From Particle Filters to Malliavin Filtering with Application to Target Tracking

Sérgio Pequito

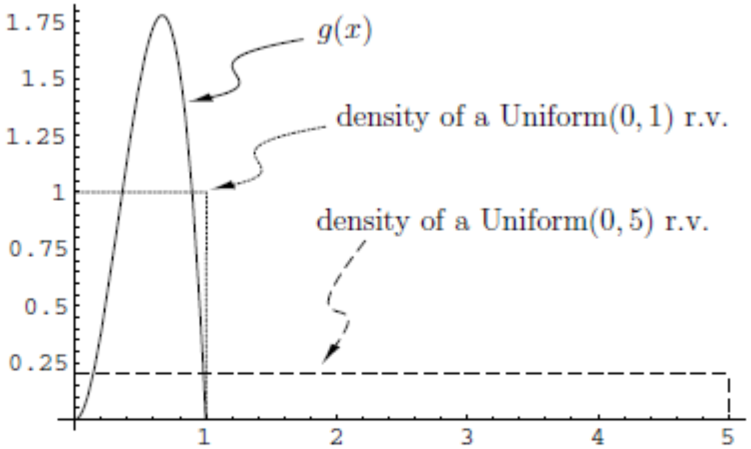
Phd Student

Final project of Detection, Estimation and Filtering course
July 7th, 2010

Guidelines

- **Motivation**
- **Monte Carlo Methods**
  - Importance Sampling
  - Control Variates
- **Particle filters**
- **Malliavin Estimator**
- **Example**

Motivation - Barely relevant sampling



$$W \sim Uniform(0,5)$$

$$\int_0^1 g(x)dx = 5E[g(W)] \rightarrow \frac{5}{N}\sum_{i=1}^{N} w_i, w_i \overset{iid}{\sim} Uniform(0,5)$$

$$U \sim Uniform(0,1)$$

$$\int_0^1 g(x)dx = E[g(U)] \rightarrow \frac{1}{N}\sum_{i=1}^{N} u_i, u_i \overset{iid}{\sim} Uniform(0,1)$$

Monte Carlo Methods

Importance sampling

Importance sampling involves a change of probability measure. Instead of taking X from a distribution with pdf $p_1(X)$, we instead take it from a different distribution with pdf $p_2(X)$

$$E_1\left[f(X)\right] = \int f(X)p_1(X)dX$$

$$= \int f(X)p_1(X)dX = \int f(X)\frac{p_1(X)}{p_2(X)}p_2(X)dX$$

$$= E_2\left[f(X)W(X)\right]$$

where $W(X) = \frac{p_1(X)}{p_2(X)}$ is the Radon-Nikodym derivative.

Importance sampling

We want the new variance $V_2\left[f(X)W(X)\right]$ to be smaller than the old variance $V_1\left[f(X)\right]$

How to achieve this? By making $W(X)$ small where $f(X)$ is large, and making $W(X)$ large when $f(X)$ is small.

Small $W(X) \Leftrightarrow$ large relative to $p_2(X)$ so more random samples in region where $f(X)$ is large

Particularly important for rare event simulation where $f(X)$ is zero almost everywhere

Control Variates

Suppose we want to approximate $E[f]$ using a simple Monte Carlo average $\bar{f}$

If there is another payoff g for which we know $E[g]$, can use $\bar{g} - E[g]$ to reduce error in $\bar{f} - E[f(X)]$

How? By defining a new estimator

$$\hat{f} = \bar{f} - \lambda\left(\bar{g} - E[g]\right)$$

Unbiased since $E\left[\hat{f}\right] = E\left[\bar{f}\right] = E[f]$

Control Variates

For a single sample

$$Var\left[f-\lambda\left(\overline{g}-E[g]\right)\right]=Var[f]-2\lambda Cov[f,g]+\lambda^2 Var[g]$$

For a average of *N* samples

$$Var\left[f-\lambda\left(\overline{g}-E[g]\right)\right]=N^{-1}\left(Var[f]-2\lambda Cov[f,g]+\lambda^2 Var[g]\right)$$

To minimize this, the optimum value for λ is

$$\lambda=\frac{Cov[f,g]}{Var[g]}$$

Control Variates

The resulting variance is

$$N^{-1}Var[f]\left(1-\frac{(Cov[f,g])^2}{Var[f]Var[g]}\right)=N^{-1}Var[f](1-\rho^2)$$

Where ρ is the correlation between f and g .

The challenge is the choose a good g which is well correlated with f - the covariance, and hence the optimal λ , can be estimated from the data.

Remarks

Importance sampling – very useful for applications with rare events, but again needs to be fine-tuned for each application

Control variates – easy to implement and can be very effective
But requires careful choice of control variate in each case

Overall, a tradeoff between simplicity and generality on one hand, and efficiency and programming effort on the other.

Particle filters

Bayes' Theorem

$$P(AB) = P(BA) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

$$P(A \mid BC) = \frac{P(BC \mid A)P(A)}{P(BC)} = \frac{P(B \mid CA)P(C \mid A)P(A)}{P(B \mid C)P(C)} = \frac{P(B \mid AC)P(A \mid C)}{P(B \mid C)}$$

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{\sum_C P(ABC)}{P(B)} = \sum_C \frac{P(ACB)P(CB)}{P(CB)P(B)} = \sum_C P(A \mid CB)P(C \mid B)$$

Nonlinear Filtering

- **State evolution**
  - discrete time stochastic model (Markovian)

$$x_{k+1} = f_k(x_k, v_k)$$

- **Measurement equation**

$$z_k = h_k(x_k, w_k)$$

- **Model prediction given past measurements**

$$p(x_{k+1} \mid z_{1:k}) = \int p(x_{k+1} \mid x_k, z_{1:k}) p(x_k \mid z_{1:k}) dx_k$$

$$= \int p(x_{k+1} \mid x_k) p(x_k \mid z_{1:k}) dx_k$$

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{\sum_C P(ABC)}{P(B)} = \sum_C \frac{P(ACB)P(CB)}{P(CB)P(B)} = \sum_C P(A \mid CB) P(C \mid B)$$

Nonlinear Filtering

- **Prediction update using current measurement**

$$p(x_{k+1} \mid z_{1:k+1}) = p(x_{k+1} \mid z_{k+1}, z_{1:k})$$

posterior

$$= \frac{p(z_{k+1} \mid x_{k+1}, z_{1:k})\, p(x_{k+1} \mid z_{1:k})}{p(z_{k+1} \mid z_{1:k})}$$

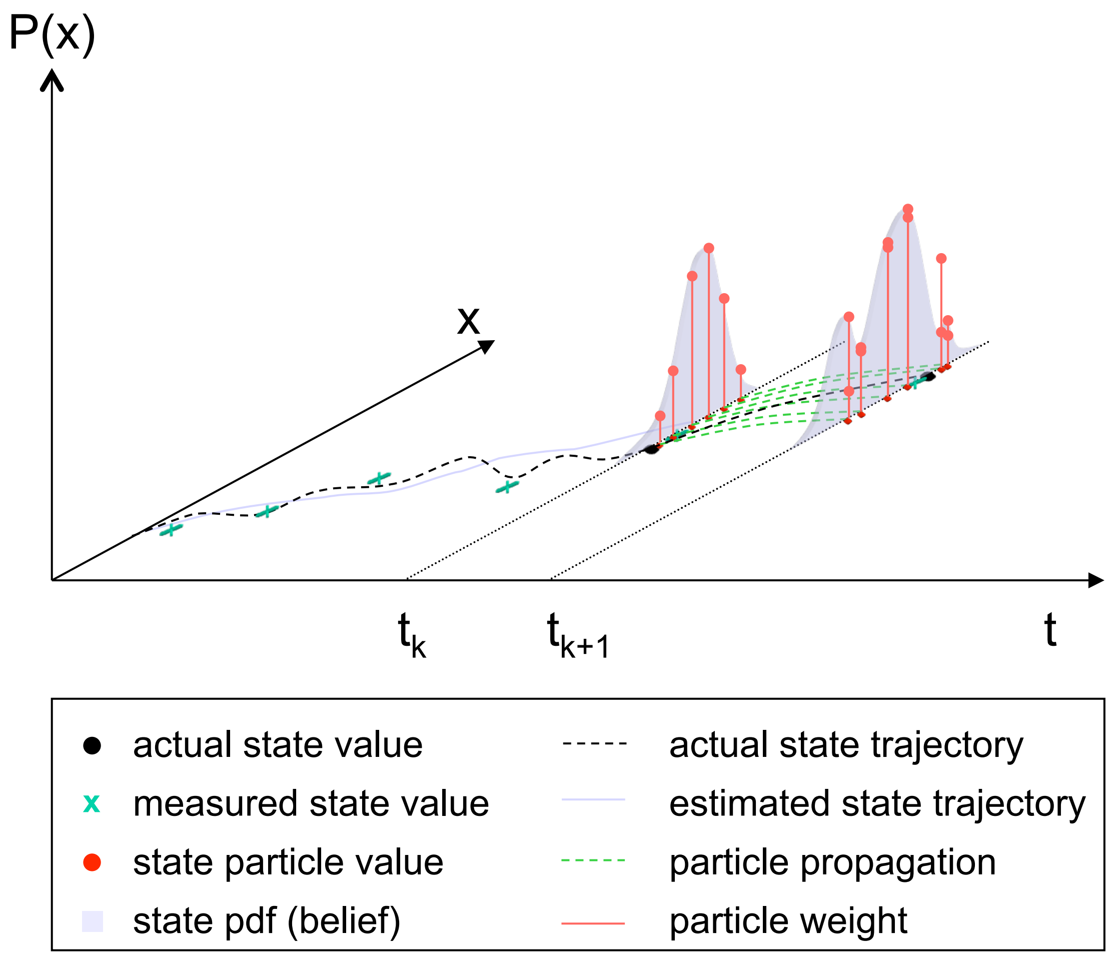
$$P(A \mid BC) = \frac{P(B \mid AC)P(A \mid C)}{P(B \mid C)}$$

$$= \frac{p(z_{k+1} \mid x_{k+1})\, p(x_{k+1} \mid z_{1:k})}{p(z_{k+1} \mid z_{1:k})}$$

$$p(z_{k+1} \mid z_{1:k}) = \int p(z_{k+1} \mid x_{k+1})\, p(x_{k+1} \mid z_{1:k})\, dx_{k+1}$$

likelihood | dynamic prior

Particle Filtering



- **represent state as a pdf**
- **sample the state pdf as a set of particles and associated weights**
- **propagate particle values according to model**
- **update weights based on measurement**

Particle Filtering

- **Prediction step:** use the state update model

$$p(\mathbf{x}_k \mid \mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} \mid \mathbf{z}_{1:k-1}) d\mathbf{x}_{k-1}$$

- **Update step:** with measurement, update the prior using Bayes' rule:

$$p(\mathbf{x}_k \mid \mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k \mid \mathbf{x}_k) p(\mathbf{x}_k \mid \mathbf{z}_{1:k-1})}{p(\mathbf{z}_k \mid \mathbf{z}_{1:k-1})}$$

Particle Filtering

- **A particle filter iteratively approximates the posterior *pdf* as a set:**

$$S_k = \{\langle x_k^i, w_k^i \rangle \mid i = 1, \ldots, n\}$$

$$p(x_k \mid z_{1:k}) \approx \sum_{i=1}^{n} w_k^i \delta(x_k - x_k^i)$$

where:

x_k^i is a point in the state space

w_k^i is an importance weight associated with the point

Particle Filtering

- **Implementation steps**
  - propose $x_0{}^i$ and propagate particles

    $$x_k^i = f_k(x_{k-1}^i, \omega_k)$$

  - compute likelihood of measurement w.r.t. each particle

    $$z_k = h_k(x_k^i, \upsilon_k) \quad \leftrightarrow \quad p(z_k \mid x_k^i)$$

  - update particle weights based on likelihood

    $$w_k^i = w_{k-1}^i \frac{p(z_k \mid x_k^i)\, p(x_k^i \mid x_{k-1}^i)}{q(x_k^i \mid x_{0:k-1}^i, z_{1:k})}$$

    $$q(x_k^i \mid x_{0:k-1}^i, z_{1:k}) \approx p(x_k^i \mid x_{k-1}^i) \Rightarrow w_k^i = w_{k-1}^i p(z_k \mid x_k^i)$$

  - normalize weights $w_k^i = w_k^i / \sum_{i=1}^{n} w_k^i$

Resampling

- **Particle weights degenerate over time**
  - measure of degeneracy: effective sample size

$$\hat{n}_{eff} = 1 / \sum_{i=1}^{n} (w_k^i)^2$$

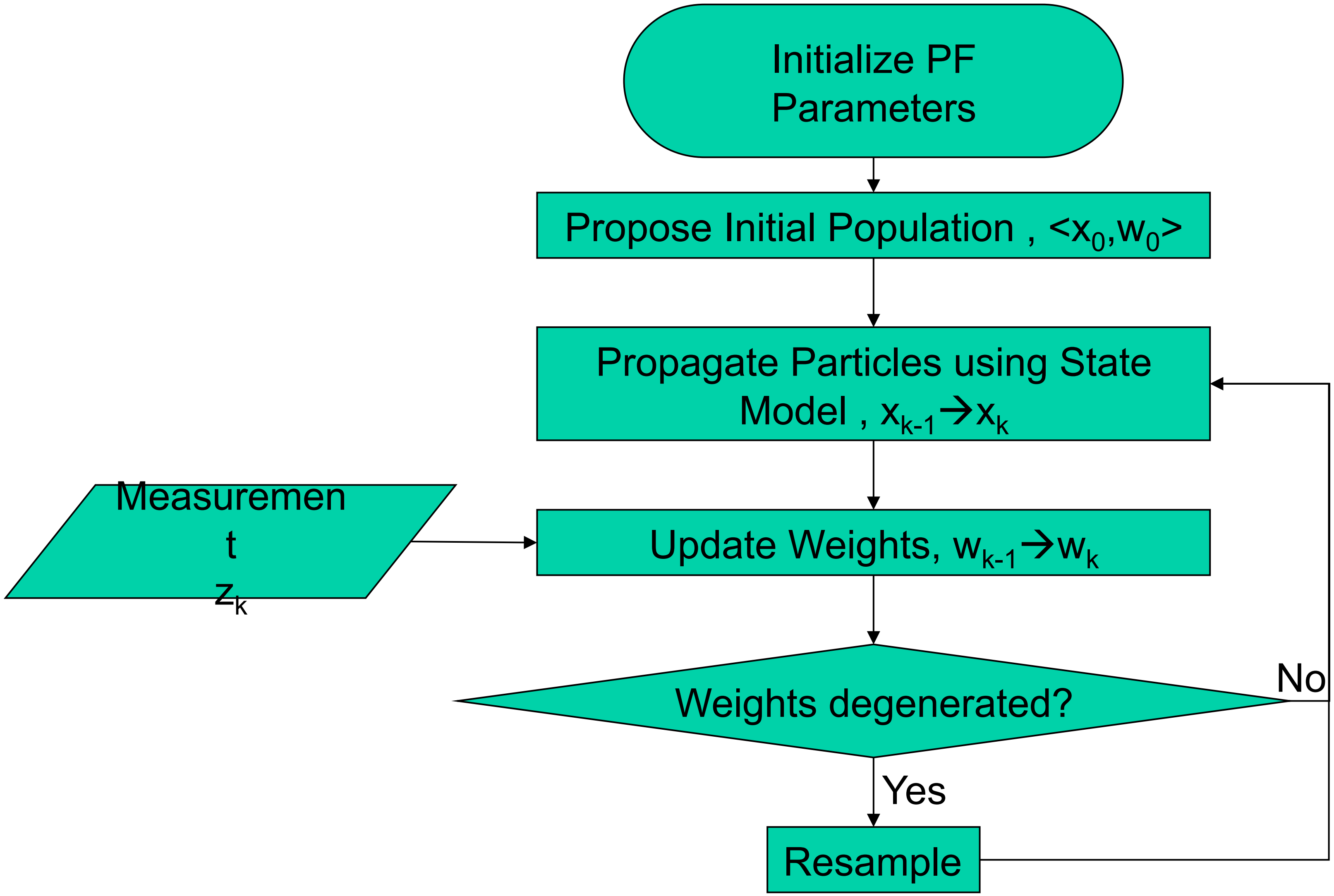
use normalized weights

$$1 \le \hat{n}_{eff} \le n$$

  - resample whenever $\hat{n}_{eff} < n_{thr}$
  - new set of particles have same statistical properties

$$\{x_k^i, w_k^i\} \Leftrightarrow \{x_k^{i*}, 1/n\}$$

PF Flowchart



Malliavin Estimator

Malliavin Estimator

Particle Filter $\oplus$ "Grid based method"

$$p\left(X_{k-1} = x_{k-1}^i \mid z_{1:k-1}\right) = w_{k-1|k-1}^i$$

$$p\left(X_{k-1} \mid z_{1:k-1}\right) = \sum_{i=1}^{N} w_{k-1|k-1}^i \delta\left(X_{k-1} - x_{k-1}^i\right)$$

$$p\left(X_k \mid z_{1:k-1}\right) = \sum_{i=1}^{N} w_{k|k-1}^i \delta\left(X_k - x_k^i\right)$$

$$p\left(X_k \mid z_{1:k}\right) = \sum_{i=1}^{N} w_{k|k}^i \delta\left(X_k - x_k^i\right)$$

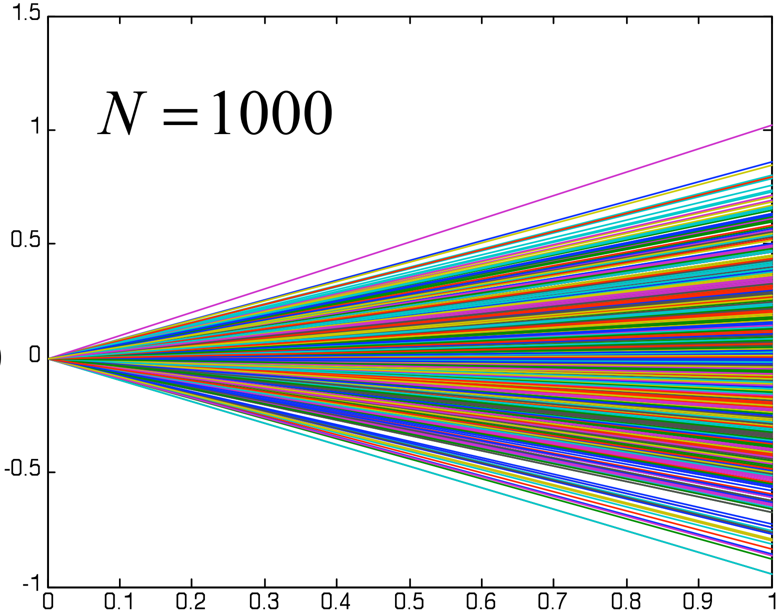
$$w_{k|k-1}^i = \sum_{j=1}^{N} w_{k-1|k-1}^j p\left(x_k^i \mid x_{k-1}^j\right)$$

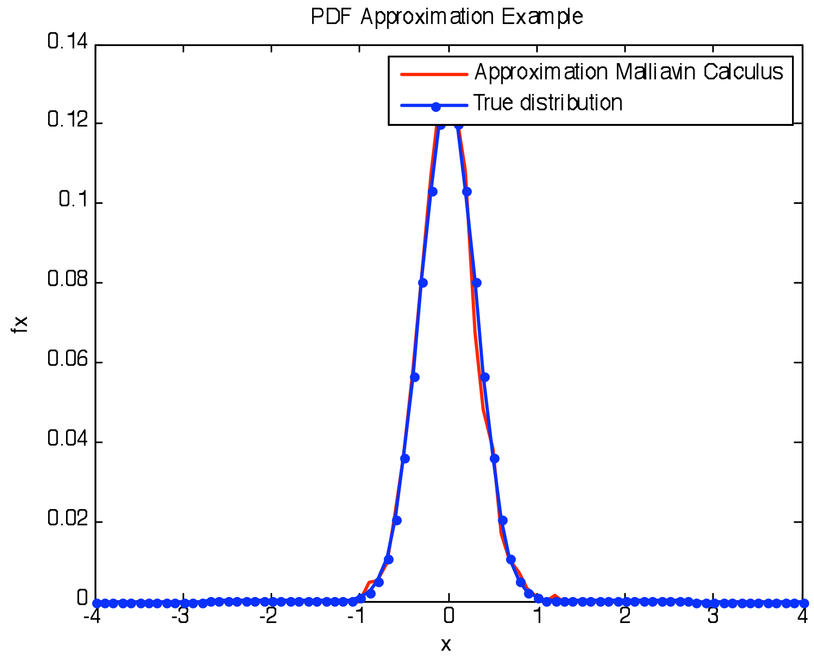
$$w_{k|k}^i = \frac{w_{k|k-1}^i p\left(z_k \mid x_k^i\right)}{\sum_{j=1}^{N} w_{k|k-1}^i p\left(z_k \mid x_k^i\right)}$$

Malliavin Estimator

$$p\left(x_k \mid x_{k-1}^i\right)$$

$\delta\left(x_0 = 0\right)$





$$dX_t = \sigma\, dW_t$$

Malliavin Estimator

$$E\left[f'(X)\right]=\int f'(x)p(x)dx$$

$$=-\int f(x)\frac{p'(x)}{p(x)}p(x)dx=E\left[f(X)\frac{p'(x)}{p(x)}\right]$$

$$=E\left[f(X)\frac{p'(x)}{p(x)}\right]=E\left[f(X)H(X,1)\right]$$

$$H(X,1)=\frac{p'(x)}{p(x)}$$

Malliavin Estimator

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s, t \in [0,T]$$

If Hörmander hypothesis is satisfied then X_T density exists and is smooth

Use Malliavin calculus to develop an expression for H(X,1) that can be Simulated. So we get a Monte Carlo for (with variance reduction)

$$E[f'(X)] = \frac{1}{N}\sum f(\tilde{X}^i)H(\tilde{X}^i,1)$$

Where $\tilde{X}^i$ are independent Euler approximations of X .

For us $f'(X) = \delta(X)$ then $E[\delta(X)] = p(x)$ (pdf of X)

Malliavin Estimator

The same simulated paths give good estimates for densities at any point. That is, one can compute the density over the whole real line with the same number of paths.

Let $\varphi, \frac{d\varphi}{dx} \in L^2(\mathbb{R}), \varphi(0)=1$ a localization function and r a parameter and $c \in L^2(\mathbb{R})$ a "control variate".

The density function

$$p(x) = E\left[\xi_{c,r}(x)\right]$$

where

$$\xi_{c,r}(x) = \left(1_{\{X \geq x\}} - c(x)\right) H\left(X, \varphi\left(\frac{X-x}{r}\right)\right)$$

Malliavin Estimator

where $$H\left(X,\varphi\left(\frac{X-x}{r}\right)\right)=\varphi\left(\frac{X-x}{r}\right)H(X,1)-\frac{1}{r}\varphi'\left(\frac{X-x}{r}\right)$$

and

$$H(X,1)=\frac{\int_0^T dW_t}{\int_0^T D_s X\,ds}+\frac{\int_0^T\int_0^T D_t D_s X\,ds\,dt}{\left(\int_0^T D_s X\,ds\right)^2}$$

The variance of $\xi_{c,r}(x)$ is minimized for

$$c(x)=\frac{E\left[1_{\{X\geq x\}}H\left(X,\varphi\left(\frac{X-x}{r}\right)\right)^2\right]}{E\left[H\left(X,\varphi\left(\frac{X-x}{r}\right)\right)^2\right]}$$

$$r=\sqrt{\frac{\int_0^\infty \varphi'(z)^2\,dz}{E\left[H(X,1)^2\right]\int_0^\infty \varphi(z)^2\,dz}}$$

$$\varphi(x)=e^{-\lambda|x|},\lambda>0$$

Malliavin Estimator

For the numerical implementation one have to discretize the following By "Euler" scheme

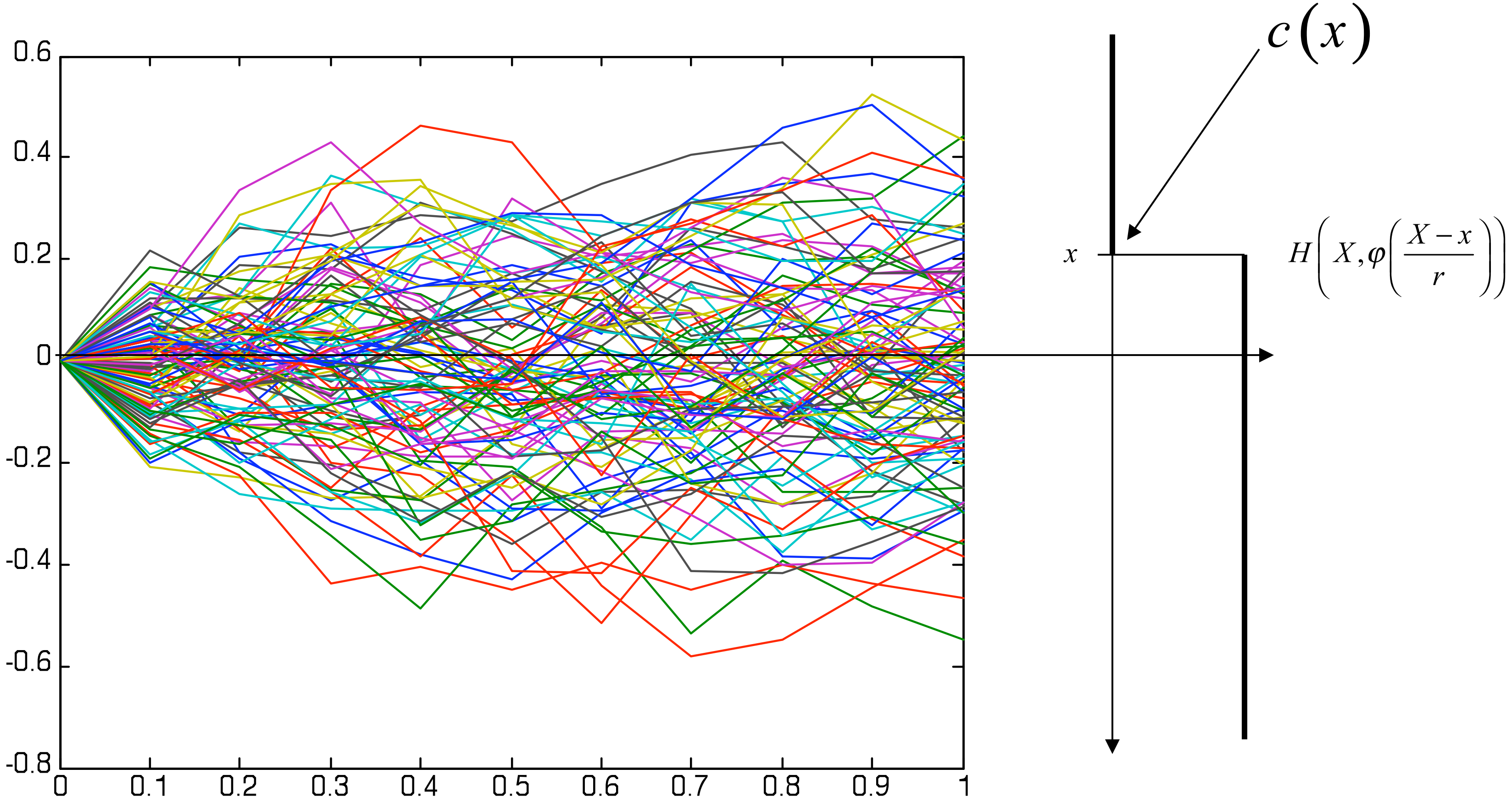
$$D_s X_t = \begin{cases} \sigma e^{\int_s^t \bar{b}'(X_v)dv + \int_s^t \sigma'(X_v)dW_v}, & s \le t \\ 0 & s > t \end{cases} \qquad \bar{b}'(X_v) = b'(X_v) - \frac{1}{2}\sigma'(X_v)^2$$

Moreover

$$D_s D_t X_T = D_s(X_t)\sigma'(X_t)e^{\int_t^T \bar{b}'(X_v)dv + \int_t^T \sigma'(X_v)dW_v}$$

$$+\left[\sigma'(X_t)1_{\{t\le s\}} + \int_t^T \bar{b}''(X_v)D_s X_v dv + \int_t^T \sigma''(X_v)D_s X_v dW_v\right] D_t X_T$$

Malliavin Estimator



$$\xi_{c,r}(x) = \left(1_{\{X \geq x\}} - c(x)\right) H\left(X, \varphi\left(\frac{X-x}{r}\right)\right)$$

Example

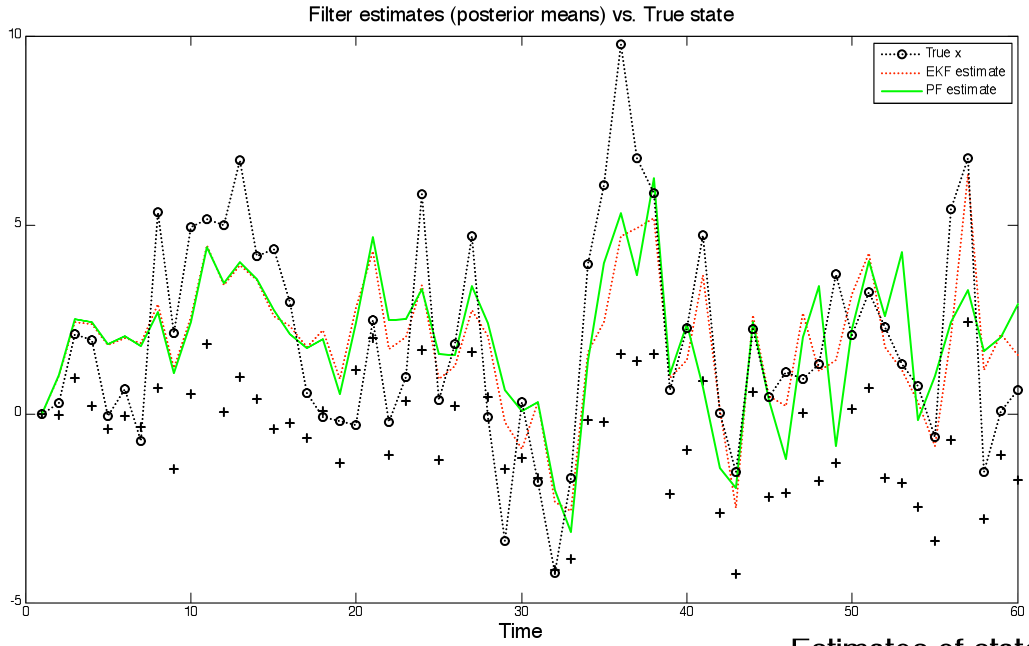
Malliavin Estimator

$$x_k = 1 + \sin\left(4 * 10^{-2} * \pi * k\right) + 0.5 * x_{k-1} + \sqrt{5} * w$$

$$y_k = \begin{cases} \dfrac{x_k}{5} + v & k \leq 30 \\ -2 + \dfrac{x_k}{2} + v & k > 30 \end{cases}$$

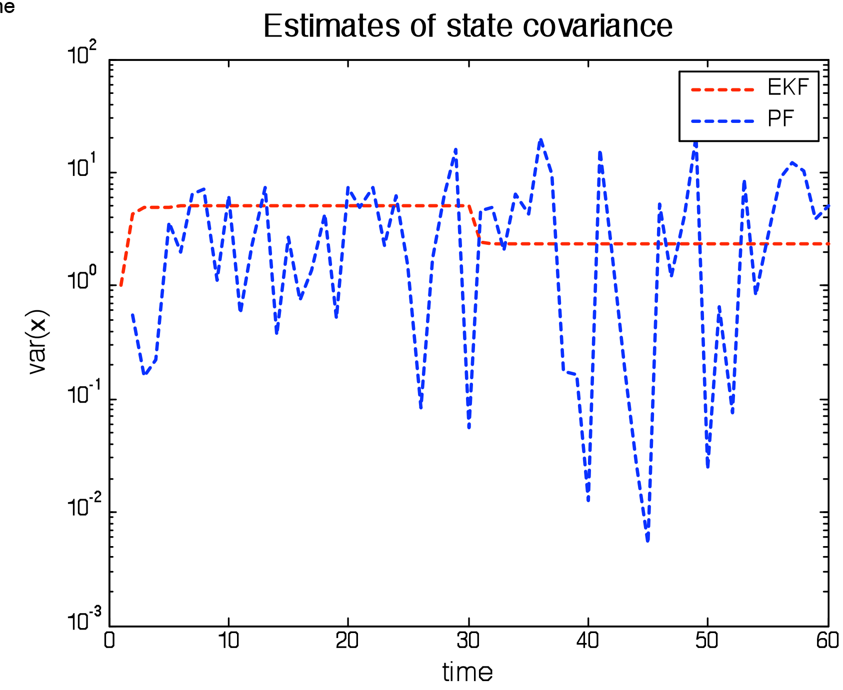
v, w are W.G.N.

N=1000 particles



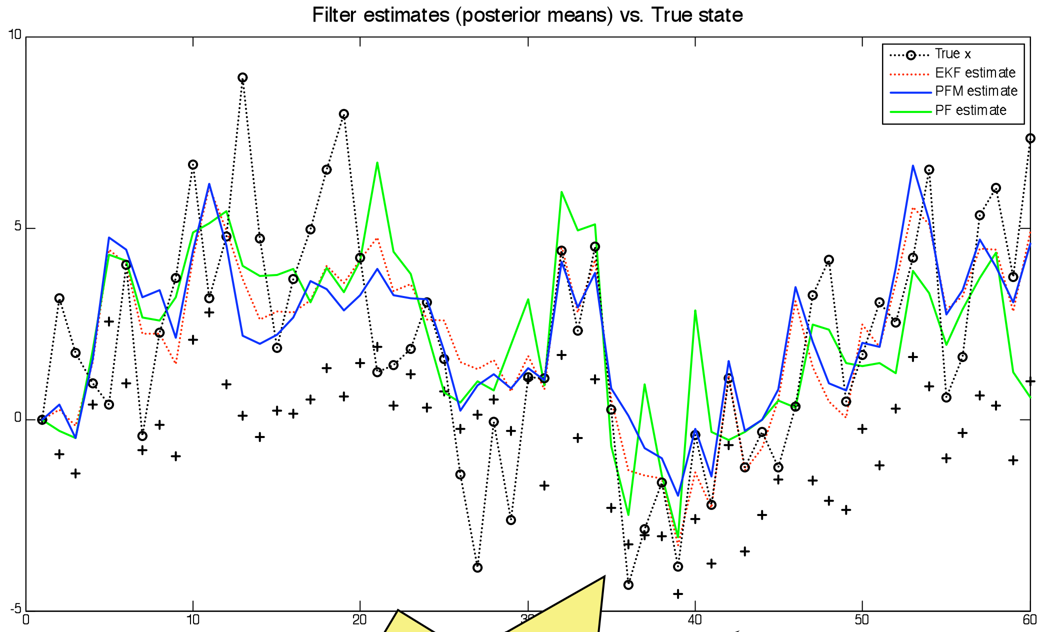
Command Window

```
************* FINAL RESULTS *****************

RMSE : mean and variance
---------

EKF             = 1.8681 (0.0403)
PF              = 2.0968 (0.053418)


Execution time  (seconds)
-------------------------

PF              = 3.7459
```



N=20 particles

Brownian Paths: 50



1000 runs!!!

Command Window

************* FINAL RESULTS ******

RMSE : mean and variance

EKF = 1.8663 (0.043806)
PFM = 1.9872 (0.051971)
PF = 2.4894 (0.079156)

Execution time (seconds)

PF = 0.12731
PFM = 3.801

****** FINAL RESULTS ****************

RMSE : mean and variance

EKF = 1.8681 (0.0403)
PF = 2.0968 (0.053418)

Execution time (seconds)

PF = 3.7459

Conclusion and Further Research

- Different SMC method with variance reduction (no need to calibration)
- Possibility to extend to nonlinear case and extend to higher dimensions

References

B. Bouchard, I. Ekeland and N. Touzi (2002). On the Malliavin approach to Monte Carlo approximation of conditional expectations.

B. Bouchard and N. Touzi (2002), Discrete time approximation and Monte-Carlo simulation of Backward Stochastic Differential Equations

E. Fournié, J.M.Larsy, J. Lebouchoux, P.L. Lions, N. Touzi (1999). Applications of Malliavin calculus to Monte Carlo methods in Finance I

E. Fournié, J.M.Larsy, J. Lebouchoux, P.L. Lions (2001). Applications of Malliavin calculus to Monte Carlo methods in Finance II

D. Nualart (1995). The Malliavin Calculus and Related Topics.

Doucet, A.; De Freitas, N.; Gordon, N.J. (2001). Sequential Monte Carlo Methods in Practice

Doucet, A.; Johansen, A.M.; (2008). A tutorial on particle filtering and smoothing: fifteen years later

Arulampalam, M.S.; Maskell, S.; Gordon, N.; Clapp, T.; (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking