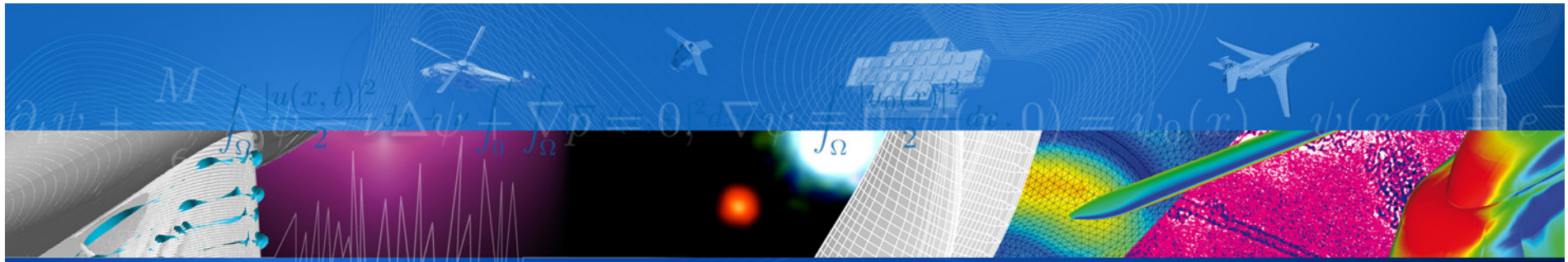
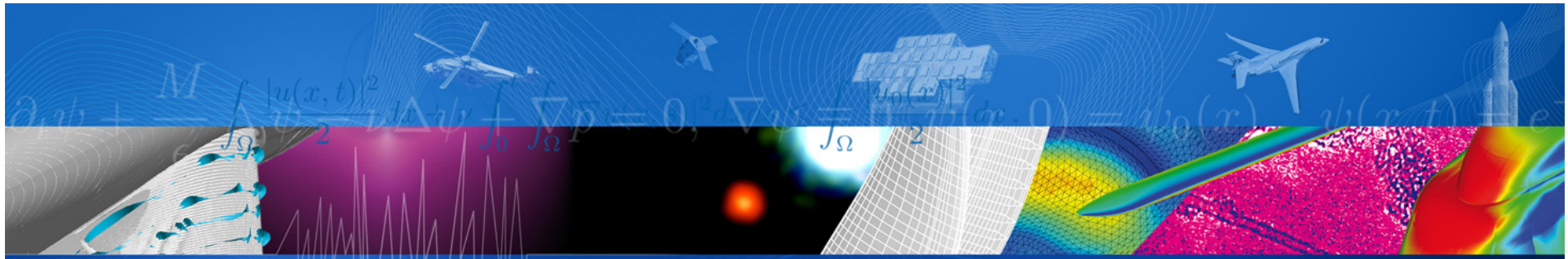


ONERA

THE FRENCH AEROSPACE LAB

retour sur innovation

www.onera.fr



Robustness and accuracy in particle filtering

Application to navigation and tracking

Christian Musso (Onera/DTIM)

PhD thesis :

Achille Murangira. Supervision Onera & UTT. Grant DGA/Onera
Paul Bui Quang. Supervision Onera & INRIA. Grant Onera

ONERA
THE FRENCH AEROSPACE LAB

retour sur innovation

Outline

- Behavior of the PF in case of multimodality
- Mixture filters
- Application to terrain navigation

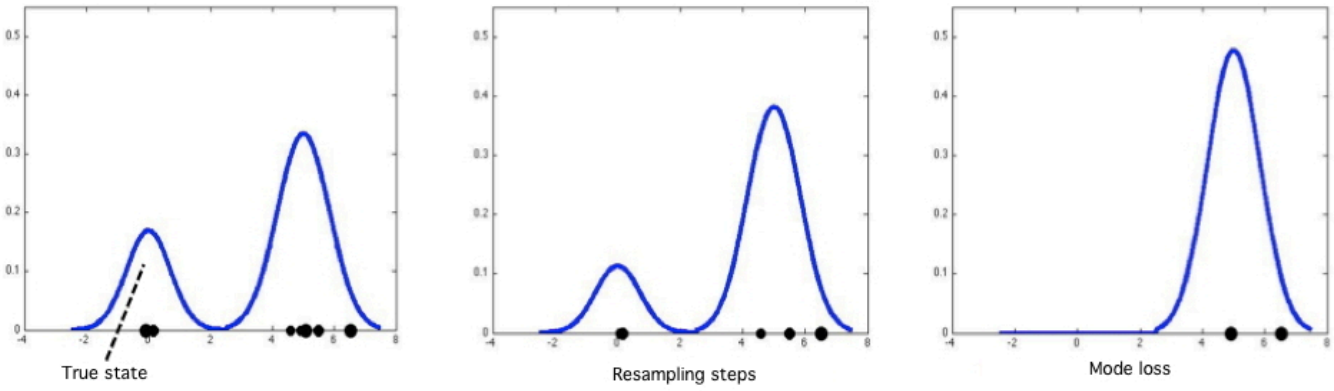
- Analysis of the Monte Carlo error with the Laplace method
- Importance density
- Application to tracking

Multimodality : behavior of the PF

In theory PF deals with multimodality when N is « large » enough.

In practice we observe mode losses :
divergence (navigation), loss of a target (multi-target in radar, one target with low SNR)

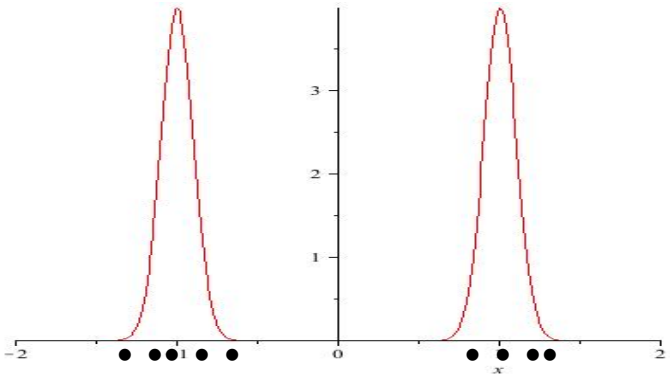
Cause : successive resampling



Theorical analysis in a simple case : bimodal density

Multimodality : behavior of the PF

$$p = \frac{1}{2}N(-1,0.1) + \frac{1}{2}N(1,0.1)$$



$$(X_1^1,\ldots,X_N^1)$$

Resampling
(with replacement)

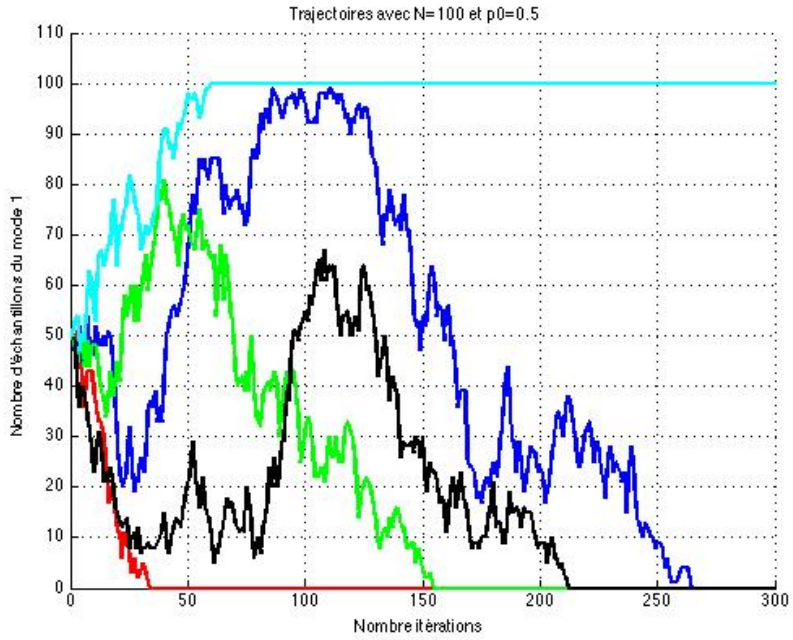
$$(X_1^2,\ldots,X_N^2)$$

$$(X_1^k,\ldots,X_N^k)$$

How many iterations before losing a mode ? Wright Fisher model (genetic)

Multimodality : behavior of the PF

A_k : sample size from mode #1



Markov chain : $$p_{mj} = P(A_k = j / A_{k-1} = m) = C_N^j \left(\frac{m}{N}\right)^j \left(1 - \frac{m}{N}\right)^{N-j}$$

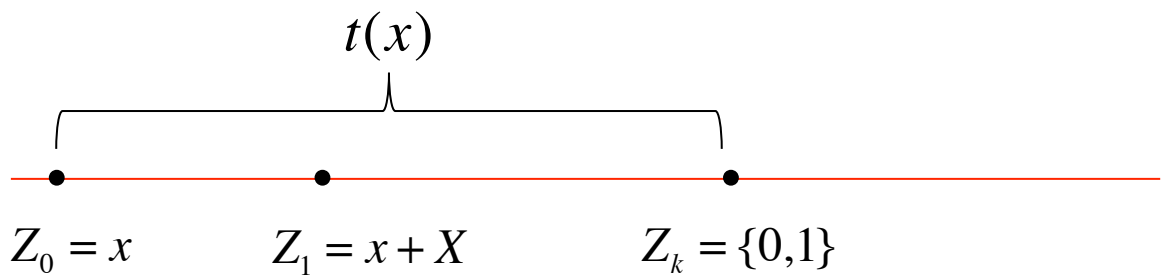
Absorbant states : {N} and {0}

Absorption time: $$\tau = \inf\{k / A_k \in \{0, N\}\}$$

Multimodality : behavior of the PF

Mean time of absorption : $E_i[\tau] = E[\tau / A_0 = i]$

Denote $Z_n = A_n / N$. Computing $t(x) = E[\tau / Z_0 = x]$ (N large)



Markov : $t(x) = 1 + E[t(x+X)]$

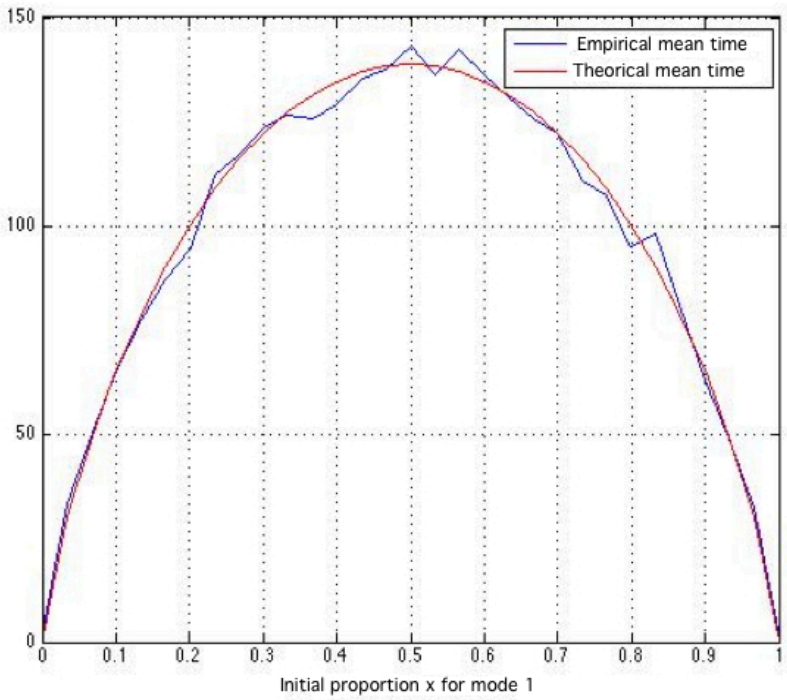
With : $E[X] = 0$, $E[X^2] = \dfrac{x(1-x)}{N}$ $N(x+X)$: binomial distribution

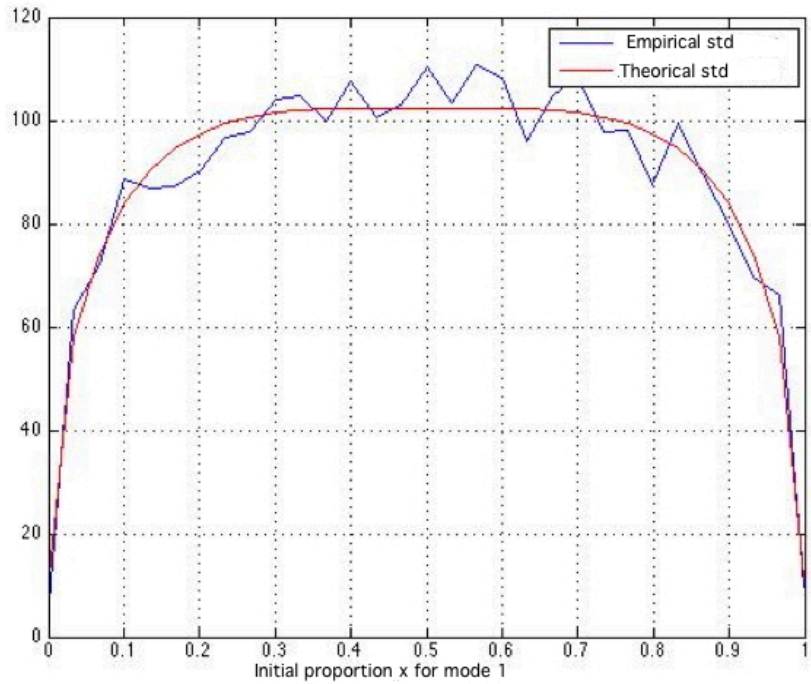
$$t(x+X) \approx t(X) + t'(x)X + \frac{t''(x)}{2}X^2 \longrightarrow t''(x) = \frac{-2N}{x(1-x)}$$

Multimodality : behavior of the PF

Expectation time of absorption : $t(x) \approx -2N[x\log(x)+(1-x)\log(1-x)]$

Variance time of absorption (x=1/2) : $\left(\frac{2}{3}\pi^2 - 8\log(2)\right)N^2 + 2\log(2)N$



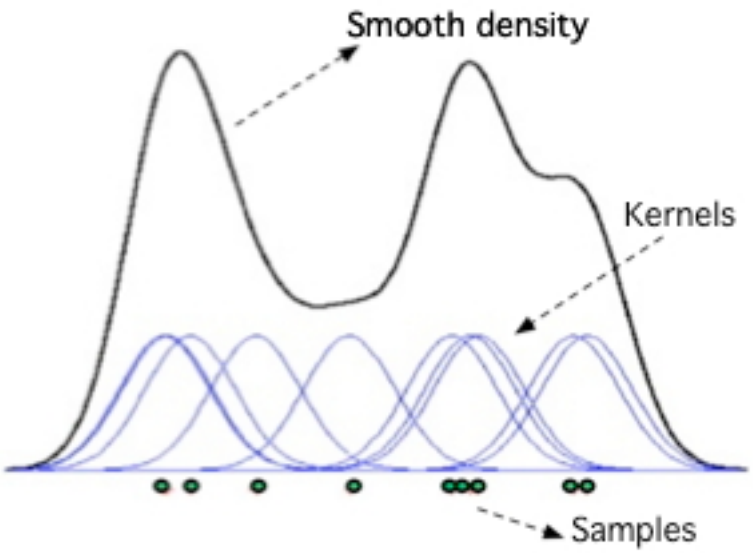


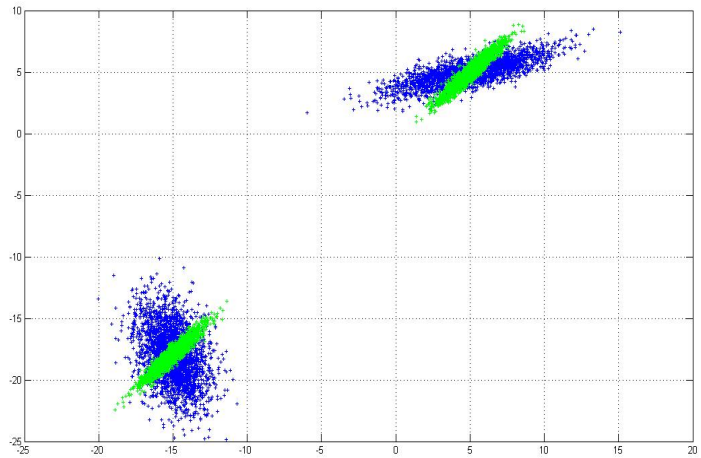
N=100

PF used in the sequel : regularized PF (RPF)

$$\hat{p}_h(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{h^d}K\left(\frac{A^{-1}(x-X_i)}{h}\right)$$

- Sample X_k according the weights among $(X_1,\ldots,X_N)$
- Sample ξ according to the kernel K
- Add noise after whitening $X_k^* = X_k + hA^{-1}\xi$





Distortion of the multimodal density

Multimodality : mixture filters

Maintaining the modes : mixture of particle filters

$$p_k = p\left(X_k \middle| Y_{0:k}\right) = \sum_{l=1}^{L} \alpha_{l,k} p_l\left(X_k \middle| Y_{0:k}\right)$$

where $\sum_{l=1}^{L} \alpha_{l,k} = 1$ and $\alpha_{l,k}$ likelihood of component l

- Propagation $p_{k+1|k} = p\left(X_{k+1} \middle| Y_{0:k}\right) = \sum_{l=1}^{L} \alpha_{l,k} p_l\left(X_{k+1} \middle| Y_{0:k}\right)$

- Correction/regularization

$$p_{k+1} = \sum_{l=1}^{L} \alpha_{l,k+1} p_l\left(X_{k+1} \middle| Y_{0:k+1}\right) \qquad \alpha_{l,k+1} = \frac{\alpha_{l,k} p_l\left(Y_{k+1} / Y_k\right)}{\sum_{j=1}^{L} \alpha_{j,k} p_j\left(Y_{k+1} / Y_k\right)}$$

Correction and prediction are done locally

Multimodality : mixture filters

- Mixture component mix p_l → estimated by $\hat{p}_l$ with N_l particules

$$\hat{p}_l = \sum_{i=1}^{N_l} \omega_{l,k}^i \delta_{x_{l,k}^i} \qquad \forall l = 1 \ldots L$$

$$\Rightarrow \hat{p}_k^N = \sum_l \alpha_{l,k} \left(\sum_{i=1}^{N_l} \omega_{l,k}^i \delta_{x_{l,k}^i} \right)$$

- Total number of particuls remains unchanged $\sum_{i=1}^{L} N_l = N$

- main pros: local resampling for each component of $\hat{p}_l$

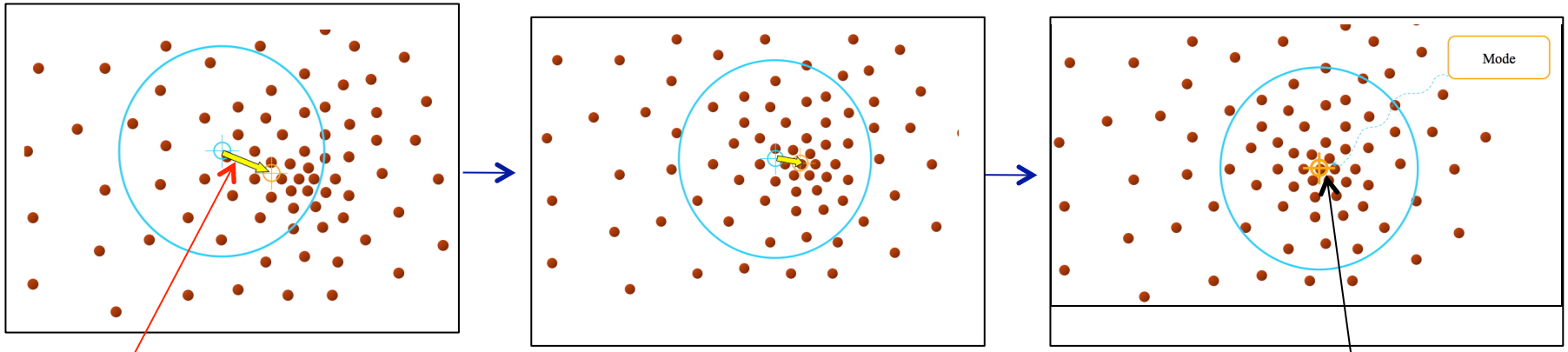
  → Maintaining the mode associated insured if $\alpha_{l,k} > threshold$

- In terrain navigation the number of modes evolves with time

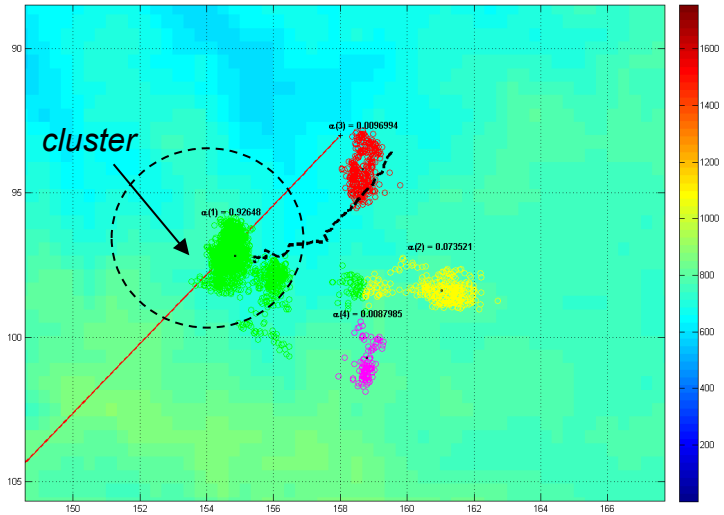
  → Grouping particules to the same mode with clustering algorithm

Multimodality : mixture filters

Mean shift clustering used after correction step : grouping particules associated to the same mode



Meanshift vector



Grouping particules around the mode

Multimodality : mixture filters

- Nb of modes changes : clusterisation with meanshift at each correction
- Nb of particules constant : $\sum_{i=1}^{L} N_l = N$
- Threshold on the likelihood of the mode

$$\hat{p}_k^N = \sum_{l=1}^{L} \alpha_{l,k} \sum_{i=1}^{N_l} \omega_{l,k}^i \delta_{x_{l,k}^i}$$

Prediction / Correction

$$\hat{p}_{k+1}^N = \sum_{l=1}^{L} \alpha_{l,k} \sum_{i=1}^{N_l} \omega_{l,k+1}^i \delta_{x_{l,k+1}^i}$$

Clustering

$$\hat{p}_{k+1}^N = \sum_{l=1}^{M} \beta_{l,k+1} \sum_{i=1}^{N_l} \upsilon_{l,k+1}^i \delta_{x_{l,k+1}^i}$$

Actual cost actuel x 2 but alternative in progress (morphological) : x 1.3

Application to terrain navigation

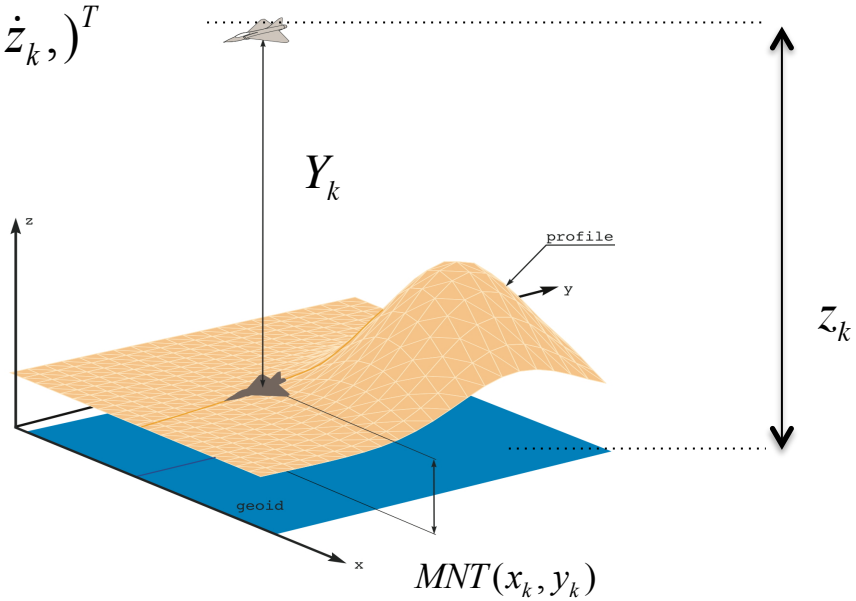
- State model $X_k = (x_k, y_k, z_k, \dot{x}_k, \dot{y}_k, \dot{z}_k,)^T$

$$X_k = f(X_{k-1}) + w_k$$

- Measurement model

$$Y_k = z_k - MNT(x_k, y_k) + v_k = h(X_k) + v_k$$

h : strongly nonlinear



Aim of PF: compute recursively the conditional density $p(x_k | y_{0:k})$

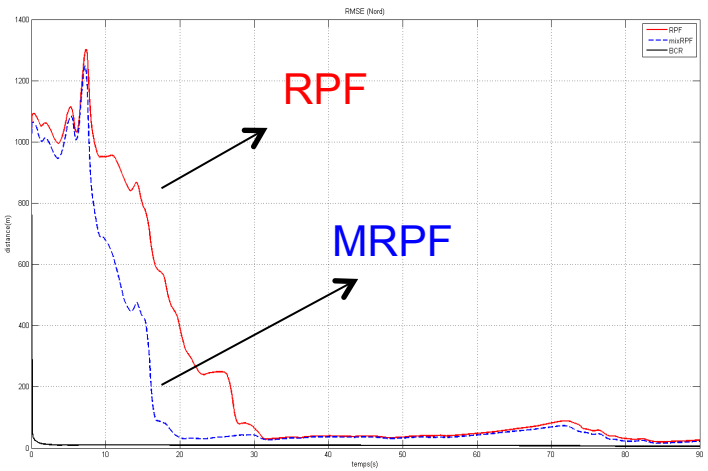
Ambiguous map: multimodal density

Application to terrain navigation

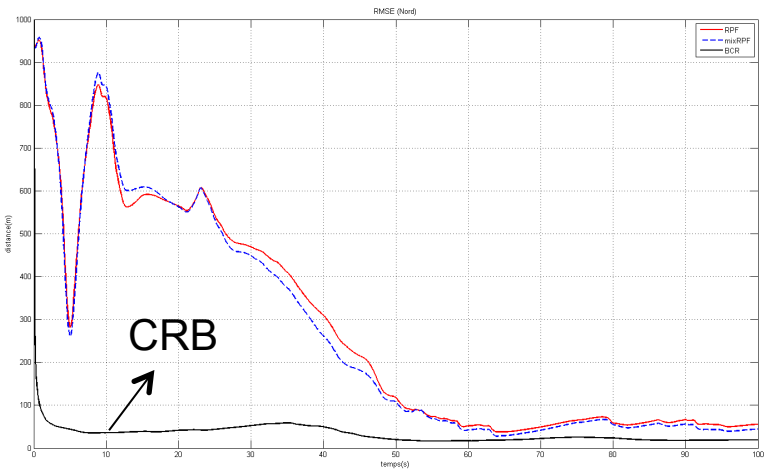
Performances in 3 scenarios : Roughness (weak, average, high)

Standard algorithm : RPF (Regularized Particle Filter)

Proposed algorithme : MRPF (Mixture Regularized Particle Filter)



RMSE position Nord - – High Roughness



RMSE position Nord - – Weak Roughness

Application to terrain navigation

Roughness	MRPF	RPF
weak	0%	0%
average	3%	7%
high	1%	2%

Divergence rate of the filters

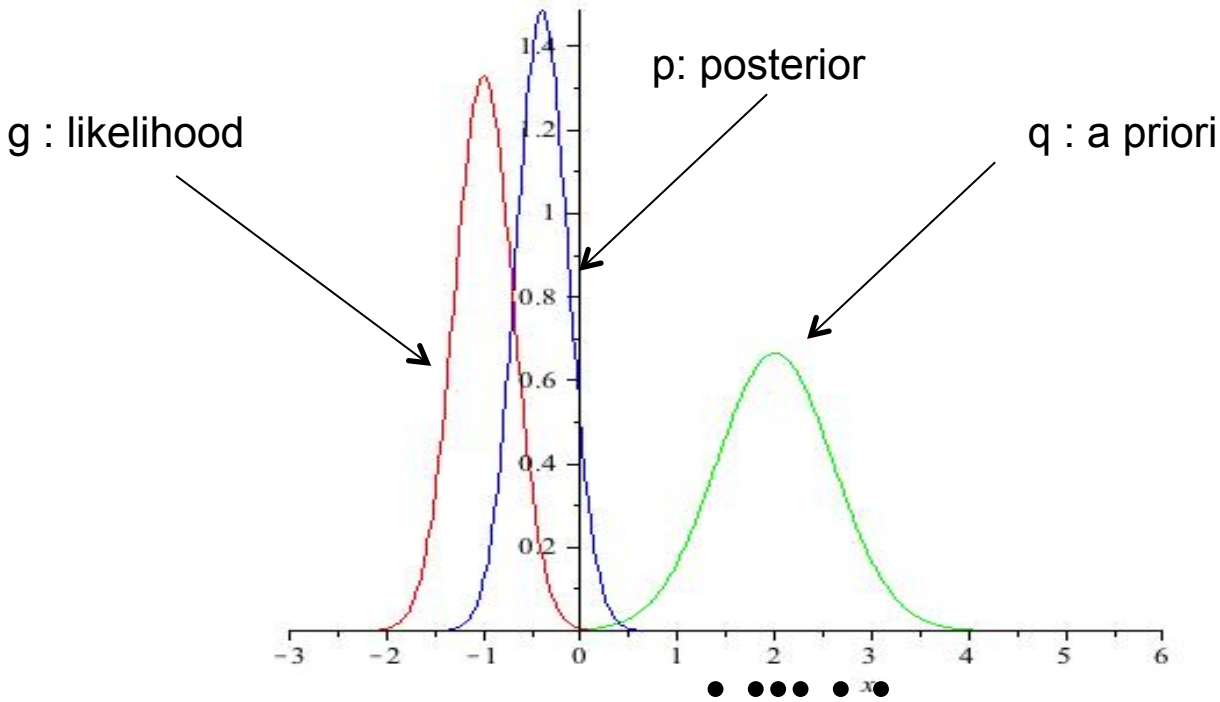
- MRPF (Mixture Regularized Particle Filter) converges faster on terrain with high roughness
- MRPF is more robust on ambiguous terrains
  - → local redistributions
  - → local regularization
- Additional tuning parameters for clustering algorithm
- Reasonable additional computing time when dealing with the same N

Analysis of the Monte Carlo error with the Laplace method

Main error : correction step. Analysis of the local error.

Large error if weak overlapping likelihood/prediction

$$p(x) \propto p(y/x)q(x) = g(x)q(x)$$



Analysis of the Monte Carlo error with the Laplace method

Samping according to q : $(X_1,\ldots,X_N) \sim q$

Weighting : $w_i \propto g(X_i)$

$\{w_i, X_i\}$: approximate sample of p (asymptotically unbiased)

Example : posterior expectation $E[X/Y]$ estimated by $\sum_{i=1}^{N} w_i X_i$

Asymptotic Variance (weight) : reliable criterion for the MC error

$$\frac{1}{N}\underbrace{\frac{\int g^2(x)q(x)\,dx}{(\int g(x)q(x)\,dx)^2}}_{\geq 1} \approx \sum_{i=1}^{N} w_i^2 - \frac{1}{N}$$

Analysis of the Monte Carlo error with the Laplace method

Laplace approximation $(\lambda \to +\infty)$

$$\int_{R^d} e^{-\lambda h(x)}\, dx = (2\pi)^{d/2} e^{-\lambda h(\hat{x})} \det\left[\lambda\, \nabla^2 h(\hat{x})\right]^{-1/2} \underbrace{(1 + O(\lambda^{-1}))}_{\text{Relative error}}$$

$$\hat{x} = \arg\min(h(x))$$

- Global unique minimum, regularity of h
- Formula related to saddle-point, tilted density
- In general very accurate even if λ small

$$\longrightarrow \qquad V = \frac{\int g^2(x) q(x)\, dx}{\left(\int g(x) q(x)\, dx\right)^2}$$

Analysis of the Monte Carlo error with the Laplace method

Thanks to Laplace approximation :

$$V = \frac{\int g^2(x)q(x)\,dx}{(\int g(x)q(x)\,dx)^2} \propto \det\left[\frac{1}{2}J(\hat{x})\right]^{1/2}$$

where $J(x) = -(\log g)''(x) - (\log q)''(x)$ is the information matrix a posteriori

and $\hat{x} = \arg\max\{g(x)q(x)\}$ the MAP

Highly informative model (small measurement std or high measurement variation)

$\longrightarrow$ Large Monte carlo error

$\longrightarrow$ Accurate Laplace approximation (information plays role of λ)

Importance density : shifting and scaling

Find an importance fonction $\tilde{q}$ which reduces the variance

$$V(\tilde{q}) = \frac{\int g^2(x)\frac{q^2(x)}{\tilde{q}(x)}dx}{(\int g(x)q(x)dx)^2}$$

Approximate the optimal (V=1) : $\tilde{q}(x) = \frac{g(x)q(x)}{\int g(x)q(x)dx} = p(x)$

By shifting and scaling:

$$q\left[\frac{x - E(X/Y)}{\sigma(X/Y)}\right] \longrightarrow$$ Easy to sample from

Computing method for posterior expectation

$$E[X/Y] = \frac{\int x\, g(x)\, q(x)\, dx}{\int g(x)\, q(x)\, dx} = \frac{dm_X}{da}(0)$$

$$m_X(a) = E[e^{a^T X}] = \frac{\int e^{a^T x}\, g(x)\, q(x)\, dx}{\int g(x)\, q(x)\, dx} \approx m_X^{Lap}(a) = \frac{e^{a^T \tilde{x}} g(\tilde{x}) q(\tilde{x})}{g(\hat{x}) q(\hat{x})} \left(\frac{\det[J(\hat{x})]}{\det[J(\tilde{x})]} \right)^{1/2}$$

$$\tilde{x} = \tilde{x}(a) = \arg\max \{e^{a^T x} g(x)\, q(x)\} \qquad \hat{x} = \arg\max \{g(x)\, q(x)\} \qquad \tilde{x}(0) = \hat{x}(0)$$

$$\frac{dm_X^{Lap}}{da} = \frac{\partial m_X^{Lap}}{\partial \tilde{x}} \frac{d\tilde{x}}{da} + \frac{\partial m_X^{Lap}}{\partial a}$$

1- $a + \frac{\partial \log g}{\partial x}(\tilde{x}(a)) + \frac{\partial \log q}{\partial x}(\tilde{x}(a)) = 0 \longrightarrow \frac{d\tilde{x}}{da} = -J(\tilde{x}(a))^{-1}$

2- $\frac{\partial \det(J(x))}{\partial x_1} = \det(J(x))\, tr[J^{-1} \frac{\partial J(x)}{\partial x_1}]$

3- $\frac{\partial m_X^{Lap}}{\partial a}(0) = \hat{x}$

Importance density : a method to compute posterior moments

Computing method for posterior expectation

$$E[X/Y] = \frac{\int x\, g(x)\, q(x)\, dx}{\int g(x)\, q(x)\, dx}$$

$$E[X/Y] \approx \hat{x} - \frac{1}{2} J^{-1}(\hat{x}) \underbrace{\begin{pmatrix} tr[J^{-1}(\hat{x}) \frac{\partial J}{\partial x_1}(\hat{x})] \\ . \\ tr[J^{-1}(\hat{x}) \frac{\partial J}{\partial x_d}(\hat{x})] \end{pmatrix}}_{\text{Related to skewness}}$$

$$V[X/Y] \approx J^{-1} + \frac{1}{2} J^{-1} \left[\frac{dJ}{dx}\right]^T (J^{-1} \otimes J^{-1}) \frac{dJ}{dx} J^{-1} + \frac{1}{2}\left(I_d \otimes vec(J^{-1})^T \frac{dJ}{dx}\right)(J^{-1} \otimes J^{-1}) \frac{dJ}{dx} J^{-1} - \frac{1}{2} J^{-1}\left(I_d \otimes vec(J^{-1})^T\right) \frac{d^2 J}{dx^2} J^{-1}$$

Accurate approximation as function of the MAP

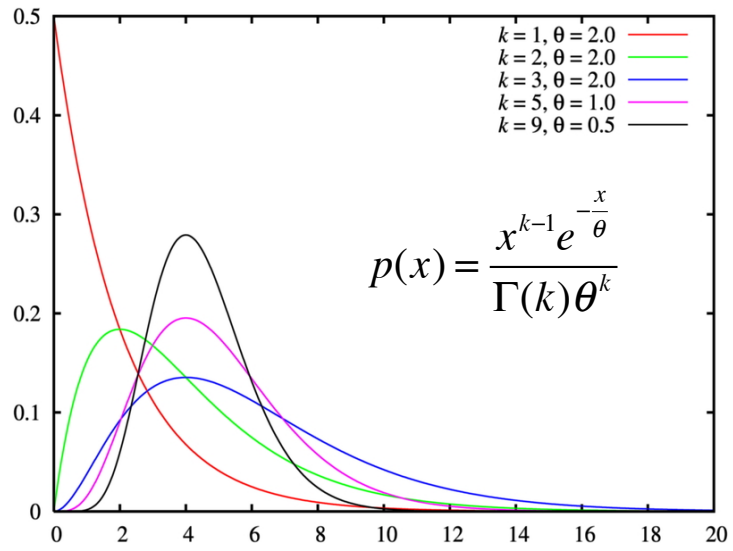
Importance density : a method to compute posterior moments

Computing method for posterior expectation, an example : gamma distribution

$$J = -\frac{d^2 \log(p)}{dx^2} = \frac{k-1}{x^2}$$

$$\frac{dJ}{dx} = \frac{-2(k-1)}{x^3}$$

$$\frac{d^2 J}{dx^2} = \frac{6(k-1)}{x^4}$$

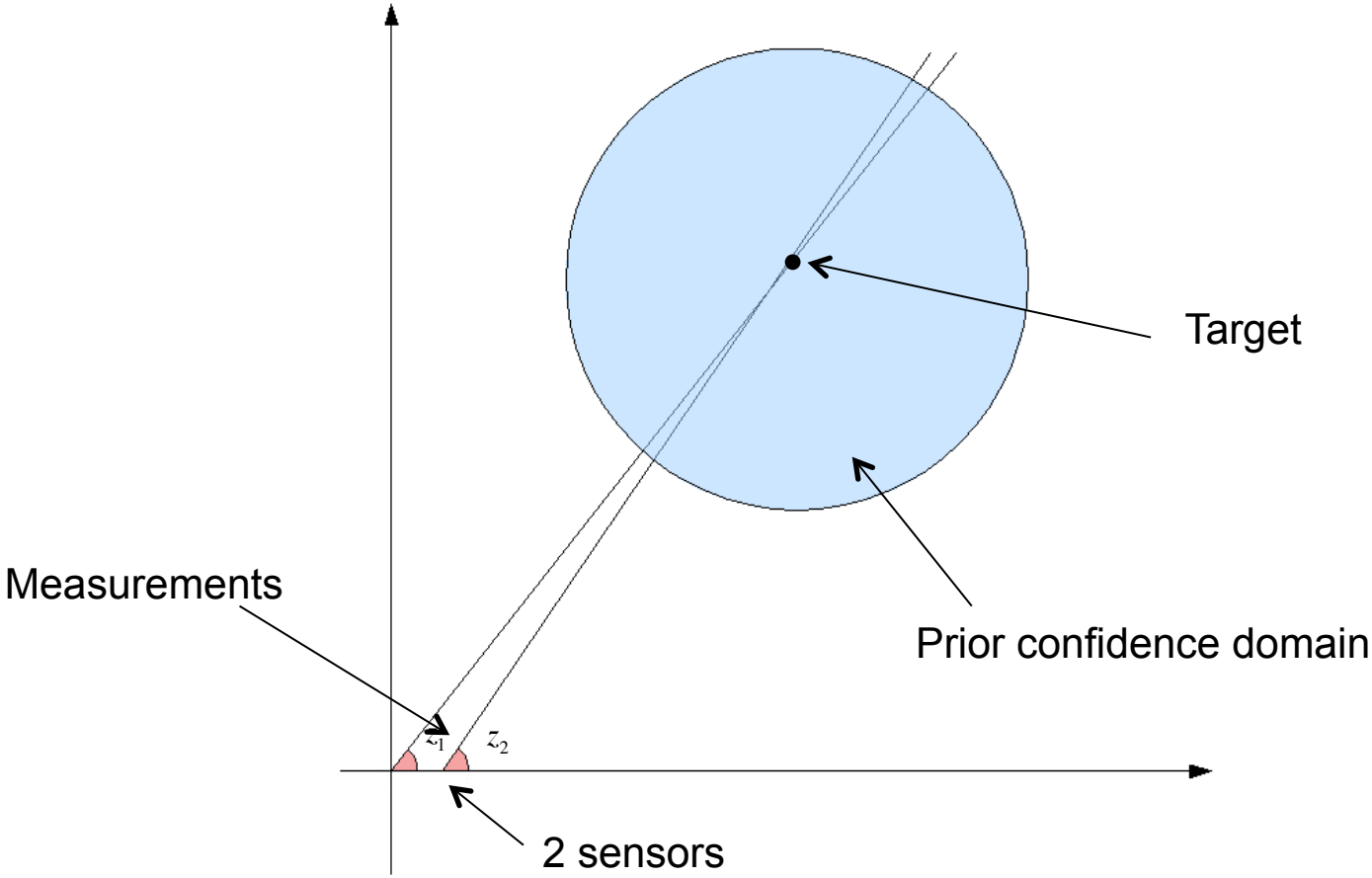


$$E[X/Y] \approx \hat{x} - \frac{1}{2} J^{-2}(\hat{x}) \frac{dJ}{dx}(\hat{x}) = k\theta$$

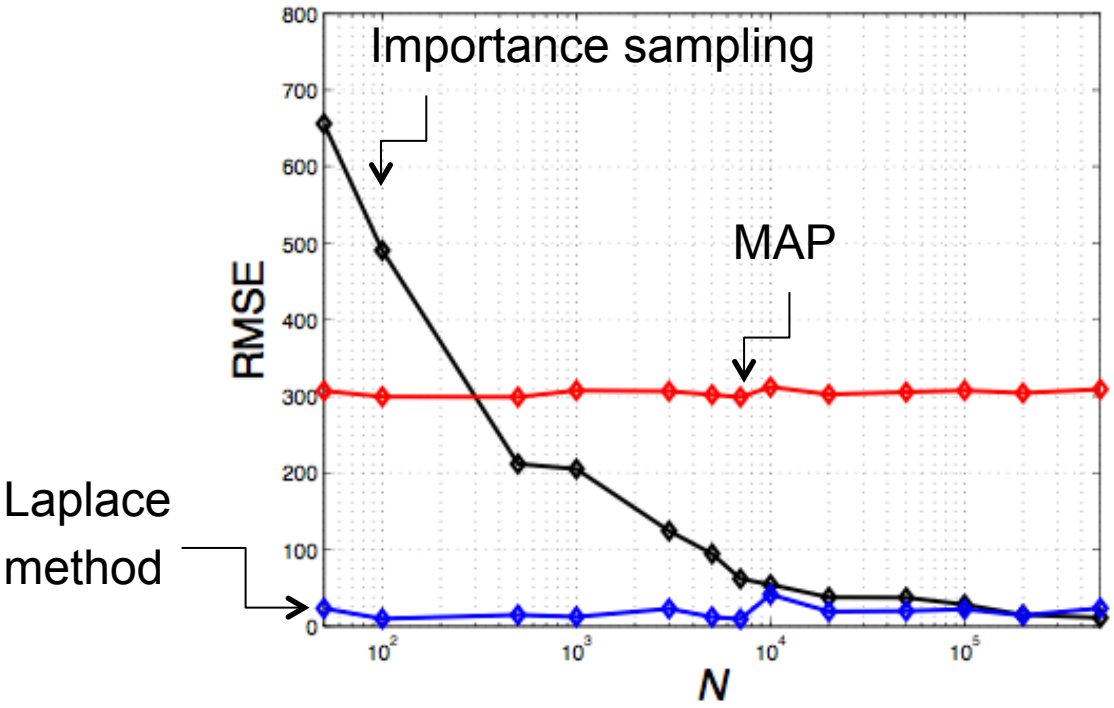
$$V[X/Y] \approx \frac{J^{-4}(\hat{x})}{2} \left(2J^3 + 2\left(\frac{dJ}{dx}\right)^2 - J\frac{d^2 J}{dx^2} \right)_{x=\hat{x}} = k\theta^2$$

True values

An other example : triangulation



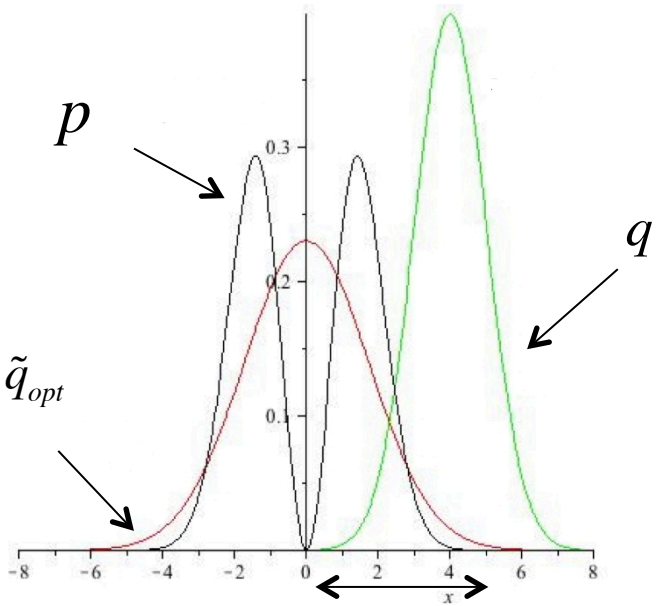
Importance density : a method to compute posterior moments



Gain in terms of accuracy and computing time

Importance density : shifting and scaling the prior

Illustration on a simple example $p(x) = x^2 e^{-\frac{x^2}{2\theta^2}}$ $\tilde{q}(x) = e^{-\frac{(x-\mu)^2}{2\sigma^2}}$



Optimal shift and scale ?

$\mu_{opt} = 0 \quad \sigma_{opt} = \sqrt{3}\,\theta$ (coincide with the posterior mean and std).

Asympt. variance = 1.5 (optimal = 1)

Warning : strong degradation of the asympt. var. if $\sigma < \sigma_{opt}$

Application to PF

q a priori : predicted density considered for the MAP as a Gaussian

$$p(x) \propto p(y/x)q(x) = g(x)q(x)$$

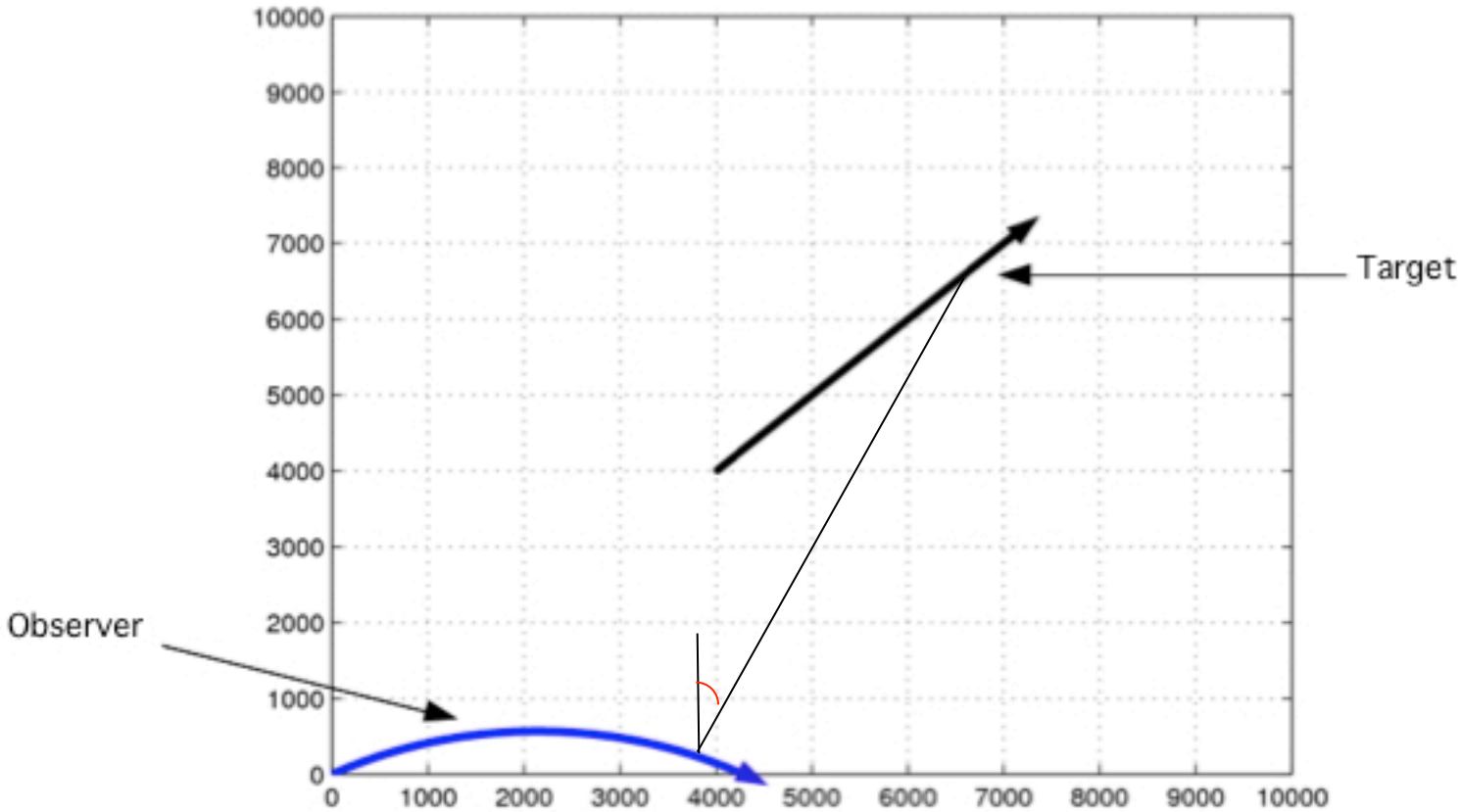
Depends on a small number of components of the state vector

Computed with the predicted particles

→ MAP computed on a low-dimensional state space

Application to tracking

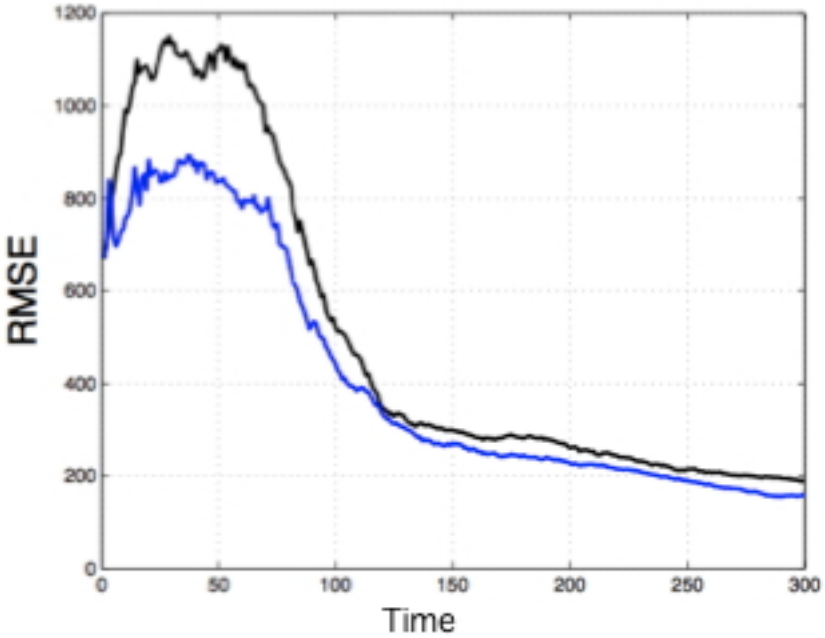
Bearing only application with accurate measurements



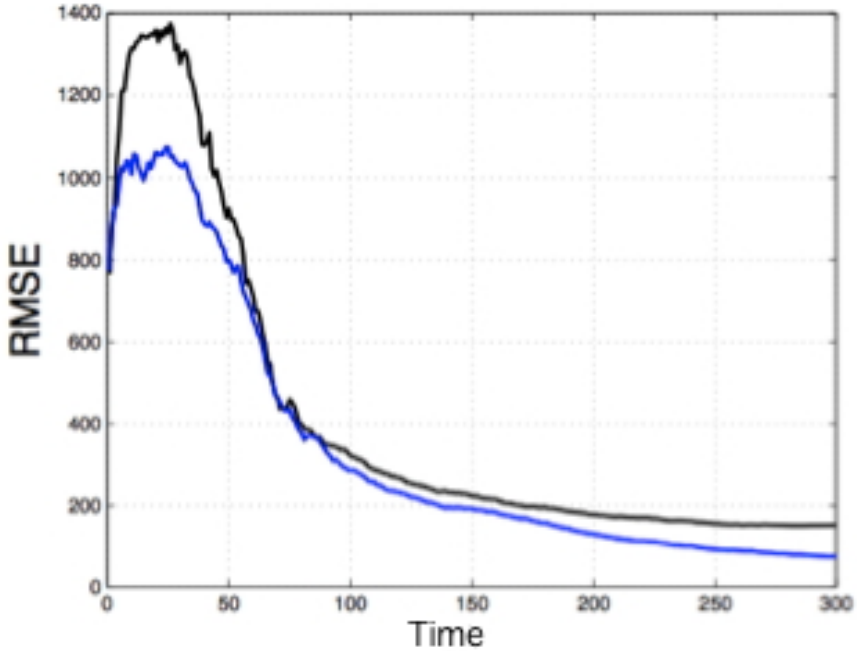
Application to tracking

$$N = 3000 \quad \sigma = 0.05°$$

Root Mean Square Error : x-position



Root Mean Square Error : y-position



With Laplace (LRPF): 0% divergence

Without Laplace (RPF) : 12% divergence

Improved particle filters

- Robustness in case of multimodality
- With a shifted/rescaled importance function

- Bui Quang, P., Musso, C, Le Gland, F. (2010). « An insight Into the issue of dimensionality in particle filtering ». Fusion 2010.
- Murangira, A., Musso, C., Dahia, K.; Allard, JM. (2011). « Robust regularized particle filter for terrain navigation ». Fusion 2011.
- Musso, C., Bui Quang, P., Le Gland, F. (2011). « Introducing the Laplace approximation in particle filtering ». Fusion 2011.