PARTICLE FILTERING: THE FIRST 25 YEARS AND BEYOND

Simon Godsill

Signal Processing and Communications (SigProC) Laboratory
University of Cambridge, UK

ABSTRACT

This paper presents a survey of the ideas behind the particle filtering, or sequential Monte Carlo, method, from at least 1930 up to the present day. The particle filter, which is now 25 years old, has been an immensely successful and widely used suite of methods for filtering and smoothing in state space models, and it is still under research today. The key ideas that led to the development in 1993 of the original particle filter, the *bootstrap filter*, were Monte Carlo integration, Importance Sampling, Bayesian updating, Probabilistic State Space models, and Sampling-Importance-Resampling. We survey these methods within their historical context and then provide a general framework for description of most current variants on the particle filtering methodology, based upon updating the joint smoothing distribution of the states. This framework aids in the understanding of the various elements of a particle filter, including resampling. prediction and weighting. We further summarise recent developments and look to the future of the methodology.

Index Terms— Importance sampling, Monte Carlo, MCMC, filtering, smoothing, bootstrap filter

1. INTRODUCTION

In the year of writing this paper the bootstrap filter of [1] is 25 years old, and there is much to celebrate about this fruitful development. Research is still active in the area, generating new methods for Big Data and for high-dimensional models, as well as applications in many surprising fields. The historical development of the techniques required for particle filtering was long and complex, commencing in the 1930s at least. The educational approach we take in this paper is to couple the required methodological developments with the historical commentary: the methodological basics are introduced alongside notes on their historical development, with the aim of producing an informal tutorial to the topic, while setting the methods in the bigger picture of the historical process. This I regard as important for students since it is very easy to focus on just one algorithm, say the classic bootstrap filter [1] without an appreciation of where it sits amongst other developments, thus limiting the possibility to see new directions and potential combinations with other techniques. There are many good tutorials now available on particle filtering, see [2, 3, 4, 5, 6, 7, 8, 9, 10, 11] to which we refer the reader for more detailed coverages than are possible here. And with only one page of citations possible, we necessarily neglect to cite many great papers on this topic, for which we apologise. In particular we do not have space to discuss the excellent theoretical analysis papers by authors such as Del Moral, Crisan, Lyons to name but a few of the founders.

The paper presents first ideas from before the 'Monte Carlo' era, including Bayes' Theorem. Then we present in historical order all of the ingredients that were required for the development of the particle filter. We next present a general framework for describing most of the current variants of particle filtering, followed by a look to recent trends and the future.

2. PREHISTORY

The 'Laws of Chance', or probability, were formulated and studied by De Moivre and Bernoulli, and following this early work the fundamentals for the Bayesian updating required for particle filters arrived as early as the 18th century. In 1763,[1] a posthumous essay *towards solving a Problem in the Doctrine of Chances* by the Rev. Thomas Bayes [12] appeared, giving the first known example of 'inverse probability' in which a parameter of a distribution is inferred by what we now term 'Bayesian Reasoning'. A more general presentation was later given by Laplace [13] and now we have the formulation in enough generality for our purposes today: if we observe a random outcome y from an experiment, then the probability of a second, linked, random event x can be determined as:

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \tag{1}$$

where $P(x)$ is the so-called *prior* probability of x, the probability that x occured in the absence of any information about whether y has occurred, and $P(y|x)$ is the *likelihood*, the probability that y occurs *given* that x has occurred. A simple example of linked events of this type would take x as the event that it rained *yesterday* and y being the event that it has rained *today*. The same formula of course occurs when x and y are dependent random variables and probability mass functions $P()$ are replaced by probability density functions $f()$ $g()$ and $p()$.

3. EARLY DEVELOPMENTS: 1930S TO 1990S

Sadly for us, Bayesian theory was largely discredited for much of the 20th century, partly because of the difficulty in determining the prior distribution $P(x)$, which could be subjective and could bias the results of reasoning, and instead 'classical statistics' in the Fisherian mode was dominant. Nevertheless, an increasing group of academics and practitioners pioneered the use of Bayesian theory and applications, including for example such revered names as Harold Jeffreys, Richard Cox, Jimmy Savage, Bruno de Finetti, Denis Lindley and Adrian Smith, a lineage that we have chosen partly as it takes us all the way to the early 1990s and the origins of particle filtering with Adrian Smith's research group at Imperial College London. These researchers report that it was in fact very difficult to have Bayesian methods published in the top journals owing to fierce resistance from the mainstream statistical community. It is interesting to observe that

[1] two days before Christmas, presented by a Mr Price to the Royal Society in London

the (probabilistic) needle has (finally) swung back in the opposite direction, with Bayesian theory usually being the starting point for understanding of Machine Learning and Artificial Intelligence methods in the current climate. A personal observation is that the Engineering and specifically the Signal Processing community was much more open than the statistical community to Bayesian approaches, from the 1990s onwards, provided issues of prior modelling were carefully justified.

3.1. State space models, filtering and Kalman filters

We are now in a position to state the main underpinning equations for Bayesian updating of probabilities in time series models. It is convenient to specify our models as Markovian state space models (a term that originated in the Control Engineering discipline through Rudolf Kalman in 1960) with time index t, in which a hidden (unobserved) random state x_t is to be inferred from some data y_t.[2] Specifically, the state is assumed to evolve randomly over time, starting at $t = 0$, conditioned only on where it was at the previous time step, and the observation depends only upon the value of the current state, summarised as a discrete time state space model:

$$f(x_t|x_{0:t-1}) = f(x_t|x_{t-1}), \quad g(y_t|x_{0:t}, y_{0:t-1}) = g(y_t|x_t)$$

Here the notation $x_{t_1:t_2}, t_2 \geq t_1$, is used to specify a vector of states, $[x_{t_1}\ x_{t_1+1}\ ...\ x_{t_2}]$. Note that we will use lower case notation for random variables and their realised values from now on, as there should not be ambiguity. Note that the form of the two densities $f(x...)$ and $g(y...)$ are generally designed to match the problem at hand as accurately as possible, based on theoretical analysis or empirical observation of datasets.

Having outlined the model, the inference task can now be specified. The canonical inference objective is termed *filtering*, not to be confiused with deterministic digital filtering. Instead, we pose a *statistical* filtering problem, which in its most general version determines the conditional distribution of the current state x_t, given all observations, i.e. $p(x_t|y_{0:t})$, which may be obtained directly in a sequential form from an adaptation of Bayes' Theorem:

$$p(x_t|y_{0:t}) = \frac{p(x_t|y_{0:t-1})g(y_t|x_{0:t})}{p(y_t|y_{0:t-1})} = \frac{p(x_t|y_{0:t-1})g(y_t|x_t)}{p(y_t|y_{0:t-1})} \tag{2}$$

This is often known as the *correction* step, preceded by the *prediction* step:

$$p(x_t|y_{0:t-1}) = \int f(x_t|x_{t-1})p(x_{t-1}|y_{0:t-1})dx_{t-1} \tag{3}$$

These equations were first obtained, independently in the 1930s, by Andrey Kolmogorov and Sydney Chapman, the celebrated Chapman-Kolmogorov equations for Markov processes. Sequential application of these two equations as time t increments is the basis for the *Bayesian filtering* task. Having determined the filtering density, $p(x_t|y_{0:t})$, estimates of x_t may be computed, for example the posterior mean $\hat{x}_t = \int_{x_t} x_t p(x_t|y_{0:t})dx_t$ or the Maximum *a posteriori* (MAP) estimate $x_t^{\text{MAP}} = \text{argmax}_{x_t} p(x_t|y_{0:t})$. Unfortunately only a few special cases can be solved in closed form. The first case is the classic Kalman Filter, which applies exactly when the state space model is linear and Gaussian. The resulting filtering density can be shown to be Gaussian, $p(x_t|y_{0:t}) = \mathcal{N}(x_t|m_t, P_t)$, although we do not give the details here. The Kalman filter does however play a substantial role in particle filtering methods, and more extensive surveys give details.

In order to gain a more advanced interpretation that will be useful for understanding particle filters we now introduce the sequential update to the so-called *smoothing* density $p(x_{0:t}|y_{0:t})$, a much more complex quantity because of the sequentially growing size of the vector $x_{0:t}$. The prediction-correction steps are however simpler to obtain as:

$$p(x_{0:t}|y_{0:t-1}) = f(x_t|x_{t-1})p(x_{0:t-1}|y_{0:t-1}) \tag{4}$$

$$p(x_{0:t}|y_{0:t}) = \frac{p(x_{0:t}|y_{0:t-1})g(y_t|x_t)}{p(y_t|y_{0:t-1})} \tag{5}$$

since they do not involve an integration, and we will find it helpful to make use of this version later.

3.2. Monte Carlo Methods

Now we have the basic schemes and 'in principle' methods for solving general filtering problems. However, for most applied scenarios the linear Gaussian assumptions of the Kalman filter are not realistic and numerical procedures must be employed.

In this section we outline three key developments during this period that led to all the right tools being available for a particle filter: the basic Monte Carlo (MC) method, the importance sampling (IS) method, and the sampling-importance-resampling (SIR) method.

Monte Carlo methods originated in the physics community, with scientists such as Enrico Fermi experimenting in the 1930s and then others during the 1940s and 50s such as von Neuman, Ulam and Metropolis. The basic Monte Carlo method attempts to calculate expectations with respect to intractable probability densities. Suppose we wish to calculate the expectation $\hat{\phi}(x) = E[\phi(x)] = \int \phi(x)p(x)dx$ where $p(x)$ is the probability density function in question and $\phi(x)$ is some measurable function. Basic Monte Carlo simply generates a large number of independent samples $x^{(i)} \sim p(x)$, $i = 1, ..., I$ and uses these to approximate the expectation. Here it is useful to introduce the *empirical* approximation of $p(x)$. If we have samples $x^{(i)} \sim p(x)$, $i = 1, ..., I$ then informally $p(x)$ can be approximated as:

$$p(x) \approx \frac{1}{I}\sum_{i=1}^{I} \delta_{x^{(i)}}(x)$$

where $\delta_{x^{(i)}}(x)$ is the Dirac function, i.e. an atom with unity volume at $x = x^{(i)}$. Of course this summation does not really approximate the value of the density function in any meaningful way. However, it does approximate expectations wrt measurable functions since the resulting expectation

$$\hat{\phi}(x) \approx \int \phi(x)\left(\frac{1}{I}\sum_{i=1}^{I}\delta_{x^{(i)}}(x)\right)dx = \frac{1}{I}\sum_{i=1}^{I}\phi(x^{(i)})$$

can be shown to be unbiased and reducing in variance as $1/I$. Although this is a slow rate of convergence, it does show that the method has good properties for large I.

The basic Monte Carlo scheme does not however solve the sequential updating problem required for Bayesian updating. The next step in developing a particle filter is the *importance sampling* step. Importance sampling was invented certainly by the early 1950s by Kahn and Harris [14], although they imply it was already known to statisticians in their writings. Suppose

[2] In fact, much less restrictive assumptions are possible for Bayesian updating, and especially for particle filtering, but the Markovian asumption is a simple and commonly used starting point.

now that we cannot generate from $p(x)$ itself, but rather we can generate from another density function $q(x)$ which may be more tractable. Monte Carlo samples from $q()$ now yield the approximation $q(x) \approx \frac{1}{I}\sum_{i=1}^{I}\delta_{x^{(i)}}(x)$, and the expectation may be rewritten as $\int \phi(x)p(x)dx = \int \phi(x)\frac{p(x)}{q(x)}q(x)dx$, leading to the modified *importance sampling* estimate:

$$\int \phi(x)\frac{p(x)}{q(x)}\left(\frac{1}{I}\sum_{i=1}^{I}\delta_{x^{(i)}}(x)\right)dx = \frac{1}{I}\sum_{i=1}^{I}\frac{p(x^{(i)})}{q(x^{(i)})}\phi(x^{(i)})$$

where $w(x) = \frac{p(x)}{q(x)}$ is known as the *importance weight*.

This method was widely used over the decades for sampling of rare events in physics and communications systems. In Bayesian statistics it was popularised in the late 1980s by John Geweke [15] as part of the computational revolution in Bayesian methods. Now, to see how to carry out Bayesian updating with importance sampling, take $q()$ to be the prior distribution, $p(x)$, in Bayes' Theorem Eq. (1). The target distribution $p()$ will be the posterior density $p(x|y)$. If we generate many samples from the prior $p(x)$ the importance weight will be, from Eq. (1):

$$w(x) = \frac{p(x|y)}{p(x)} = \frac{\frac{p(y|x)p(x)}{p(y)}}{p(x)} = \frac{p(y|x)}{p(y)}$$

One extra item required for many Bayesian updating problems is that we are often unable to compute exactly the term $p(y)$, the *total*, or *marginal* probability of y. It turns out that we can approximate this also by Importance sampling using the same collection of samples from $p(x)$:

$$p(y) = \int p(y|x)p(x)dx \approx \frac{1}{I}\sum_{i=1}^{I}p(y|x^{(i)})$$

and the resulting empirical approximation of the posterior density is

$$p(x|y) \approx \sum_{i=1}^{I}\tilde{w}^{(i)}\delta_{x^{(i)}}(x), \quad \tilde{w}^{(i)} = \frac{p(y|x^{(i)})}{\sum_{i'=1}^{I}p(y|x^{(i')})}$$

where $\tilde{w}^{(i)}$ is the so-called *self-normalised* importance weight.

One final step was required to have in place everything required for the particle filter. This final item is the resampling step, and this was proposed in the Bayesian setting by Donald Rubin [16] and popularised by Adrian Smith [17]. We can now take the *self-normalised* importance weights $\tilde{w}$ and use them to *resample* I samples with replacement from the original sample set $\{x^{(i)}\}$, which we will now refer to as *particles*. The weights disappear, or more correctly, are all set equal to $1/I$. The resulting algorithm is the *Sampling-Importance-Resampling* (SIR) procedure, and it was one of the enablers of the practical implementation of the particle filter.

It can be argued that SIR is a retrograde step from updating using a weighted IS representation without resampling, since it introduces an extra step of randomness which will make the Monte Carlo (finite I) error larger on average. However, it proves to be a key element when IS is repeatedly applied, for example in the multiple sequential time update steps of the particle filter.

Now, placing the SIR algorithm within the setting of a state-space model, we finally have all the required ingredients to reach 1993 and present the bootstrap filter...

4. 1993-2018: THE FIRST 25 YEARS

Bootstrap Filter [1]. In order to apply SIR to the sequential state space setting we take the *prediction* step as the proposal distribution $q()$, and the *corection* step is then implemented with SIR. Conceptually the simplest way to see this is through the joint prediction/correction step, Eq. (4)-(5). The prediction step is implemented as a joint random draw from the prediction density Eq. (4). The key innovation here is that, just like the Kalman filter, we will assume that the problem has been *solved* up to time $t-1$, which in a Monte Carlo scheme means that we have a weighted set of Monte Carlo samples $\{x_{0:t-1}^{(i)}\}$ drawn from the smoothing density up to time $t-1$:

$$p(x_{0:t-1}|y_{0:t-1}) = \sum_{i=1}^{I}\tilde{w}_{t-1}^{(i)}\delta_{x_{0:t-1}^{(i)}}(x_{0:t-1}) \tag{6}$$

We can easily make a random draw from this discrete distribution by choosing one of the stored sequences $x_{0:t-1}^{(i)}$ with probability $\tilde{w}_{t-1}^{(i)}$. I independent random draws from the same discrete distribution correspond to a random sample of size I with replacement from the population of particles $\{x_{0:t-1}^{(i)}\}$ (the 'resampling' step - note that this happens first rather than last in this presentation) - call these the *resampled* sequences $\{\tilde{x}_{0:t-1}^{(i)}\}$. The reason why this step is important in the sequential setting is that it tends to replicate high-weighted sequences many times and to under-represent very low weighted samples. Hence we will tend not to waste computation updating paths that have very low chance of success, while we expend most of our computational effort on high-weighted sequences. The prediction step is completed by making a random sample from the term $x_t^{(i)} \sim f(x_t|\tilde{x}_{t-1}^{(i)})$ for each of the resampled sequences $\tilde{x}_{0:t-1}^{(j)}$. Taken together, the resampled sequences and extended state values $x_{0:t}^{(i)} = [x_t^{(i)}\ \tilde{x}_{0:t-1}^{(i)}]$ form a Monte Carlo approximation of the prediction density required in Eq. (4):

$$p(x_{0:t}|y_{0:t-1}) \approx \frac{1}{I}\sum_{i=0}^{I}\delta_{x_{0:t}^{(i)}}(x_{0:t})$$

The approximation arises because we are only taking a finite number of samples I to make the representation. This is now treated as the proposal distribution to an importance sampler with target distribution $p(x_{0:t}|y_{0:t})$. The IS weight is, by Eq. (5):

$$w_t = \frac{p(x_{0:t}|y_{0:t})}{p(x_{0:t}|y_{0:t-1})} = \frac{g(y_t|x_t)}{p(y_t|y_{0:t-1})}$$

and, since $p(y_t|y_{0:t-1})$ is constant but unknown in general, we take the self-normalised version of IS to get:

$$p(x_{0:t}|y_{0:t}) \approx \sum_{i=0}^{I}\tilde{w}_t^{(i)}\delta_{x_{0:t}^{(i)}}(x_{0:t})$$

where $\tilde{w}_t^{(i)} = \frac{g(y_t|x_t^{(i)})}{\sum_j g(y_t|x_t^{(j)})}$. The same procedure can be iterated now at times $t+1$, $t+2$, ... making a direct Monte Carlo analogue of the Kalman filtering equations. This method is precisely the *bootstrap* filter of Gordon, Salmond and Smith [1]. Just as with the Kalman filter, its origins are quite complex and we should cite at least the near contemporary developments by Genshiro Kitagawa [18]. There are also similar algorithms proposed in the physics literature from earlier years. It is worth noting that some sequential IS algorithms *without* the crucial resampling step appeared in the control literature from the late 60s, see [19, 20]. The bootstrap algorithm has shown great success in solving hard nonlinear/ non-Gaussian problems from many application domains. However, it is limited in a number of ways, and we now give a whistle-stop tour of some of the developments from 1993 to now.

Revisiting the update. First let us return to the joint prediction-correction updates Eq. (4)-(5). Some insights can be gained from writing these as a single equation and substituting the approximation from Eq. (6), which leads after a little rearrangement to:

$$p(x_{0:t}|y_{0:t}) = \frac{\sum_{i=0}^{I} f(x_t|x_{t-1}^{(i)}) g(y_t|x_t^{(i)}) \tilde{w}_{t-1}^{(i)} \delta_{x_{0:t-1}^{(i)}}(x_{0:t-1})}{p(y_t|y_{0:t-1})} \tag{7}$$

This equation now leads to many of the important variants we have seen developing over recent years. The first observation is that this updated joint distribution is a combination of a discrete distribution over past sequences $x_{0:t-1}$ taking possible values only in the set $\{x_{0:t-1}^{(i)}\}$, and a conditional distribution for x_t, as follows:

$$\Pr(x_{0:t-1} = x_{0:t-1}^{(i)}|y_{0:t}) = \frac{\tilde{w}_{t-1}^{(i)} \int f(x_t|x_{t-1}^{(i)}) g(y_t|x_t^{(i)}) dx_t}{p(y_t|y_{0:t-1})} \tag{8}$$

$$p(x_t|x_{0:t-1}, y_{0:t}) = \frac{f(x_t|x_{t-1}) g(y_t|x_t)}{p(y_t|x_{t-1})} \tag{9}$$

This means that we have a target distribution over the entire sequence $x_{0:t}$ which may be sampled using IS with an essentially arbitrary discrete proposal function for $x_{0:t-1}$, say:

$$q(x_{0:t-1}|y_{0:t}) = \sum_{i=0}^{I} v_{t-1}^{(i)} \delta_{x_{0:t-1}^{(i)}}(x_{0:t-1}), \;\; v_{t-1}^{(i)} \geq 0,$$

where $v_{t-1}^{(i)}$ (satisfying $\sum_i v_{t-1}^{(i)} = 1$) are carefully chosen proposal weights; and a conditional proposal for a new state x_t, $q(x_t|x_{0:t-1}, y_{0:t})$, noting that both proposals can be designed to depend on the entire available data, $y_{0:t}$ in this case. The self-normalised importance weights for this general framework are then

$$\tilde{w}_t \propto \frac{\tilde{w}_{t-1}}{v_{t-1}} \frac{f(x_t|x_{t-1}) g(y_t|x_t)}{q(x_t|x_{0:t-1}, y_{0:t})}, \;\; \sum \tilde{w}_t^{(i)} = 1$$

Setting $v_{t-1} = \tilde{w}_{t-1}$ and $q(x_t|x_{0:t-1}, y_{0:t}) = f(x_t|x_{t-1})$ gets us precisely back to the bootstrap filter.

General Importance Sampling Filter [3, 21]. However, alternative choices are possible for both elements of the proposal and can lead to substantial improvements compared to the 'vanilla' bootstrap version. Setting $v_{t-1} = \tilde{w}_{t-1}$ but choosing $q(x_t|x_{0:t-1}, y_{0:t})$ in some other carefully chosen way leads to the general importance sampling framework. In the 'optimal' case we can set $q(x_t|x_{0:t-1}, y_{0:t}) = p(x_t|x_{0:t-1}, y_{0:t})$ and then the weights do not depend on the value of x_t at all: $\tilde{w}_t \propto p(y_t|x_{t-1}), \;\; \sum \tilde{w}_t^{(i)} = 1$.

While such a choice has good properties, it is used practically more as a guide to design parameters, since it is rarely possible both to sample x_t in this way and compute the required weight.

Auxiliary Particle Filter (APF) [22]. Meanwhile, choosing alternative forms for $q(x_{0:t-1}|y_{0:t})$ leads to the class of methods known as Auxiliary Particle Filtering. In particular it is proposed there to make the proposal weights v_{t-1} depend upon the new observation y_t. In the so-called 'fully adapted' case of the APF, v_{t-1} is chosen proportional to $\tilde{w}_{t-1} p(y_t|x_{t-1})$ and $q(x_t|x_{0:t-1}, y_{0:t}) = p(x_t|x_{0:t-1}, y_{0:t})$ (as in the 'optimal' proposal), which leads to a uniformly weighted sample set with $\tilde{w}_t = 1/I$. Again, this 'ideal' case is only usually possible as a guide for more practical proposal designs. See [23] for further discussion of these approaches.

Sequential MCMC. The general update equation (7) allows for additional insights and further methods. One which is of current interest for challenging scenarios takes the Monte Carlo update Eq. (7) and applies a different style of Monte Carlo. The Gibbs sampling MCMC framework allows the problem to be split into smaller individual conditional sampling steps, with convergence to the joint target after many such steps, see [24, 25]. A simple way to split the conditionals is as follows:

$$p(x_{0:t-1}|y_{0:t}, x_t) \propto \sum_{i=0}^{I} f(x_t|x_{t-1}^{(i)}) \tilde{w}_{t-1}^{(i)} \delta_{x_{0:t-1}^{(i)}}(x_{0:t-1}) \tag{10}$$

which again involves drawing from a simple discrete distribution with probability masses proportional to $f(x_t|x_{t-1}^{(i)}) \tilde{w}_{t-1}^{(i)}$, and a conditional sampling step for the new state exactly as in Eq. (9). This latter step will typically be split into several sub-conditionals, and implemented using Metropolis-Hastings (MH) where necessary. Typically an additional joint MH sampling mechanism targetting (7) directly is included to aid mobility of the sampler around the state-space. These methods have appeared in various guises since the late 90s [26, 22, 27, 28, 29] and are one of the promising directions for challenging scenarios currently being investigated.

Particle Smoothing [30]. It turns out that the conditional distribution also facilitates a backward smoothing procedure: first draw a sample randomly from the discrete particle filtering representation of $p(x_{0:t}|y_{0:t})$ and keep just the last time point x_t, then repeateadly draw from the conditional density Eq. (10) from $t-1, t-2, ...\ 0$. The set of random samples returns an improved sequence random draw from the entire smoothing density $p(x_{0:t}|y_{0:t})$, as compared with the initial draw. See also the related marginal smoothing methods given by [3, 2].

Other Developments and the Future. There have been many important developments to the methods over the years and here we can only name a few. We stick with inference for dynamical models, but we note that Sequential Monte Carlo methods have been developed for much more general applications including static parameter estimation and as a replacement for other, typically MCMC-based methods in batch scenarios, see especially the SMC Sampler work of [31]. A fundamental advance that occurred fairly early was the Rao-Blackwellised particle filter[3, 32], in which the power and simplicity of the Kalman filter is used to compute any linear/ Gaussian component of the state-space model, while the particle filter is used to compute the remaining non-Gaussian/ nonlinear components. An advance which proved particularly successful was the addition of MCMC steps to a standard IS-based particle filter, see the Resample-Move approach of [33], not to be confused with the Sequential MCMC approaches described above. The intuitive idea there is that any sampled sequence $x_{0:t}^{(i)}$ from a particle filter is nominally drawn from the correct distribution $p(x_{0:t}|y_{0:t})$ (subject to the finite sample-size error). Any MCMC kernel targetting $p(x_{0:t}|y_{0:t})$ can be applied to this sampled sequence without changing that distribution, and in fact a suitable kernel will both move particles closer to the true posterior density, and also increase diversity amongst different particles to which MCMC is applied. A more recent development that has sparked a new field of research in its own right is the Particle MCMC approach of [34]. Again, not to be confused with Sequential MCMC or Resample-Move schemes above, Particle MCMC embeds particle filters within (typically) intractable MCMC methods, and a remarkable result is proven that shows the resulting samplers are still precisely valid as MCMC procedures. Finally we note that effort is being made to parallelise and to extend particle filters to much larger state spaces, and in recent years more heuristic approaches that split the state-space into smaller chunks have some validity in terms of their approximation errors, see e.g. [35].

5. REFERENCES

[1] N. Gordon, D. Salmond, and A. F. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proc. F, Radar Signal Process.*, vol. 140, pp. 107–113, 1993.

[2] M. Hürzeler and H. R. Künsch, "Monte Carlo approximations for general state-space models," *J. Comput. Graph. Statist.*, vol. 7, pp. 175–193, 1998.

[3] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte-Carlo sampling methods for Bayesian filtering," *Stat. Comput.*, vol. 10, pp. 197–208, 2000.

[4] A. Doucet, N. De Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, Springer, New York, 2001.

[5] P.M. Djuric ; J.H. Kotecha ; Jianqui Zhang ; Yufei Huang ; T. Ghirmai ; M.F. Bugallo ; J. Miguez, "Particle filtering," *IEEE Signal Processing Magazine*, vol. 20, no. 5, 2003.

[6] O. Cappé, S.J. Godsill, and E.Moulines, "An overview of existing methods and recent advances in Sequential Monte Carlo," *Proc. IEEE*, vol. 95, no. 5, May 2007.

[7] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fiteen years later," *Oxford Handbook of Nonlinear Filtering, OUP, D. Crisan and B. Rozovsky (eds.)*, 2011.

[8] F. Gustafsson, "Particle filter theory and practice with positioning applications," *IEEE Aerospace and Electronic Systems Magazine*, vol. 10, no. 7, 2010.

[9] M.S. Arulampalam, S. Maskell, N.J. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[10] B. Ristic, M. Arulampalam, and A. Gordon, *Beyond Kalman Filters: Particle Filters for Target Tracking*, Artech House, 2004.

[11] J.S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer, New York, 2001.

[12] T. Bayes, "An essay towards solving a problem in the doctrine of chances," *Phil. Trans. of the Royal Soc. of London*, vol. 53, pp. 370–418, 1763.

[13] P. S. Laplace, *Essai philosophique sur les probabilities*, Courcier, Paris, 1814, Reprinted (1812) in English, F.W. Truscott amd F. L. Emory (Trans.) by Wiley, New York.

[14] H. Kahn and T. E. Harris, "Estimation of particle transmission by random sampling," *National Bureau of Standards applied mathematics series*, , no. 12, 1951.

[15] John Geweke, "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica*, vol. 57, no. 6, pp. pp. 1317–1339, 1989.

[16] D. B. Rubin, "A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when the fraction of missing information is modest: the SIR algorithm (discussion of Tanner and Wong)," *J. Am. Statist. Assoc.*, vol. 82, pp. 543–546, 1987.

[17] Adrian FM Smith and Alan E Gelfand, "Bayesian statistics without tears: a sampling–resampling perspective," *The American Statistician*, vol. 46, no. 2, pp. 84–88, 1992.

[18] G. Kitagawa, "A Monte Carlo filtering and smoothing method for non-Gaussian nonlinear state space models," *Proceedings of the 2nd U.S.-Japan Joint Seminar on Statistical Time Series Analysis*, pp. 110–131, 1993.

[19] J. Handschin and D. Mayne, "Monte Carlo techniques to estimate the conditionnal expectation in multi-stage non-linear filtering," in *Int. J. Control*, 1969, vol. 9, pp. 547–559.

[20] H. Akashi and H. Kumamoto, "Random sampling approach to state estimation in switching environment," *Automatica*, vol. 13, pp. 429–434, 1977.

[21] J. S. Liu and R. Chen, "Sequential monte carlo methods for dynamic systems," *Journal of the American Statistical Association*, vol. 93, no. 443, pp. 1032–1044, 1998.

[22] M. K. Pitt and N. Shephard, "Filtering via simulation: Auxiliary particle filters," *J. Am. Statist. Assoc.*, vol. 94, no. 446, pp. 590–599, 1999.

[23] S. J. Godsill and T. C. Clapp, "Improvement strategies for Monte Carlo particle filters," in *Sequential Monte Carlo Methods in Practice*, A Doucet, J. F. G. De Freitas, and N. J. Gordon, Eds. New York: Springer-Verlag, 2001.

[24] W.R. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC Interdisciplinary Statistics. Taylor & Fran(RM)cis, 1995.

[25] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, *Handbook of Markov Chain Monte Carlo*, CRC press, 2011.

[26] C. Berzuini, N. G. Best, W. R. Gilks, and C. Larizza, "Dynamic conditional independence models and markov chain monte carlo methods," *Journal of the American Statistical Association*, vol. 92, no. 440, pp. 1403–1412, 1997.

[27] A. Golightly and D. J. Wilkinson, "Bayesian sequential inference for nonlinear multivariate diffusions," *Statistics and Computing*, vol. 16, no. 4, pp. 323–338, Dec 2006.

[28] S.-K. Pang, J. Li, and S.J. Godsill, "Detection and tracking of coordinated groups," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 47, no. 1, pp. 472–502, January 2011.

[29] F. Septier and G. W. Peters, "Langevin and Hamiltonian based Sequential MCMC for Efficient Bayesian Filtering in High-dimensional Spaces," *IEEE Journal of Selected Topics in Signal Processing*, 2016.

[30] S. J. Godsill, A. Doucet, and M. West, "Monte Carlo smoothing for non-linear time series," *J. Am. Statist. Assoc.*, vol. 99, no. 465, pp. 156–168, 2004.

[31] P. Del Moral, A. Doucet, and A. Jasra, "Sequential Monte Carlo samplers," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 68, no. 3, pp. 411–436, 2006.

[32] R. Chen and J. S. Liu, "Mixture Kalman filters," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 62, no. 3, pp. 493–508, 2002.

[33] W. R. Gilks and C. Berzuini, "Following a moving target—Monte Carlo inference for dynamic Bayesian models," *J. Roy. Statist. Soc. Ser. B*, vol. 63, no. 1, pp. 127–146, 2001.

[34] C. Andrieu, A. Doucet, and R. Holenstein, "Particle markov chain Monte Carlo methods," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 72, no. 3, pp. 269–342, 2010.

[35] P. Rebeschini and R. van Handel, "Can local particle filters beat the curse of dimensionality?," *Ann. Appl. Probab.*, vol. 25, no. 5, pp. 2809–2866, 10 2015.