

RISK SENSITIVE PARTICLE FILTERS FOR MITIGATING SAMPLE IMPOVERISHMENT

Umut Orguner and Fredrik Gustafsson

Department of Electrical Engineering
Linköping University
SE-581 83 Linköping, Sweden
{umut, fredrik}@isy.liu.se

ABSTRACT

Risk-sensitive filters (RSF) put a penalty to higher-order moments of the estimation error compared to conventional filters as the Kalman filter minimizing the mean square error. The result is a more cautious filter, which can be interpreted as an implicit and automatic way to increase the state noise covariance. On the other hand, the process of jittering, or roughening, is well-known in particle filters to mitigate sample impoverishment. The purpose of this contribution is to introduce risk-sensitive particle filters (RSPF) as an alternative approach to mitigate sample impoverishment based on constructing explicit risk functions from a general class of factorizable functions.

Index Terms— Risk sensitive, particle filter

1. INTRODUCTION

Recursive implementations of Monte Carlo based statistical signal processing [1] are known as *particle filters*, see [2, 3]. The research has since the paper [4] steadily intensified, see the article collection [2], the survey [5], and the monograph [6]. The particle filters may be a serious alternative for real-time applications classically approached by the (extended) Kalman filter. The more non-linear model, or the more non-Gaussian noise, the more potential particle filters have, especially in applications where computational power is rather cheap and the sampling rate is moderate.

The basic particle filter approach suffers from the problem of *sample degeneracy*, also known as *depletion* or *sample impoverishment*. This means that after a while all particles but a few ones will have negligible weights. By introducing an additional noise to the samples the depletion problem can be reduced. This technique is called *jittering* in [7], but a similar approach was introduced in [4] under the name *roughening*. In [8], the problem is handled by introducing an additional Markov Chain Monte Carlo step to separate the samples.

In [4], the so-called *prior editing* method is discussed. The estimation problem is delayed one time-step, so that the likelihood can be evaluated at the next time step. The idea

is to reject particles with sufficiently small likelihood values, since they are not likely to be re-sampled. The *auxiliary particle filter* [9] is constructed in such a way that particles with large predictive likelihoods are simulated directly.

Another technique is *regularization*. The basic idea is to convolve each particle with a diffusion kernel before resampling. This will prevent multiple copies of a few particles. One may for instance use a Gaussian kernel where the variance acts as the bandwidth. The problem is that this variance will increase the variance of the posterior distribution.

Recently, a risk-sensitive particle filter has been proposed in [10] using practical arguments and a specific risk function which measures how risky not tracking a particular area of state space is. We must note that although the name "risk-sensitive particle filter" is a correct one since, in [10], the sampling density of the standard particle filter is modified so that more samples are generated in the "risky" regions of the state-space, this interpretation is not the same as the so called "risk-sensitive estimation" in especially control theory and signal processing where one minimizes the expected exponential of (cumulative quadratic) estimation error.

In this study, we propose a general theoretical framework yielding particle filters which minimizes a product cost function and we call the resulting filters as *product particle filters*. Under this framework, the particle filter of [10], which we will call as Thrun's filter from this point on, can be shown to minimize a product cost and the risk-sensitive particle filters (as understood in the context of control theory and as described above) are a special case. In order to derive the product particle filters, the reference probability method [11] is used. Our approach for obtaining the recursively calculated information state can be considered as a combination and/or generalization of [12] where risk-sensitive filters are derived for nonlinear systems and [13] where product estimators are proposed for hidden Markov models.

2. RISK SENSITIVE FILTERING

We here explain the basic concepts of risk-sensitive filtering in terms of the Kalman filter for a linear system

$$x_{k+1} = Ax_k + w_k, \quad \text{Cov}(w_k) = Q_k, \quad (1a)$$

$$y_k = Cx_k + v_k, \quad \text{Cov}(v_k) = R_k. \quad (1b)$$

The authors gratefully acknowledge fundings from SSF (Swedish Foundation for Strategic Research) Strategic Research Center MOVIII and the Vinnova/FMV TAIS project ARCUS.

Defining the product form

$$\Upsilon_k(\xi) = \left[\prod_{l=0}^{k-1} \eta_l(\hat{x}_{0:l-1}, x_l, \hat{x}_{l|l}) \right] \eta_k(\hat{x}_{0:k-1}, x_k, \xi) \quad (2)$$

$$\eta_k(\hat{x}_{0:k-1}, x_k, \xi) = \exp \left\{ \frac{\theta}{2} (x_k - \xi)^T \Sigma_k (x_k - \xi) \right\}, \quad (3)$$

where $\hat{x}_{0:k} \triangleq \{\hat{x}_{0|0}, \hat{x}_{1|1}, \dots, \hat{x}_{k|k}\}$, we consider the problem $\hat{x}_{k|k} = \arg \min_{\xi} E[\Upsilon_k(\xi)|y_{0:k}]$.

Using a Taylor expansion of the cost function and letting the exponential parameter θ approach zero, it is realized that only the quadratic cost is minimized, which leads to the Kalman filter solution. That is, the Kalman filter is a special case of a risk-sensitive filter. When θ increases, higher order moments appear in the cost, and the resulting filter can be expected to become more robust, in particular to unlikely events in the tails of the involved noise distributions. On the other hand, if θ is chosen too large, the RS filter becomes unstable.

It can be shown that the solution gives a recursion that only modifies the time update of the covariance matrix in the Kalman filter algorithm [12],

$$P_{k+1|k} = A_k \left(P_{k|k}^{-1} - \theta \Sigma_k^{-1} \right)^{-1} A_k^T + Q_k \quad (4a)$$

$$= A_k P_{k|k} (I + \theta \Sigma_k^{-1} P_{k|k} + \theta^2 \Sigma_k^{-1} P_{k|k} \Sigma_k^{-1} P_{k|k} + \dots) A_k^T + Q_k. \quad (4b)$$

The latter series expansion holds if θ is small enough. This expression clearly shows that the exponential risk function can be interpreted as a roughening instrument in that it increases the covariance after the time update.

3. RISK-SENSITIVE NON-LINEAR FILTERING

We consider the following discrete-time nonlinear state space model defined on a probability space (Ω, \mathcal{F}, P)

$$\begin{aligned} x_{k+1} &= f(x_k) + w_{k+1} \\ y_k &= h(x_k) + v_k \end{aligned} \quad (5)$$

where $\{x_k \in \mathbb{R}^n\}$ is the state sequence with initial distribution $x_0 \sim p_0(x_0)$, $\{y_k \in \mathbb{R}^m\}$ is the noisy observation sequence, $\{w_k \in \mathbb{R}^n\}$ is a white process noise sequence with distribution $w_k \sim p_w(w_k)$, $\{v_k \in \mathbb{R}^m\}$ is a white measurement noise sequence independent from the process noise w_k with distribution $v_k \sim p_v(v_k)$. The functions $f(\cdot)$ and $h(\cdot)$ are measurable and nonlinear functions of the state x_k . The classical exponential form of the risk function can be motivated by the explicit algebra leading to (4). In a general nonlinear filtering framework, any function that punishes higher order moments can be used. We propose the following product form.

$$\Gamma_{0:k}(\zeta) = \hat{\Gamma}_{0:k-1} \gamma_k(\hat{x}_{0:k-1}, x_k, \zeta) \quad (6)$$

$$\hat{\Gamma}_{0:k} = \prod_{i=0}^k \gamma_i(\hat{x}_{0:i-1}, x_i, \hat{x}_{i|i}) \quad (7)$$

The aim is to calculate the state estimates $\hat{x}_{k|k}$ defined as

$$\hat{x}_{k|k} \triangleq \arg \min_{\zeta} E[\Gamma_{0:k}(\zeta)|y_{0:k}]. \quad (8)$$

If we work with a different probability measure \bar{P} under which the measurement sequence $\{y_k\}$ is independent identically distributed with density function $y_k \sim p_v(y_k)$ for all k , we see that

$$\hat{x}_k \triangleq \arg \min_{\zeta} \bar{E}[\bar{\Lambda}_k \Gamma_{0:k}(\zeta)|y_{0:k}]. \quad (9)$$

where $\bar{\Lambda}_k \triangleq \frac{p(y_k|x_k)}{p_v(y_k)}$, $\bar{\Lambda}_k \triangleq \prod_{l=0}^k \bar{\Lambda}_l$ and \bar{E} denotes the expectation operation under the probability measure \bar{P} . The recursive calculation of the expected value in (9) can be done using the unnormalized density function $\alpha_{k|k-1}(x)$ defined as

$$\alpha_{k|k-1}(x) dx \triangleq \bar{E}[\bar{\Lambda}_{k-1} \hat{\Gamma}_{0:k-1} \mathcal{I}_{\{x_k \in dx\}} | y_{0:k-1}] \quad (10)$$

where the function $\mathcal{I}_A(\omega)$ denotes the indicator function of the set A . Defining the normalized densities $\beta_{k|k-1}(x) \triangleq \alpha_k(x) / \int_{\mathbb{R}^n} \alpha_k(\xi) d\xi$ we can obtain the following lemma.

Lemma 1 *The normalized density $\beta_{k|k-1}(x)$ satisfies the following recursion.*

$$\begin{aligned} \beta_{k|k-1}(x_k) &= \frac{1}{c_k} \int_{\mathbb{R}^n} p(x_k|x_{k-1}) p(y_{k-1}|x_{k-1}) \\ &\quad \times \gamma_{k-1}(\hat{x}_{0:k-2}, x_{k-1}, \hat{x}_{k-1|k-1}) \beta_{k-1|k-2}(x_{k-1}) dx_{k-1} \end{aligned} \quad (11)$$

where c_k is a normalization factor.

Proof: Omitted in this draft version.

The density function $\beta_{k|k-1}(\cdot)$ (as well as $\alpha_{k|k-1}(\cdot)$) is actually an information state for the expectation in (9) and hence the estimate $\hat{x}_{k|k}$.

4. RISK-SENSITIVE PARTICLE FILTER

The infinite dimensional recursion can be used to generate a particle filter as follows: At each time step k , the information state $\beta_{k|k-1}(x_k)$ is approximated with N samples (particles) $x_k^{(i)}$ as $\beta_{k|k-1}(x_k) \approx \frac{1}{N} \sum_{i=1}^N \delta(x_k - x_k^{(i)})$ where $\delta(\cdot)$ is the delta-Dirac function. Substituting this approximation into the recursion (11), we obtain the following particle filter.

Algorithm 1

1. *Initialize the particles $\{x_0^{(i)}\}_{i=1}^N \sim p_0(x_0)$. Calculate the estimate $\hat{x}_{0|0}$. Set $k = 1$.*
2. *Measurement Update: Calculate importance weights $\{\pi_{k-1}^{(i)}\}_{i=1}^N$ as*

$$\pi_{k-1}^{(i)} = p(y_{k-1}|x_{k-1}^{(i)}) \gamma_{k-1}(\hat{x}_{0:k-2}, x_{k-1}^{(i)}, \hat{x}_{k-1|k-1})$$
and normalize $\bar{\pi}_{k-1}^{(i)} = \pi_{k-1}^{(i)} / \sum_{j=1}^N \pi_{k-1}^{(j)}$.

3. *Resampling: Draw N particles $\{x_{k-1|k-1}^{(i)}\}_{i=1}^N$ with replacement according to*

$$P\left(x_{k-1|k-1}^{(i)} = x_{k-1}^{(j)}\right) = \bar{\pi}_{k-1}^{(j)}. \quad (12)$$

4. *Time Update: Obtain predicted particles $\{x_k^{(i)}\}_{i=1}^N$ according to $x_k^{(i)} \sim p(x_k|x_{k-1|k-1}^{(i)})$.*
5. *Calculate the estimate $\hat{x}_{k|k}$. Set $k = k + 1$, go to step 2.*

The algorithm described above differs from the standard particle filter only at Step 2 where the importance weights are formed using both measurement likelihood and the risk function. The use of the risk function modifies the resampling step of the algorithm so that more samples will be generated from those parts of the state space for which the risk function $\gamma_k(\cdot)$ assigns more risk.

4.1. Generalizations

Once the problem is transferred into particle filtering domain, the following generalizations easily follow:

1. The product functions $\gamma_k(\cdot)$ appearing can be allowed to change from particle to particle i.e., the function can be $\gamma_k^{(i)}(\cdot)$.
2. The (previous) particles can be input to functions $\gamma_k(\cdot)$ i.e., we can add $x_{0:k-1}^{(1:N)}$ defined by

$$x_{0:k-1}^{(1:N)} \triangleq \{\{x_0^{(i)}\}_{i=1}^N, \{x_1^{(i)}\}_{i=1}^N, \dots, \{x_{k-1}^{(i)}\}_{i=1}^N\} \quad (13)$$

and $x_k^{(1:N)}$ defined as $x_k^{(1:N)} \triangleq \{x_k^{(i)}\}_{i=1}^N$ as additional inputs to functions $\gamma_k(\cdot)$ (or $\gamma_k^{(i)}(\cdot)$ considering the first generalization.).

4.2. Special Cases

Thrun's filter [10] is obtained by making the specific selection

$$\gamma_k^{(i)}(\dots) = \begin{cases} r(x) & k = 0 \\ \frac{r(x)}{r(x_{k-1}^{(i)})} & k > 0 \end{cases} \quad (14)$$

It then follows that the samples $\{x_{k|k}^{(i)}\}_{i=1}^N$ generated by the resampling step of (the generalized version of) Algorithm 1 are distributed according to $\frac{1}{\tau_k} r(x_k) p(x_k|y_{0:k})$ where τ_k is the normalization constant.

Moreover, a risk-sensitive particle filter is found if we make the risk function an exponential one as follows.

$$\gamma_k(\hat{x}_{0:k-1}, x_k, \hat{x}_{k|k}) = \exp\left\{\frac{\theta}{2}(x_k - \hat{x}_{k|k})^T M_k (x_k - \hat{x}_{k|k})\right\}$$

where $M_k > 0 \in \mathbb{R}^{n \times n}$ and $\theta > 0 \in \mathbb{R}$.

5. REFERENCES

- [1] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in practice*, Chapman & Hall, 1996.
- [2] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, Springer Verlag, 2001.
- [3] A. Doucet, S.J. Godsill, and C. Andrieu, "On sequential simulation-based methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.
- [4] N.J. Gordon, D.J. Salmond, and A.F.M. Smith, "A novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *IEE Proceedings on Radar and Signal Processing*, 1993, vol. 140, pp. 107–113.
- [5] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb. 2002.
- [6] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: particle filters for tracking applications*, Artech House, London, 2004.
- [7] P. Fearnhead, *Sequential Monte Carlo methods in filter theory*, Ph.D. thesis, University of Oxford, 1998.
- [8] A. Doucet, N.J. Gordon, and V. Krishnamurthy, "Particle Filters for State Estimation of Jump Markov Linear Systems," *IEEE Trans. on Signal Processing*, vol. 49, no. 3, pp. 613–624, March 2001.
- [9] M.K. Pitt and N. Shephard, "Filtering via simulation: Auxiliary particle filters," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 590–599, June 1999.
- [10] S. Thrun, J. Langford, and V. Verma, "Risk sensitive particle filters," in *Advances in Neural Information Processing Systems 14*. 2002, MIT Press.
- [11] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models: Estimation and Control*, Springer-Verlag, NY, USA, 1994.
- [12] S. Dey and J. B. Moore, "Risk-sensitive filtering and smoothing via reference probability methods," *IEEE Trans. Automat. Contr.*, vol. 42, no. 11, pp. 1587–1591, Nov. 1997.
- [13] V. R. Ramezani, *Product Estimators for Hidden Markov Models*, Ph.D. thesis, Department of Electrical Engineering, University of Maryland, College Park, MD 20742, USA, 2001.