APPLICATION OF A RESAMPLING SCHEME TO SOLVE THE DIVERGENCE IN THE PATHWISE FILTER

Alexsandro Machado Jacob [*,1] **Takashi Yoneyama** [*]

[*] *Divisão de Engenharia Eletrônica*
Instituto Tecnológico de Aeronáutica
São José dos Campos SP 12228-900, Brazil
e-mail: {ajacob, takashi}@ita.br

Abstract:
A Monte Carlo-based approach to filtering for nonlinear systems based on the *Pathwise* theory was proposed by M. H. A. Davis in 1981. The discrete-time Markov chain used to compute the solution of a Fokker-Planck equation whose coefficients were determined by the observed process was here replaced by the simulation of an equivalent stochastic differential equation in order to make the filter implementation more clear and with less computational cost. This paper shows that the Pathwise filter for an one-dimensional Ornstein-Uhlenbeck state process with saturation in the observation has an interesting characteristic of divergence in this estimates when the signal-to-noise ratio on the state equation is low. Rewriting the filtering solution in terms of observation-based weights, it is presented that the low performance of the filter can be preliminary explained by the sudden increase of the weight variances. To solve this problem, a resampling scheme using the effective number of particles was used to smooth or, at least, maintain the weight variance controlled.

Keywords: Nonlinear filtering, Pathwise filter, Monte Carlo approximation, Simulation of SDEs.

1. INTRODUCTION

The basic continuous-time nonlinear filtering problem consists of estimating a time-homogeneous Markov process $X = \{X_t;\, t \geq 0\}$ with known law, given $Y = \{Y_t;\, t \geq 0\}$ defined by

$$Y_t = \int_0^t h(X_s)ds + V_t, \quad 0 \leq t \leq T. \qquad (1)$$

The *signal process*, X, takes values in $\mathbb{R}^d$ and the *observation one*, Y, in $\mathbb{R}^m$, with $d, m \geq 1$. The process $\{V_t; t \geq 0\}$ is a standard m-dimensional Brownian motion, $h : \mathbb{R}^d \to \mathbb{R}^m$ is a bounded continuous function, and T is a fixed final time. It is assumed that X_0 is a random variable with law ξ and $Y_0 = 0$.

The classical filtering problem can be summarized as finding the the conditional distribution of X_t with respect to $\mathcal{F}_0^t$, that is,

$$\pi_t(\varphi) = \int_{\mathbb{R}^d} \varphi(x)\pi_t(dx) = \mathbb{E}[\varphi(X_t)|\mathcal{F}_0^t], \qquad (2)$$

[1] Supported by the The State of São Paulo Research Foundation (FAPESP) under Grant 02/10632-0.

where $\mathcal{F}_0^t$ is the filtration generated by Y up to time t, and $\varphi : \mathbb{R}^d \to \mathbb{R}^m$ is a bounded continuous function.

In accordance to Kallianpur and Striebel (1968), the classical filtering problem can be rewritten as

$$\pi_t(\varphi) = \frac{\hat{\mathbb{E}}[\varphi(X_t)\,\Lambda_t|\mathcal{F}_0^t]}{\hat{\mathbb{E}}[\Lambda_t|\mathcal{F}_0^t]} = \frac{\sigma_t(\varphi)}{\sigma_t(1)}, \tag{3}$$

where $\hat{\mathbb{E}}$ is the expectation with respect to the measure defined by Λ_t, and $\sigma_t(\cdot)$, the *unnormalized* representation of $\pi_t(\cdot)$, satisfies the Zakai (Zakai, 1969) and the Pathwise (Clark, 1978; Davis, 1981*b*; Pardoux, 1981; Davis, 1981*a*) equations.

According to Davis (1981*b*) and Pardoux (1981), the main result of the *Pathwise* theory of nonlinear filtering shows that is possible to compute, for each t, the conditional distribution (3) in terms of the solution of a Fokker-Planck equation whose coefficients depend on the observed sample path $\{y(s),\, s \geq 0\}$. Clark (1978) and Clark and Crisan (2005) discussed the robustness of the Pathwise filtering, or the continuity of the filter with respect to the observation process.

Based on ideas of Kushner (1977) of approximating the trajectories of a parabolic partial differential equation by a discrete-time Markov chain, Davis (1981*a*) thus proposed a Monte Carlo-based technique to calculate (3) via a large number of independent simulations or *particles*. For an adequate choice of spatial grid spacing and number of independent simulations, (O'Loghlen and Wright, 1982) and (Souza, 1992) affirmed that the *Pathwise* filter estimates are comparable to that obtained by the extended Kalman filter for a stable one-dimensional Ornstein-Uhlenbeck signal process with saturation in the observation.

However, the most efficient and widely applicable approach to solving stochastic differential equations (SDEs) seems to be the simulation of sample paths of time discrete approximations on digital computers (Kloeden and Platen, 1999). As implemented by Jacob *et al.* (2004), the idea was to use the structure of the SDE in a natural way, in contrast to the Kushner's Markov chain approach where the state variables were discrete. An advantage of considerable practical importance of this approach is that the computational costs such as time and memory required increase only polynomially with the dimension of the problem.

The initial idea of this paper was to study implementation methods of the *Pathwise* filter and the filter performance when resampling methods were applied aiming to get a better rate of convergence. However, preliminary tests using the scenario presented by (O'Loghlen and Wright, 1982) showed that for unfavorable signal-to-noise ratio (SNR) in the state noise, the nonlinear filter started diverging on its estimates. In this way, based on the behavior of the main variables of the filter dynamics, this paper aims to give initial explanations about the divergence in the estimates. Based on a defined coefficient, a possible solution for the problem was proposed using a resampling scheme into the simulated trajectories of the filter.

Next section shows the main equations of the *Pathwise* theory. The numerical implementation and a generalized algorithm of the filter are proposed in Section 3. Section 4 contains the results referring to the filter divergence and the effectiveness of the proposed resampling-based solution. Conclusions are in Section 5.

2. THE PATHWISE FILTER

Let $(\Omega, \mathcal{F}, \mathcal{F}_{t\geq 0}, P)$ be a filtered probability space on which the signal X is described by the following SDE

$$dX_t = b(t, X_t)\,dt + g(t, X_t)\,dW_t, \tag{4}$$

where $\{W_t; t \geq 0\}$ is a standard d-dimensional Brownian motion independent of V, and

$$dY_t = h(X_t)\,dt + dV_t \tag{5}$$

is the differential form of equation (1) corresponding to the noisy observation, where

$$||h|| = \max_{1\leq i\leq m} \sup_{x\in\mathbb{R}^m} |h_i(x)|. \tag{6}$$

The functions $b : \mathbb{R}^+ \text{x} \mathbb{R}^d \to \mathbb{R}^d$ and $g : \mathbb{R}^+ \text{x} \mathbb{R}^d \to \mathbb{R}^{d^2}$ are globally Lipschitz and $X_0 \sim \xi$ is a d-dimensional square integrable random vector, $\mathcal{F}_0$-measurable and independent of W and V. These hypotheses satisfy the sufficient conditions for the existence and uniqueness of the solution of equation (4) (Lipster and Shirayaev, 1977).

Based on the concepts of the Pathwise filter presented in (Clark, 1978; Pardoux, 1981; Davis, 1981*b*), Davis (1981*a*) used some ideas described in (Kushner, 1977) to develop the robust Monte Carlo-based nonlinear filter. By fixing a continuous function $\{y(s),\, s \geq 0\}$, or *sample path* of the observed process Y, it can be shown that, given $\mathcal{F}_0^t$ and any continuous bounded function φ, the main result is

$$\mathbb{E}[\varphi(X_t)|\mathcal{F}_0^t] = \int\limits_{\mathbb{R}^d} \varphi(x) p^y(t, x) dx, \tag{7}$$

where

$$p^y(t,x) = \frac{exp\{y(t)h(x)\}q^y(t,x)}{\int_{\mathbb{R}^d} exp\{y(t)h(x')\}q^y(t,x')dx'} \quad (8)$$

and $q^y(s,x)$ is the solution of the following parabolic partial differential equation

$$-\frac{\partial q^y}{\partial s} + (L_s^y)^* q^y + c^y q^y = 0, \; s < t,$$
$$q^y(0,x) = p_0(x), \; s = 0, \quad (9)$$

which is solved forward from the initial condition $p_0(x)$ and with $(\cdot)^*$ being the conjugated operator. The parameters presented in (9) are given by

$$L_s^y\varphi = \frac{1}{2}\sum_{i,j} a_{ij}\frac{\partial^2\varphi}{\partial x_i \partial x_j} + \sum_i b_i^y(s,x)\frac{\partial \varphi}{\partial x_i} \quad (10)$$

which is the infinitesimal generator of a process Z_s whose solution is the next SDE

$$dZ_s = b^y(s, Z_s)\,ds + g(s, Z_s)\,dB_s, \quad (11)$$

where B_t is a standard vector Brownian-motion process with the same dimension and independent of V,

$$c^y(s,x) = \frac{1}{2}y^2(s)\sum_{i,j} a_{ij}(s,x)\frac{\partial h}{\partial x_i}\frac{\partial h}{\partial x_j}$$
$$-y(s)L_s h(x) - \frac{1}{2}h^2(x), \quad (12)$$

$$L_s\varphi = \frac{1}{2}\sum_{i,j} a_{ij}\frac{\partial^2\varphi}{\partial x_i \partial x_j} + \sum_i b_i(s,x)\frac{\partial \varphi}{\partial x_i}, \quad (13)$$

which is the infinitesimal generator of the signal process (4),

$$a_{ij} = a_{ij}(s,x) = g(s,x)g^*(s,x) \quad (14)$$

and

$$b_i^y(s,x) = b_i(s,x) - y(s)\sum_j a_{ij}(s,x)\frac{\partial h}{\partial x_j}. \quad (15)$$

Therefore, the nonlinear filtering consists in computing the conditional density (3) by solving equation (9), which is an equation of almost the same type as the Fokker-Planck equation whose solution for the signal realization x is the *unnormalised conditional density* given by $exp\{y(t)h(x)\}q^y(t,x)$ in (8).

3. NUMERICAL IMPLEMENTATION

Consider the time partition

$$0 = \mathcal{T}_0 < \mathcal{T}_1 < \cdots < \mathcal{T} < \cdots < \mathcal{T}_N = T \quad (16)$$

with

$$\mathcal{T}_{j+1} = \mathcal{T}_j + \delta, \;\; j = 0, 1, ..., (N-1), \quad (17)$$

and time-increment given by $\delta = \frac{T}{N}$.

3.1 Davis' Monte Carlo Filter

The parabolic partial differential equation (9) is the forward equation for a process (11) perturbed by a Feynman-Kac formula transformation. To have a clear understanding about the implementation of the filter, equation (7) can be rewritten as

$$\int_{\mathbb{R}^d} \tilde{\varphi}_i(\mathcal{T}, x)q^y(\mathcal{T}, x)dx = \mathbb{E}[K(\mathcal{T})\,\tilde{\varphi}_i(\mathcal{T}, z_{\mathcal{T}})], (18)$$

where

$$K(\mathcal{T}) = exp\{\int_0^{\mathcal{T}} c^y(r, z_r)dr\,\} \quad (19)$$

is the *modulation coefficient*, and $\tilde{\varphi}_i$, $i = 1, 2$, is described as

$$\tilde{\varphi}_1(\mathcal{T}, z) = exp\{y(\mathcal{T})h(z)\} \quad (20)$$

and

$$\tilde{\varphi}_2(\mathcal{T}, z) = exp\{y(\mathcal{T})h(z)\}\varphi(z), \quad (21)$$

in accordance to (7) and (8).

The classical Monte Carlo method consists of approximating the expectation in equation (18) by an average of P realizations or *particles.* Thus, generating P independent sample trajectories $\{z_s^k, 0 \le s \le \mathcal{T}\}$, $k = 1, 2, ..., P$, of the $Z_{\mathcal{T}}$ process of equation (11), the values

$$K_k(\mathcal{T}) = exp\{\int_0^{\mathcal{T}} c^y(r, z_r^k)dr\,\}, \quad (22)$$

must be calculated and the approximated value of the right-hand side of equation (18) is

$$\frac{1}{P}\sum_{k=1}^{P} K_k(\mathcal{T})\tilde{\varphi}_i(z_{\mathcal{T}}^k). \quad (23)$$

Finally, combining (23) with (11) into functions (20) and (21), an approximated estimate of $\pi_t(\varphi)$ is given by

$$\mathbb{E}[\varphi(x_{\mathcal{T}})|\mathcal{F}_0^{\mathcal{T}}] \simeq \sum_{k=1}^{P} \mu_k(\mathcal{T})\tilde{\varphi}_2(\mathcal{T}, z_{\mathcal{T}}^k), \quad (24)$$

where

$$\mu_k(\mathcal{T}) = \frac{K_k(\mathcal{T})}{\sum_{k=1}^{P} K_k(\mathcal{T})\tilde{\varphi}_1(\mathcal{T}, z_{\mathcal{T}}^k)} \quad (25)$$

is the *normalized weight.*

As quoted by Davis (1981*a*), the influence of the observed sample path here is twofold: first, the value of $y(\mathcal{T})$ appears explicitly on equation (24), and, secondly, the sample path $\{y(s), 0 \leq s \leq \mathcal{T}\}$ determines the generator of $z_{\mathcal{T}}$ via (15) and (12) and hence affects the distributions of $K_k(\mathcal{T})$.

This approach is feasible in the sense that one carries it out and get a return directly related to the amount of computational effort invested. However, the convergence could still be quite slow due to bad initializations of $Z_{\mathcal{T}}$, as explained by O'Loghlen and Wright (1982) and Souza (1992).

3.2 Davis' Generalized Monte Carlo Filter

Based on ideas of Del Moral and Miclo (2000), a resampling scheme of the particles must be implemented in order to kill trajectories exploring unfruitful directions of the process. This can be interesting because, according to Le Gland (1984), K_k has a sharp maximum given a certain realization. Thus, just simulations next to the referred realization contribute effectively for the estimate.

A good criterion used to apply the resampling in a given time instant $\mathcal{T}_j$ is the *effective sample size* (Doucet, 1998). It measures the degeneracy of the particles and is defined as

$$N_{eff} = \frac{1}{\sum_{k=1}^{P}(\mu_k^{\Delta})^2}, \quad (26)$$

with the integration interval being redefined as

$$\mu_k^{\Delta} = \frac{K_k^{\Delta}}{\sum_{k=1}^{P} K_k^{\Delta}\tilde{\varphi}_1(\mathcal{T}, z_{\mathcal{T}}^k)}, \quad (27)$$

where

$$K_k^{\Delta} = exp\{\int_{t_\Delta}^{\mathcal{T}} c^y(r, z_r^k)dr\} \quad (28)$$

for t_Δ being the time instant where the last resampling occurred. This scheme is just applied when $N_{eff} < N_{thres}$, for a given N_{thres}.

3.2.1. Generalized Monte Carlo Filter Algorithm

Given an initial distribution ξ, the step-by-step algorithm of the so-called *Generalized Monte Carlo Filter* (GMCF) is described as the following:

At time $\mathcal{T}_0$,

Step 0: *Initialization*

- For $k = 1, ..., P$, sample $z_{\mathcal{T}_0}^k \sim \xi$;
- Set N_{thres} and $t_\Delta = \mathcal{T}_0$.
- Set $j = 1$.

While $\mathcal{T}_j \neq \mathcal{T}_N$,

Step 1: *Evolution*

- For $k = 1, ..., P$, evolve $z_{\mathcal{T}_j}^k$ in accordance to the model described by (11).

Step 2: *Importance weights evaluation*

- For $k = 1, ..., P$, evaluate the importance weights K_k^{Δ}, K_k, μ_k^{Δ} and μ_k.

Step 3: *Algorithm degeneracy computation*

- Compute N_{eff} according to (26);
- If $N_{eff} < N_{thres}$
  - · Set $t_\Delta = \mathcal{T}_j$;
  - · For $k = 1, ..., P$, replace the particles in $z_{\mathcal{T}_j}^k$ according to μ_k^{Δ}.

Step 4: *Conditional law computation*

- Compute $\pi_{\mathcal{T}_j}(\varphi)$ according to (24);
- Set $j = j + 1$ and go to **Step 1**.

End

When the parameter N_{thres} is chosen in a form that the resampling does not occur, the nonlinear filter is called Monte Carlo filter (MCF).

4. EXPERIMENTAL RESULTS

Consider the system suggested by Davis (1981*a*) and let the filtering problem be the following one-dimensional Ornstein-Uhlenbeck process

$$dx_t = -1.0\, dt + \alpha\, dw_t \quad (29)$$

whose one-dimensional observation process is

$$dy_t = h(x_t)\, dt + dv_t \quad (30)$$

with

$$h(x) = \begin{cases} sin(x) & \text{if } |x| \leq \pi/2 \\ +1 & \text{if } x > \pi/2 \\ -1 & \text{if } x < -\pi/2 \end{cases}, \quad (31)$$

where w_t and v_t are independent one-dimensional standard Brownian motions, and $\alpha > 0$ is the time-independent amplitude of the state noise. The filtering was carried out for $t \in [0, 5]$ by using a time step of $\delta = 2^{-8}s$, and the integral presented in (28) was approximated by the traditional Euler scheme. Given α, the signal process (29) was simulated according to its analytical solution whereas the obsevation process (30) used the Euler-Maruyama scheme (Kloeden and Platen, 1999). The signal initialization was set to

$x(0) \sim N(0.0, 0.25)$ to demonstrate the performance of the nonlinear filter over different SNR.

Preliminary experiments showed that MCF started diverging when the SNR in the state process is not favorable, that is, low. The time step δ was decreased aiming to verify the influence of discretization errors over the estimates, however the error magnitudes did not change. In this way, to investigate empirically what is happening, important information about the filter can be obtained by checking the average and the variance of the following coefficients: ϵ - the estimate square error, K^{Δ} - the unnormalised weight, and μ^{Δ} - the normalized weight. To provide robustness in the error estimates, the analysis was made for 50 different filter realizations in 50 different initializations.

Figure 1 presents the results of the coefficients for the MCF with $P = 50$ for three different values of state noise magnitude α: 0.10, 0.30, 0.50. For all the coefficients, it can be seen that the variance for $\alpha = 0.50$ is the biggest one, what seems that the filter has a limited point of stable operation with respect to the amplitude of the noise state. Based on these facts, the divergence effect seems to occur when the variance of μ^{Δ} starts increasing. Thus the resampling scheme can be used to decrease the referred variance and consequently makes the estimates more stabilized.

A resampling scheme for $N_{thres} = 0.98, 0.96$ was implemented for the amplitude state noise $\alpha = 0.50$ and the results are presented in Figure 2 for a longer time range of simulation. The action of the resampling really makes the estimate of the filter more stabilized and accurate due to the control of the variance amplitude of the weights. Specifically, though a great number of resampling controls the variance amplitude as given by $N_{thres} = 0.98$, the smoother mode of operation presented by $N_{thres} = 0.96$ also had good results. That is, apparently it is not necessary a great number of resamplings to control the divergence. Thus the choice of the effective number of particles is robust and its value is not the only variable to be checked in order to control effectively the stabilization of the estimates. Though resampling has proven practically useful in different contexts, theoretical results supporting the resampling idea must be developed in future research.

5. CONCLUSIONS

The *Pathwise* filter presented divergence in estimates for a low signal-to-noise ratio in the state noise variable for the Ornstein-Uhlenbeck process with saturation on the observation. Preliminary results showed that the referred filter can be rewritten in a way where suggested weights present sudden increase in its variance when the filter starts diverging. To solve this problem, a resampling scheme based on the effective number of particles was used and results show that when the weight variance is controlled, the filter estimates become stabilized.

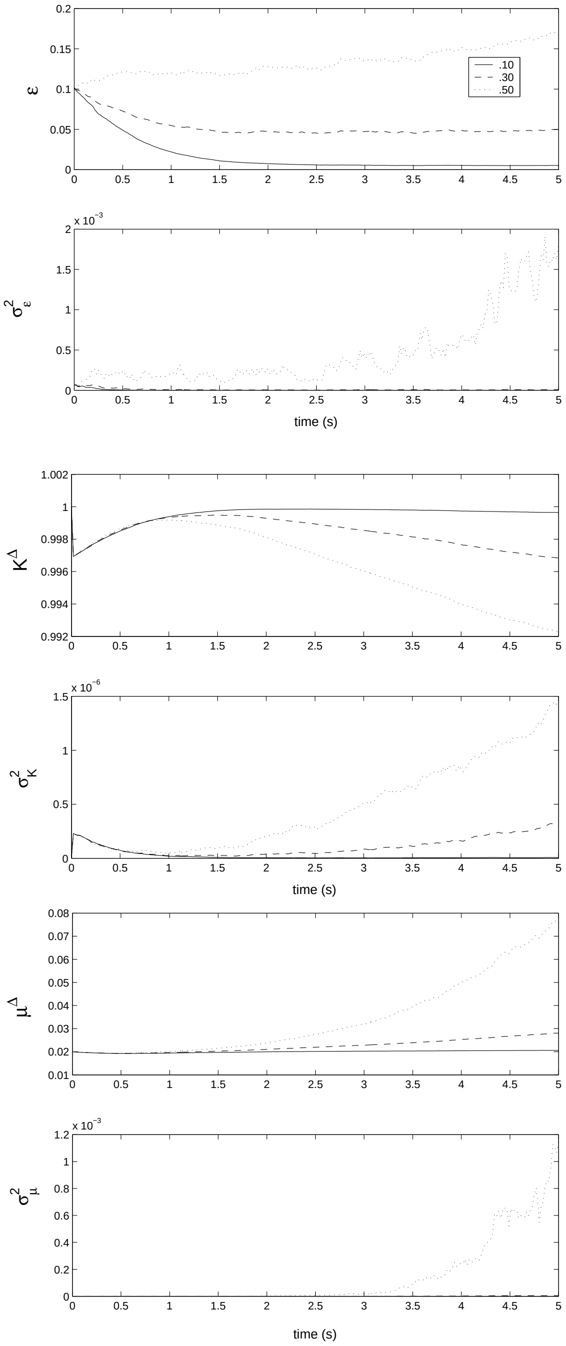

Fig. 1. Averages and variances of the square error ϵ, and weights K^{Δ} and μ^{Δ} for MCF with noise state given by $\alpha = 0.10, 0.30, 0.50$.

ACKNOWLEDGEMENT

The first author would like to thank The State of São Paulo Research Foundation (FAPESP) under Grant 02/10632-0.

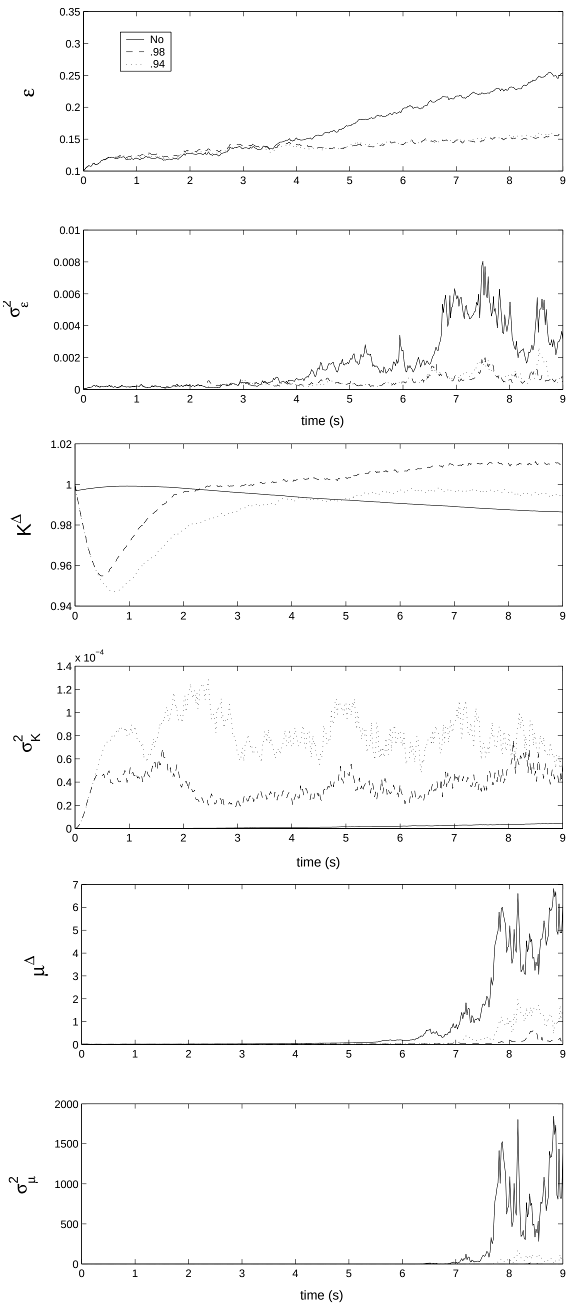

Fig. 2. Comparison of the averages and variances of the square error ϵ, and weights K^{Δ} and μ^{Δ} for MCF and GMCF with $N_{thres} = 0.98, 0.94$ and $\alpha = 0.50$.

REFERENCES

Clark, J. M. C. (1978). *Communication Systems and Random Process Theory*. Chap. The design of robust approximations to the stochastic differential equations of non-linear filtering, pp. 721–734. Sijthoff and Noordhoff.

Clark, J. M. C. and D. Crisan (2005). On a robust version of the integral representation formula of nonlinear filtering. *Probability Theory and Related Fields*. (to appear).

Davis, M. H. A. (1981*a*). New approach to filtering for nonlinear systems. *IEE Proc. D, Control Theory and Appl.* **128**(5), 166–172.

Davis, M. H. A. (1981*b*). *Stochastic systems: the mathematics of filtering and identification and applications*. Chap. Pathwise non-linear filtering, pp. 505–528. D. Reidel.

Del Moral, P. and L. Miclo (2000). *Séminaire de Probabilités XXXIV*. Chap. Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering, pp. 1–145. Vol. 1729 of *Lecture Notes in Mathematics*. Springer-Verlag. Berlin.

Doucet, A. (1998). On sequential simulation-based methods for bayesian filtering. Technical Report CUED/F-INFENG/TR-31. University of Cambridge. Cambridge.

Jacob, A. M., E. M. Hemerly and T. Yoneyama (2004). Improvements on Davis' Monte Carlo-based nonlinear filtering. In: *Congresso Brasileiro de Automática (CBA)*. Gramado.

Kallianpur, G. and H. Striebel (1968). Estimations of stochastic processes: Arbitrary system process with additive white noise observation errors. *Ann. Math. Statist.* **39**, 785–801.

Kloeden, P. E. and E. Platen (1999). *Numerical solution of stochastic differential equations*. 3rd ed.. Springer. Berlin.

Kushner, H. J. (1977). *Probability methods for approximations in stochastic control and elliptic equations*. Academic Press. New York.

Le Gland, F. (1984). Monte carlo methods in nonlinear filtering. In: *Proceedings of the 23rd IEEE Conference on Decision and Control*. Las Vegas.

Lipster, R. S. and A. N. Shirayaev (1977). *Statistics of randon processes I: general theory*. Springer-Verlag. New York.

O'Loghlen, J. W. and G. C. Wright (1982). Computer results for new nonlinear filtering algorithm. *IEE Proc. D, Control Theory and Appl.* **129**(2), 70–71.

Pardoux, E. (1981). *Stochastic systems: the mathematics of filtering and identification and applications*. Chap. Non-linear filtering, prediction and smoothing, pp. 529–557. D. Reidel.

Souza, E. T. de (1992). Métodos numéricos para filtragem não-linear. Ph.d. thesis. Instituto Tecnológico de Aeronáutica. São José dos Campos.

Zakai, M. (1969). On the optimal filtering for diffusion process. *Z. Wahrsch. Verw.* **11**, 230–243.