligned} \mathbf{dp}(\mathbf{x}_t|\mathbf{x}_{t-1}) &= \boldsymbol{\delta}(\mathbf{x}_t, \mathbf{f}_t(\mathbf{x}_{t-1})) \\ \mathbf{dp}(\mathbf{x}_t|\mathbf{z}_t, \mathbf{x}_{t-1}) &= \boldsymbol{\delta}(\mathbf{x}_t, \mathbf{f}_t(\mathbf{x}_{t-1})) \cdot \frac{\mathbf{dp}(\mathbf{z}_t|\mathbf{f}_t(\mathbf{x}_{t-1}))}{\mathbf{dp}(\mathbf{z}_t)} \end{aligned}$$

In this limit, the describing kernel integral collapses to a single nonlinear function evaluation

$$\langle \mathbf{m}_t|\mathbf{x}_{t-1}, \mathbf{z}_t\rangle = \mathbf{m}_t(\mathbf{x}_t) \cdot \frac{\mathbf{dp}(\mathbf{z}_t|\mathbf{x}_t)}{\mathbf{dp}(\mathbf{z}_t)}\bigg|_{\mathbf{x}_t=\mathbf{f}_t(\mathbf{x}_{t-1})}$$

and the parametric state density update quadrature becomes a simple sum of evaluations of the describing functions $\mathbf{m}_t(\cdot)$ at time-evolved sample points

$$\begin{aligned} &\left\langle \mathbf{m}_t|\mathbf{z}_t, \boldsymbol{\mu}_{t-1}\right\rangle \\ &= \int_{\mathbf{x}_{t-1}} \mathbf{m}_t(\mathbf{x}_t) \cdot \frac{\mathbf{dp}(\mathbf{z}_t|\mathbf{x}_t)}{\mathbf{dp}(\mathbf{z}_t)} \cdot \mathbf{dp}(\mathbf{x}_{t-1}|\boldsymbol{\mu}_{t-1})\bigg|_{\mathbf{x}_t=\mathbf{f}_t(\mathbf{x}_{t-1})} \end{aligned}$$

[2] $\boldsymbol{\delta}(\mathbf{x}, \mathbf{y})$ denotes a Dirac delta function: a singular density in $\mathbf{x}$ with unit weight at $\mathbf{y}$ and no weight elsewhere.

$$\approx \sum_i w_{t-1}^i \cdot \mathbf{m}_t(\mathbf{x}_t) \cdot \frac{\mathbf{dp}(\mathbf{z}_t|\mathbf{x}_t)}{\mathbf{dp}(\mathbf{z}_t)} \cdot \mathbf{dp}(\mathbf{x}_{t-1}^i|\boldsymbol{\mu}_{t-1})\Big|_{\mathbf{x}_t=\mathbf{f}_t(\mathbf{x}_{t-1}^i)}$$

The Bayesian re-weighting of the state density following an observation simply produces a rescaling of the quadrature weights

$$w_{t-1}^i \quad \longrightarrow \quad \frac{\mathbf{dp}(\mathbf{z}_t|\mathbf{f}_t(\mathbf{x}_{t-1}^i))}{\mathbf{dp}(\mathbf{z}_t)} \cdot w_{t-1}^i$$

In fact, the state density after an arbitrary number of time steps is simply an iterated series of nonlinear deformations and re-weightings of the input density

$$\mathbf{dp}(\mathbf{x}_t|\mathbf{z}_1,\ldots,\mathbf{z}_t) \approx \sum_i w_0^i \cdot \mathbf{dp}(\mathbf{x}_0^i|\boldsymbol{\mu}_0) \cdot$$
$$\cdot \prod_{t'=1}^{t} \frac{\mathbf{dp}(\mathbf{z}_{t'}|\mathbf{x}_{t'})}{\mathbf{dp}(\mathbf{z}_{t'})}\Big|_{\mathbf{x}_{t'}=(\mathbf{f}_{t'}\cdot\cdots\cdot\mathbf{f}_1)(\mathbf{x}_0^i)}$$

…1st order dynamical noise correction??

References

[1] Jérôme Barraquand. Numerical valuation of high dimensional multivariate european securities. Technical Report PRL-RR-26, Digital Equipment Corporation, Paris Research Laboratory, 85 avenue Victor Hugo, 92563 Rueil-Malmaison, France. FTP: gatekeeper.pa.dec.com, 1993. (Quadratic resampling, references and summary of Monte Carlo approaches for unimodal distributions).

[2] Nicolas Bouleau and Dominique Lépingle. *Numerical Methods for Stochastic Processes*. Wiley series in Probability and Mathematical Statistics, 1995?? (Detailed numerical analysis for accurate low dimensional methods – emphasis on analysis rather than applications. £71).

[3] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*, volume 23 of *Applications of Mathematics*. Springer Verlag, 1992. (Mainly concentrates on accurate ($\geq 2^{nd}$ order) techniques for low dimensional problems. £53).

[4] P. E. Kloeden, E. Platen, and H. Schurz. *Numerical Solution of Stochastic Differential Equations through Computer Experiments*. Springer Verlag Universitext, 1993. (Something like an examples book for [3]. £30).

[5] A. H. Stroud. *Approximate Calculation of Multiple Integrals*. Prentice Hall, 1971.