IAP Study Day, Mons - May 24, 2013

Particle filtering

Sequential Monte Carlo
SMC

Fabien Campillo
Montpellier

general outline

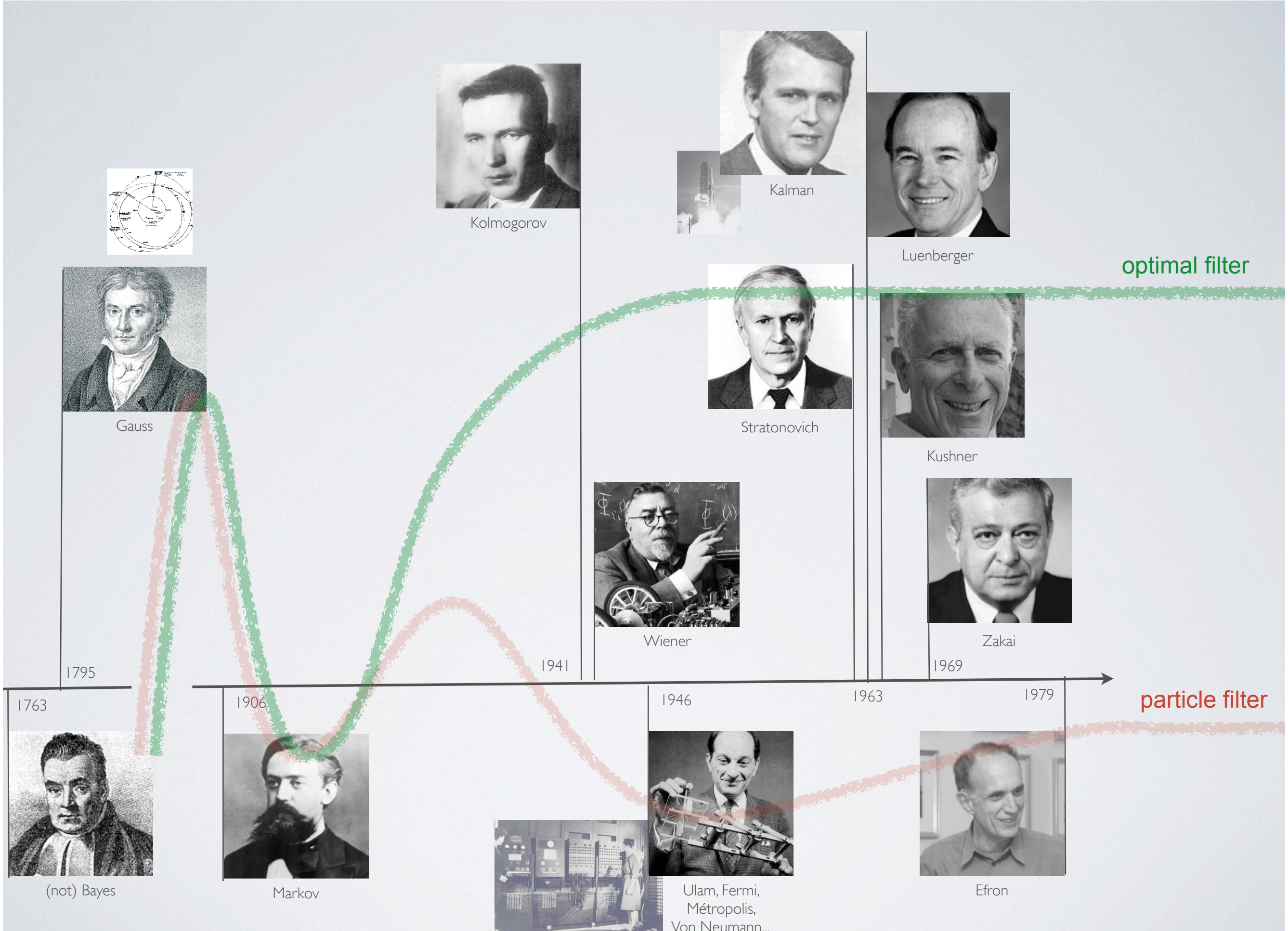
content

goal

- we consider a dynamical system that we don't observe directly, we have observations Y_k of the system at time t_k that are partial and corrupted by noise and we want to infer the state of the dynamical system
- in this approach we need to build:
  - a state space model:
    - a state model X_k
    - an observation model to rely X_k to Y_k
  - an estimator $\hat{X}_k$ of X_k as a function $\hat{X}_k = \hat{X}_k(Y_{1:k})$ of the past observations $Y_{1:k} = (Y_1, \ldots, Y_k)$ that minimizes the mean squared error (MSE)

$$\text{MSE}(\hat{X}_k) = \mathbb{E}[|X_k - \hat{X}_k|^2]$$

- on-line processing $\Rightarrow$ recursive processing (filtering) i.e. the estimate $\hat{X}_k$ should be updated from $\hat{X}_{k-1}$ using only the last available observation Y_k (opposed to batch processing)
- observers, software sensors, hidden Markov models etc.



optimal filter
particle filter
Kolmogorov
Kalman
Luenberger
Gauss
Stratonovich
Kushner
Wiener
Zakai
1795
1941
1969
1763
1906
1946
1963
1979
(not) Bayes
Markov
Ulam, Fermi,
Métropolis,
Von Neumann,..
Efron

content

joint distribution

- behavior of the couple of random variables (X,Y) determined by its distribution (probability density function) $p_{X,Y}(x,y)$

$$\mathbb{P}(X \in A, Y \in B) = \iint_{A\times B} p_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \qquad \forall A, B$$
$$\mathbb{E}[\phi(X,Y)] = \iint \phi(x,y)\,p_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \qquad \forall \phi$$

  and $p_{X,Y} \geq 0$, $\iint p_{X,Y}\,\mathrm{d}x\,\mathrm{d}y = 1$

- prior to any observation, our knowledge on X is its (marginal) distribution

$$p_X(x) \stackrel{\text{def}}{=} \int p_{X,Y}(x,y)\,\mathrm{d}y$$

  but *after observing $Y = y$ what is our knowledge about X ?*

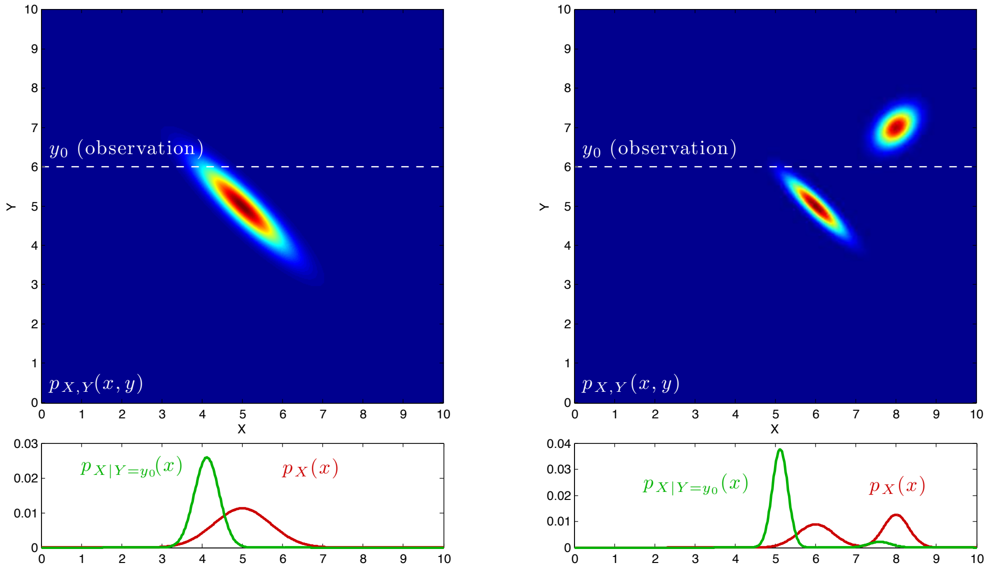
conditional distribution I

- the answer is given by conditional distribution of X given $Y = y$:

$$p_{X|Y=y}(x) \stackrel{\text{def}}{=} \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{p_{X,Y}(x,y)}{\int_{\mathbb{R}^d} p_{X,Y}(x',y)\,\mathrm{d}x'}$$

  it is *the distribution that contains all the information on X given by $p_{X,Y}(x,y)$ (the model) and $Y = y$ (the observation)*

conditional distribution II



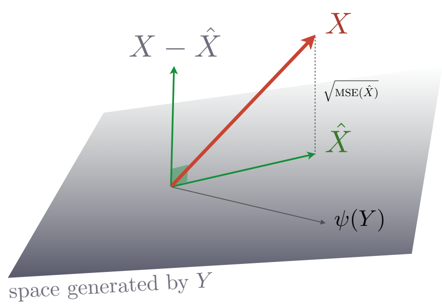
estimator of X given $Y = y$

- the conditional mean

$$\hat{X}(y) \stackrel{\text{def}}{=} \mathbb{E}[X|Y=y] = \int x\, p_{X|Y=y}(x)\, \mathrm{d}x$$

minimizes the mean square error:

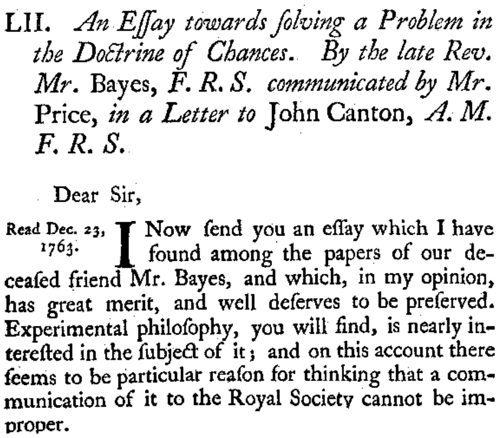
$$\mathrm{MSE}(\psi) \stackrel{\text{def}}{=} \mathbb{E}[|\psi(Y) - X|^2]$$



  - prior to observation the best estimation of X is its mean $\mathbb{E}(X)$, posteriori to the observation it is its conditional mean $\hat{X}(y)$
  - to compute $\hat{X}(y)$ we need to know the conditional distribution $p_{X|Y=y}(x)$
  - *the goal is $p_{X|Y=y}(x)$, it gives the estimation of X, the estimation of the error etc.*

Bayes formula I

Ceci n'est pas le révérend Thomas Bayes



LII. *An Eſſay towards ſolving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *F. R. S. communicated by Mr.* Price, *in a Letter to* John Canton, *A. M. F. R. S.*

Dear Sir,

Read Dec. 23, 1763.

I Now ſend you an eſſay which I have found among the papers of our deceaſed friend Mr. Bayes, and which, in my opinion, has great merit, and well deſerves to be preſerved. Experimental philoſophy, you will find, is nearly intereſted in the ſubject of it; and on this account there ſeems to be particular reaſon for thinking that a communication of it to the Royal Society cannot be improper.

Bayes formula II

- learning process: from the prior knowledge $p_X(x)$ to the posterior knowledge $p_{X|Y=y}(y)$ by integrating "$Y = y$"
- as $p_{X,Y}(x,y) = p_{X|Y=y}(x)\,p_Y(y) = p_{Y|X=x}(y)\,p_X(x)$ then

$$p_{X|Y=y}(x) = \frac{p_{Y|X=x}(y)\,p_X(x)}{p_Y(y)} = \frac{p_{Y|X=x}(y)\,p_X(x)}{\int p_{Y|X=x}(y)\,p_X(x)\,\mathrm{d}x}$$

denoted

$$p_{X|Y=y}(x) \propto \underbrace{p_{Y|X=x}(y)}_{L(y|x)} \times p_X(x)$$

posterior distribution $\propto$ likelihood $\times$ prior distribution

$L(y|x)$ measures the adequacy of x to y

Bayes formula III

$$p_{X|Y=y}(x) \propto \underbrace{p_{Y|X=x}(y)}_{L(y|x)} \times p_X(x)$$

- "bonnet blanc/blanc bonnet" what do we gain so far ?
- $p_{X,Y}(x,y)$ is equivalent to $[p_X(x), L(y|x)]$ but the latter is easier to understand than the former:
  - $p_X(x)$ state model: how the hidden state process behaves
  - $L(y|x)$ observation model: causal link between the state and the observation, example:

$$Y = h(X) + V \qquad V \sim N(0, \sigma^2)$$

    then $L(y|x) \propto \exp(-\frac{1}{2}|y - h(x)|^2/\sigma^2)$

Bayes formula IV

- to compute $\hat{X}$ we need to make 2 integrations:

$$\hat{X}(y) = \int x \, p_{X|Y=y}(x) \, \mathrm{d}x = \int x \, \frac{L(y|x)}{c(y)} \, p_X(x) \, \mathrm{d}x$$

  with $c(y) = \int L(y|x) \, p_X(x) \, \mathrm{d}x$, which is impossible analytically in most cases

- Bayes had to wait for the Manhattan project, the ENIAC, the spread of computers (...) the 70's $\rightarrow$ Monte Carlo, empirical numerics approaches

content

Monte Carlo I

- sample $\xi^1, \dots, \xi^N$ from a distribution $p_X(x)$ of a r.v. X, then:

$$\mathbb{E}[\phi(X)] \simeq \frac{1}{N} \sum_{i=1}^{N} \phi(\xi^i)$$

i.e.

$$p_X(x) \simeq p_X^N(x) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi^i}(x)$$

where $\delta_{\xi^i}(x)$ is the Dirac function $\int \phi(x)\, \delta_{\xi^i}(x)\, \mathrm{d}x = \phi(\xi^i)$

- by the law of large numbers $\frac{1}{N} \sum_{i=1}^{N} \phi(\xi^i) \to \mathbb{E}[\phi(X)]$ and the speed of cv is given by the central limit theorem

$$\frac{1}{N} \sum_{i=1}^{N} \phi(\xi^i) \simeq \mathcal{N}\Big(\mathbb{E}[\phi(X)], \frac{\sigma_X^2}{\sqrt{N}}\Big)$$

Monte Carlo II

- importance sampling:

$$\widehat{\phi(X)} \stackrel{\text{def}}{=} \mathbb{E}[\phi(X)|Y=y] = \int \phi(x)\, p_{X|Y=y}(x)\, \mathrm{d}x$$
$$= \int \phi(x)\, \frac{L(y|x)}{c(y)}\, p_X(x)\, \mathrm{d}x$$

it is difficult to sample from $p_{X|Y=y}(x)$, so:

1. sampling: $\xi^1, \ldots, \xi^N \sim p_X(x)$ (prior to observation)
2. weighting: $\omega^i = L(y|\xi^i)$ (adequacy to obs.)
3. normalizing: $\bar{\omega}^i = \omega^i / \sum_{i'} \omega^{i'}$
4. estimator: $\widehat{\phi(X)}^N = \sum_{i=1}^N \bar{\omega}^i\, \phi(\xi^i)$

content

state space model

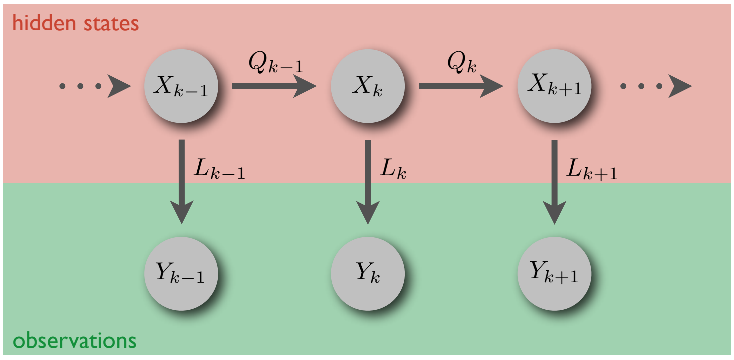
- $(X_k)_{k\geq 0}$ Markov process on $\mathbb{R}^n$ with transition kernel

$$Q_k(x'|x) = p_{X_k|X_{k-1}=x}(x') \qquad \text{(state)}$$

- Y_k observations on $\mathbb{R}^d$: likelihood $p_{Y_k|X_k=x}(y)$

$$L_k(y|x) \propto p_{Y_k|X_k=x}(y) \qquad \text{(observation)}$$

- memoryless channel hypothesis



examples

- discrete time linear/Gaussian model:

$$X_{k+1} = F\,X_k + G\,W_k\,, \quad Y_k = H\,X_k + V_k\,, \quad W_k,\, V_k,\, X_0 \text{ Gausian}$$

- discrete time nonlinear/non-Gaussian model:

$$X_{k+1} = f(X_k, W_k)\,, \qquad Y_k = h(X_k, V_k)$$

- continuous-time state process nonlinear/non-Gaussian model:

$$\dot{Z}_t = f(Z_t) + g(Z_t)\,\beta_t\,, \quad Y_k = h(Z_{t_k}, V_k)\,, \quad \beta_t \text{ white noise}$$

- remarks:
  - in all these cases we can determine Q_k, L_k
  - Q_k can be quite complex but in in SMC we don't need an analytical representation of Q_k we just need to simulate according to it (simulate the state process)

content

nonlinear filter

- we define:
  - filter: distribution of X_k given $Y_{1:k} = y_{1:k}$

$$\pi_k(x) \stackrel{\text{def}}{=} p_{X_k|Y_{1:k}=y_{1:k}}(x) \qquad \text{law}(X_k|Y_{1:k})$$

  - predicted filter: distribution of X_k given $Y_{1:k-1} = y_{1:k-1}$

$$\pi_{k^-}(x) \stackrel{\text{def}}{=} p_{X_k|Y_{1:k-1}=y_{1:k-1}}(x) \qquad \text{law}(X_k|Y_{1:k-1})$$

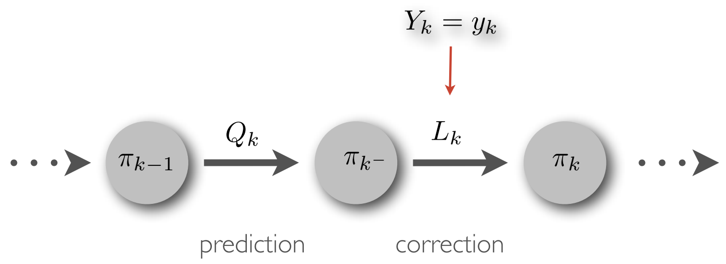
optimal filter

- iteration $\pi_{k-1} \to \pi_k$ in two steps

$$\pi_{k^-}(x) = \int Q_k(x'|x)\, \pi_{k-1}(x')\, \mathrm{d}x' \qquad \text{(prediction)}$$

$$\pi_k(x) \propto \frac{L_k(y_k|x)\, \pi_{k^-}(x)}{\int L_k(y_k|x')\, \pi_{k^-}(x')\, \mathrm{d}x'} \qquad \text{(correction)}$$

$$\hat{X}_k = \int x\, \pi_k(x)\, \mathrm{d}x \qquad \mathrm{MSE} = \int |x - \hat{X}_k|^2\, \pi_k(x)\, \mathrm{d}x$$



- remarks:
  - a lot of integrations on the state space !
  - sequential Bayes formula (dynamic learning procedure)

content

Kalman filter

- explicit solution: (almost) only in the linear/Gaussian case where $\pi_k = N(\hat{X}_k, R_k)$ et $\pi_{k^-} = N(\hat{X}_{k^-}, R_{k^-})$

Kalman-Stratonovitch filter

$\hat{X}_0 = \bar{X}_0$, $R_0 = Q_0$ {initialization}

for $k = 1, 2, 3 \dots$ **do**

$\hat{X}_{k^-} = F_{k-1}\, \hat{X}_{k-1}$ {prediction}

$R_{k^-} = F\, R_{k-1}\, F^* + G\, Q^W\, G^*$

$K_k = R_{k^-}\, H^*\, [H\, R_{k^-}\, H^* + Q^V]^{-1}$ {correction}

$\hat{X}_k = \hat{X}_{k^-} + K_k\, [Y_k - H\, \hat{X}_{k^-}]$

$R_k = [I - K_k\, H]\, R_{k^-}$

end for

content

“Gaussian-oriented” approximations

- extended Kalman filter: linearize the nonlinear system around the current estimate and use the Kalman approach
- 1994: ensemble Kalman filter (EnKF) of Evensen
  - sequential data assimilation
  - version of the Kalman filter for large problems where the covariance matrix is replaced by the sample covariance (no need to keep it in memory)
- 2000: unscented Kalman filter (UKF) of Julier and Uhlmann:
  - avoid the computation of the gradients in the EKF
  - based on quadrature formulas (not Monte Carlo)
  - great alternative to EKF (faster/lighter)

content

particle approximation

- Monte Carlo approximation of the filter, i.e.

$$\pi_k(x) \simeq \pi_k^N(x) \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^N \delta_{\xi_k^i}(x)$$

where $\xi_k^1, \dots, \xi_k^N$ are sampled from π_k, or

$$\pi_k(x) \simeq \pi_k^N(x) \stackrel{\text{def}}{=} \sum_{i=1}^N \omega_k^i \, \delta_{\xi_k^i}(x)$$

where $\xi_k^1, \dots, \xi_k^N$ are sampled from another distribution (importance sampling)

a first (almost) good idea

$\xi_0^1, \dots, \xi_0^N \overset{\text{iid}}{\sim} p_{X_0}$, $\omega_0^i = 1$ {initialization}

for $k = 1, 2, 3 \dots$ **do**

  $\xi_k^i \sim Q_k(\,\cdot\,|\xi_{k-1}^i)$, $i = 1 \cdots N$ {particle evolution}

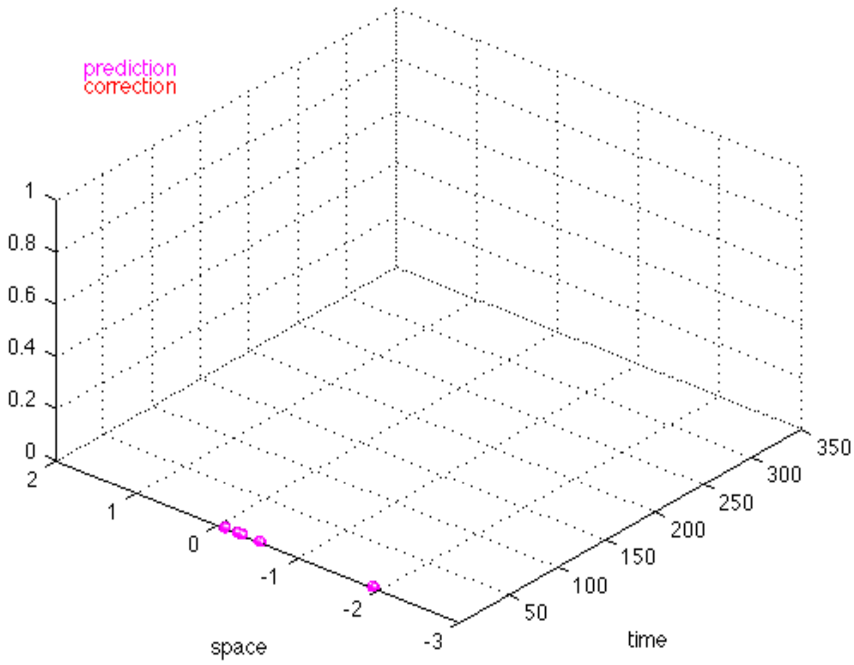
  $\omega_k^i = L_k(y_k, \xi_k^i)\, \omega_k^i$, $i = 1 \cdots N$ {weighting}

  $\omega_k^i = \omega_k^i / \sum_{i'} \omega_k^{i'}$ {normalization}

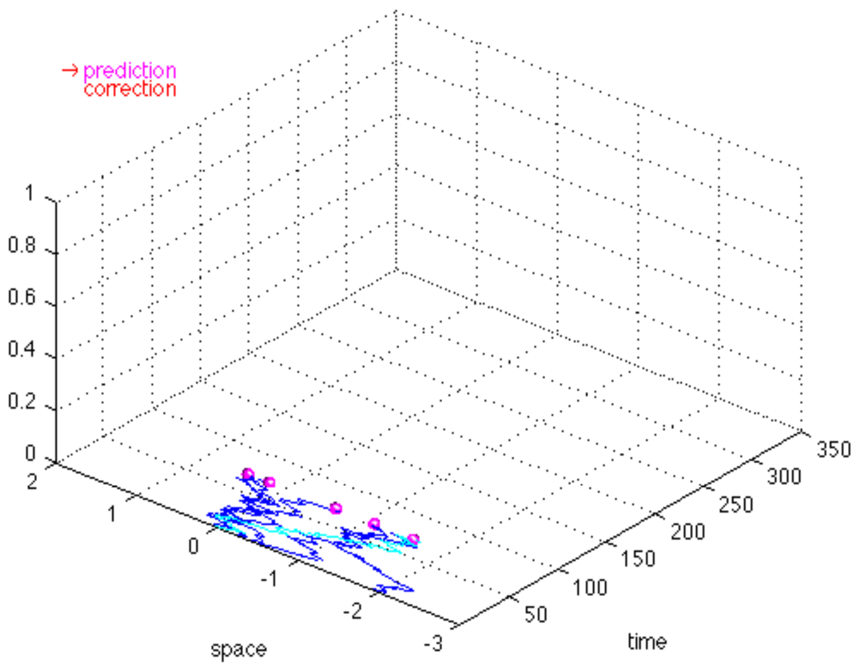
end for

- here $\pi_k(x) \simeq \pi_k^N(x) = \sum_{i=1}^N \omega_k^i \, \delta_{\xi_k^i}(x)$
- it converges mathematically: $\pi_k^N(x) \to \pi_k(x)$ as $N \to \infty$
- but practically it fails completely !!!

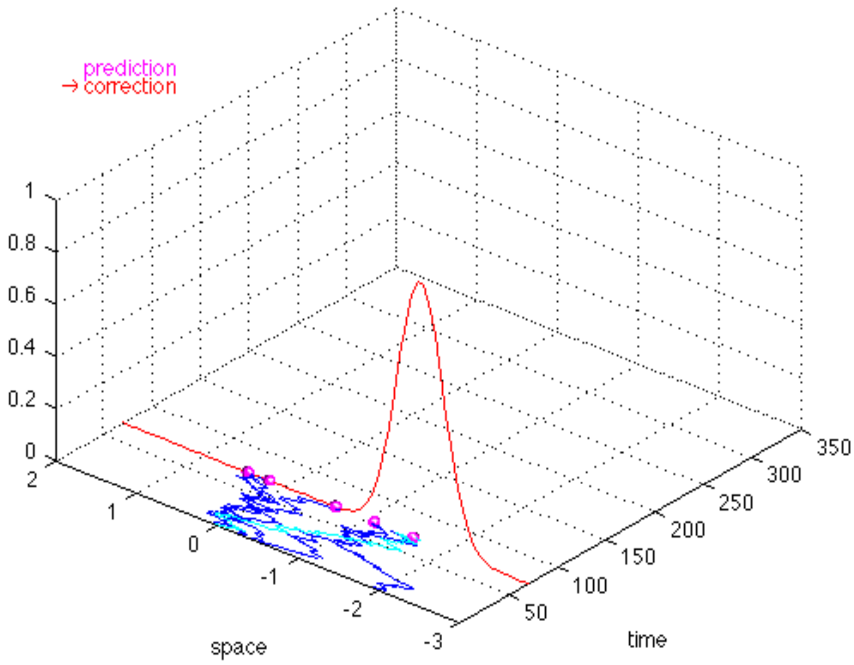
particle degeneracy



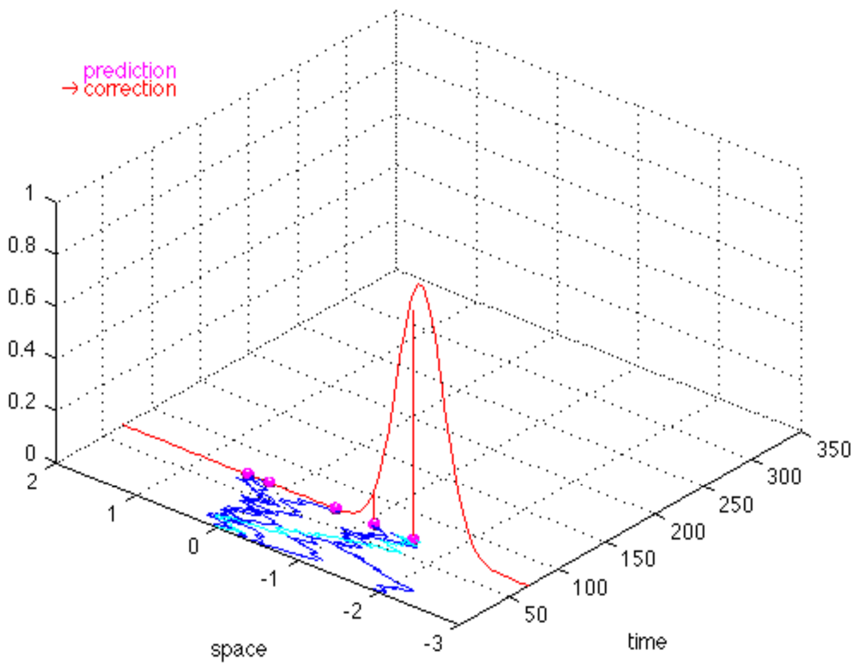
particle degeneracy



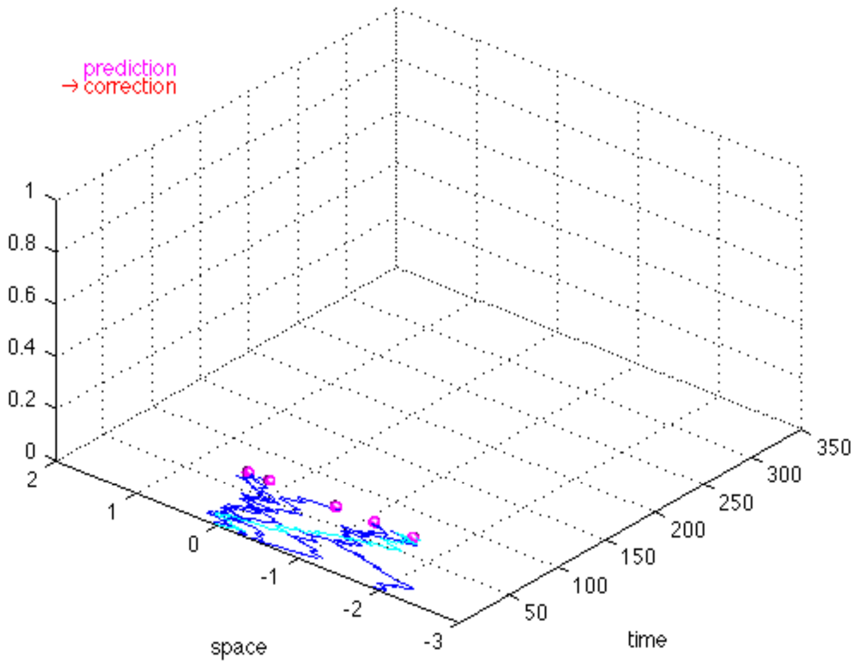
particle degeneracy



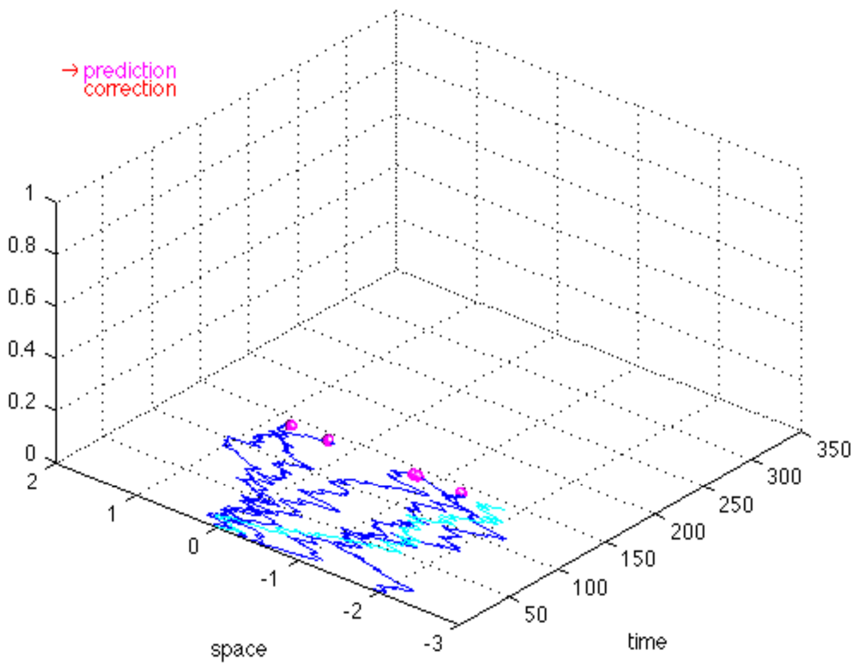
particle degeneracy



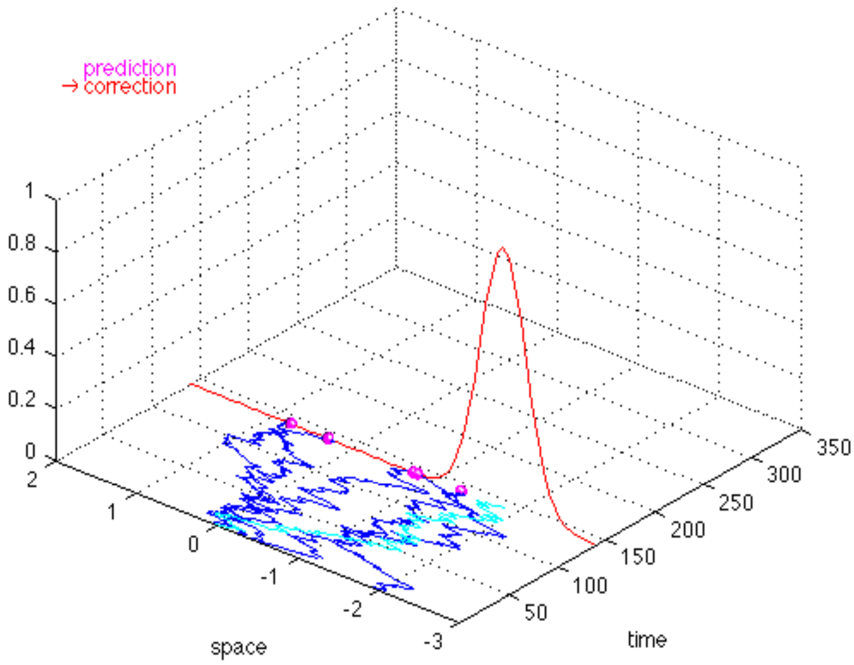
particle degeneracy



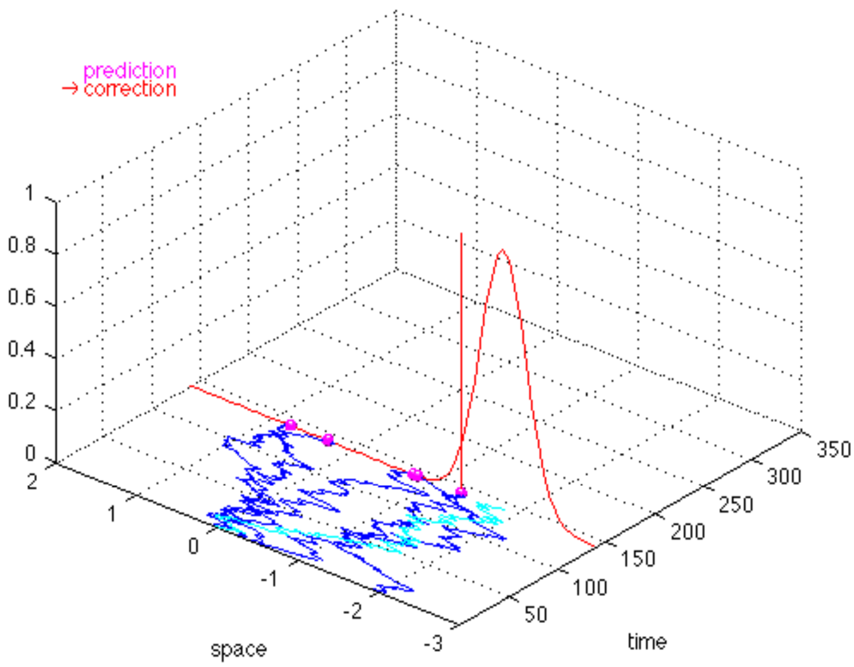
particle degeneracy



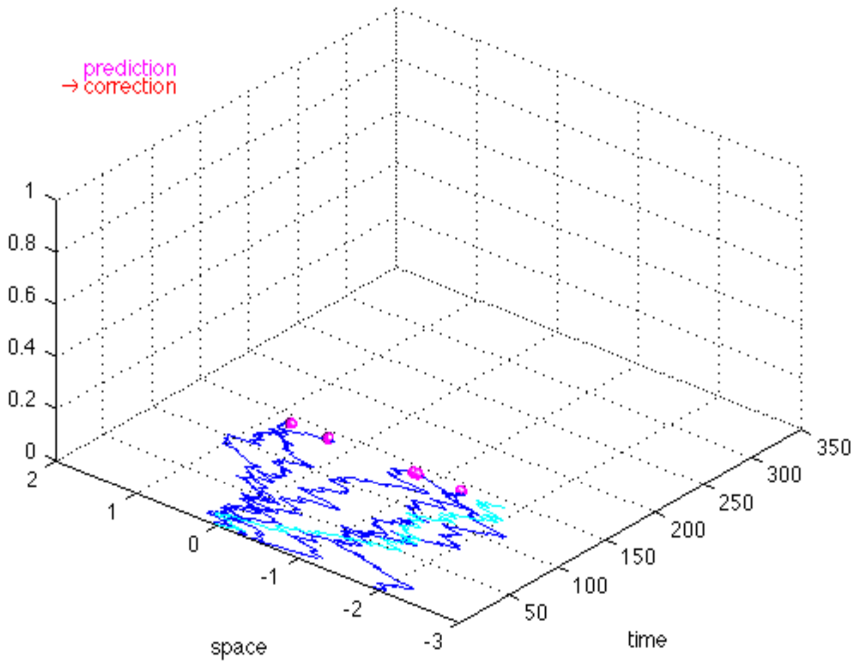
particle degeneracy



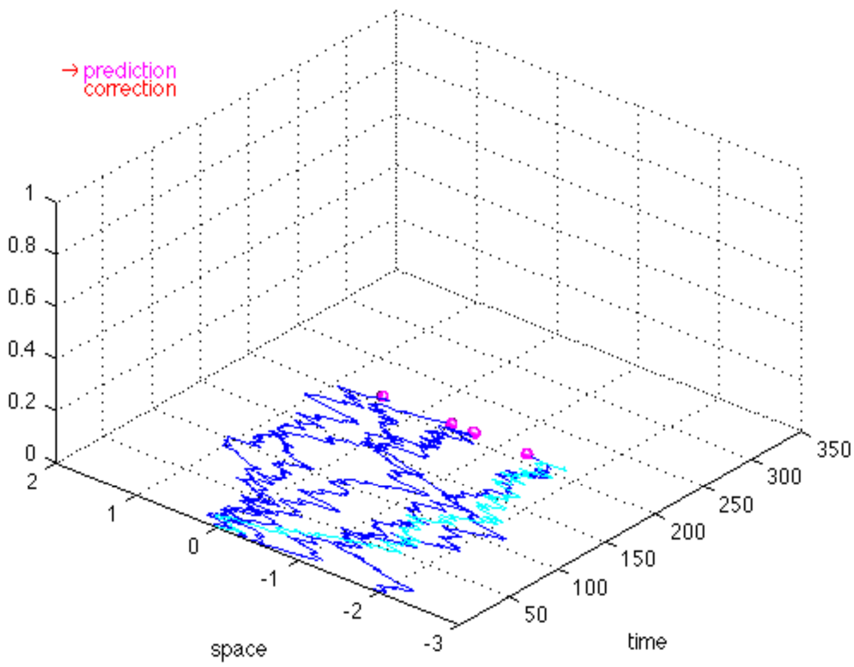
particle degeneracy



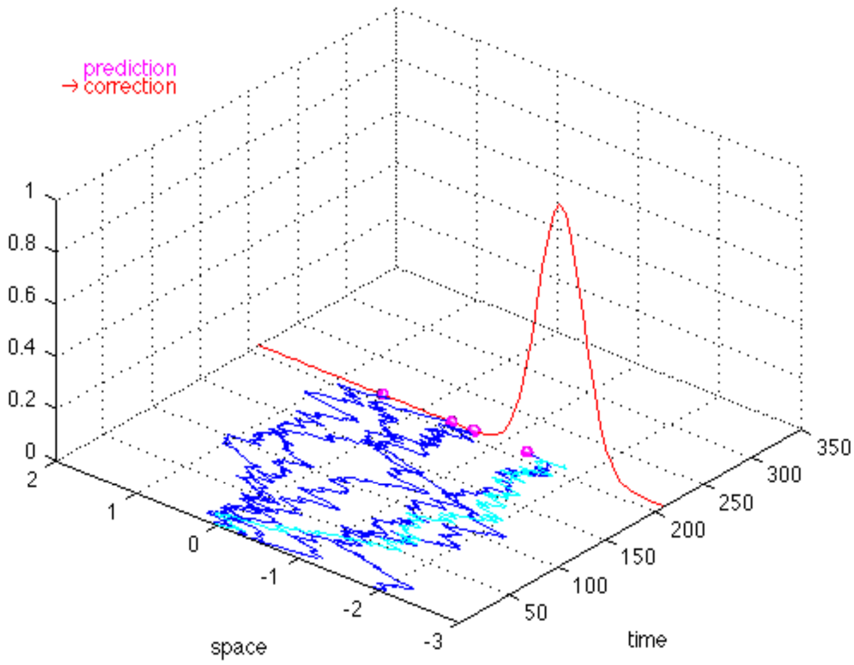
particle degeneracy



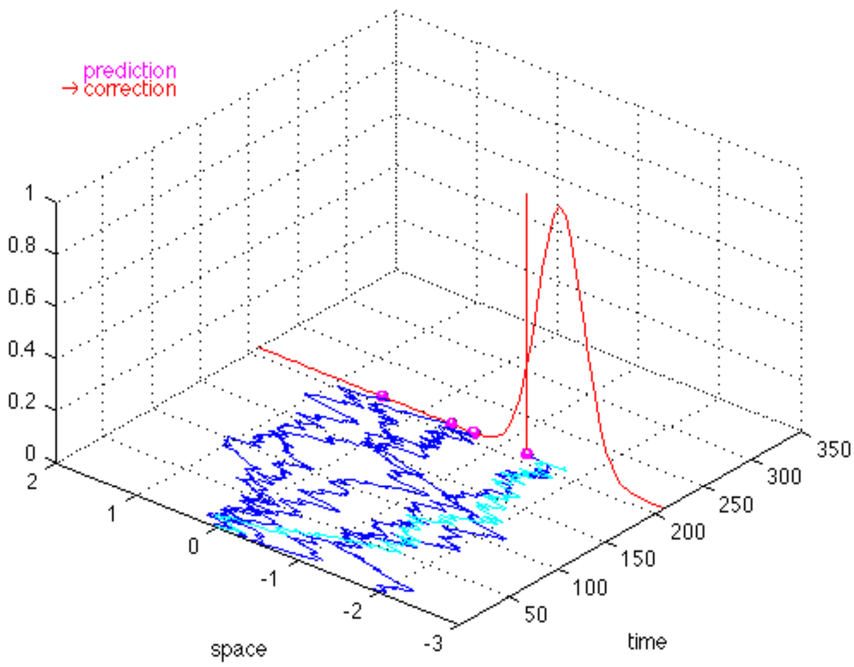
particle degeneracy



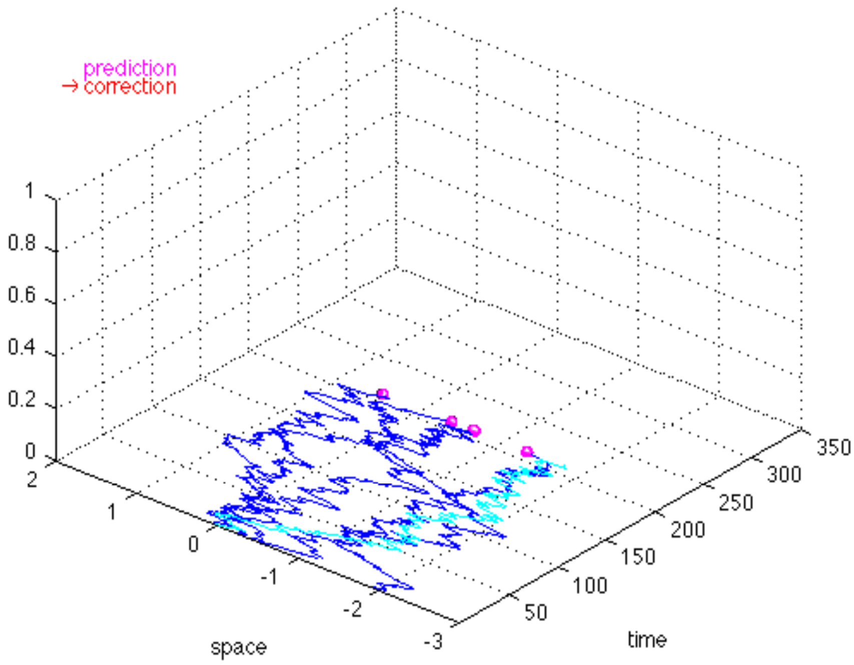
particle degeneracy



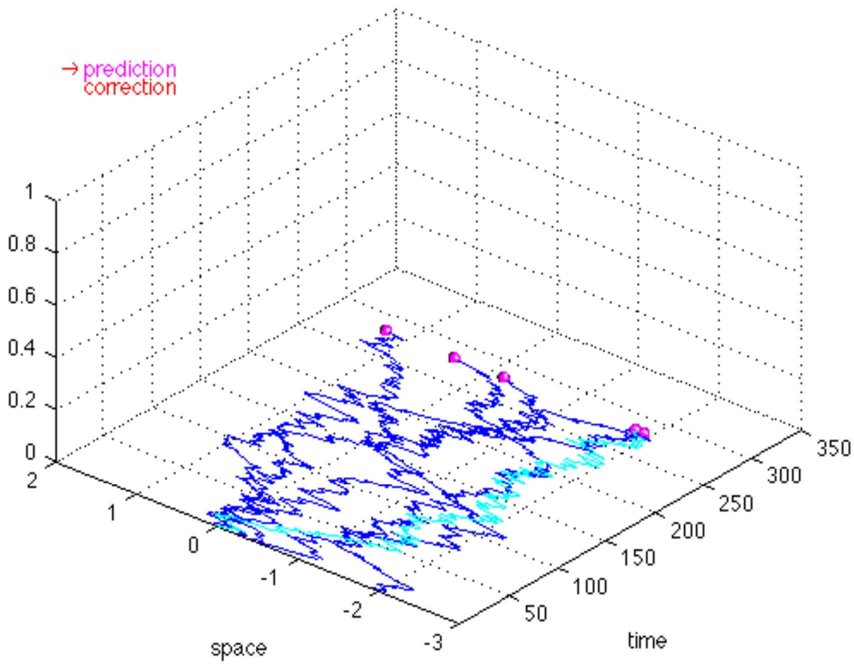
particle degeneracy



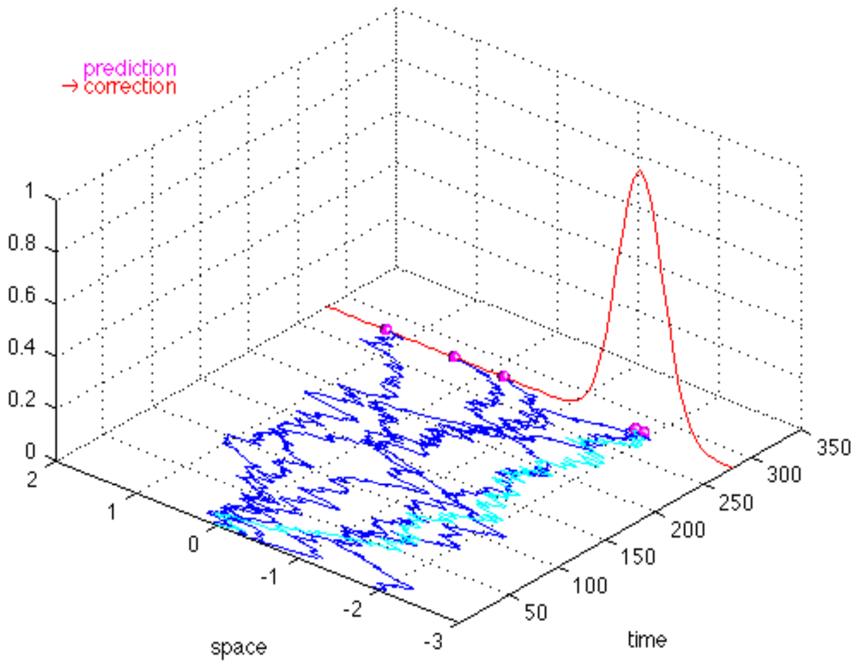
particle degeneracy



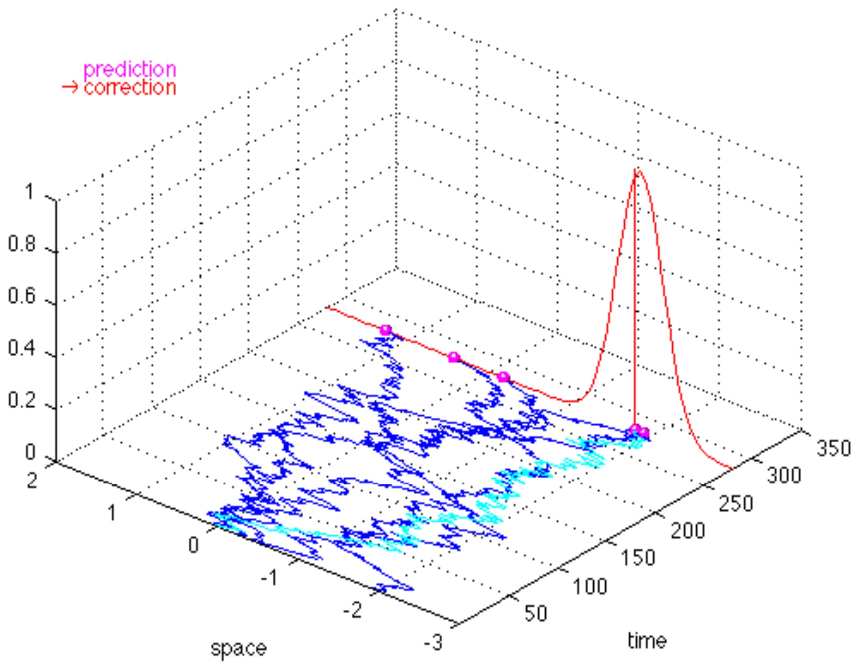
particle degeneracy



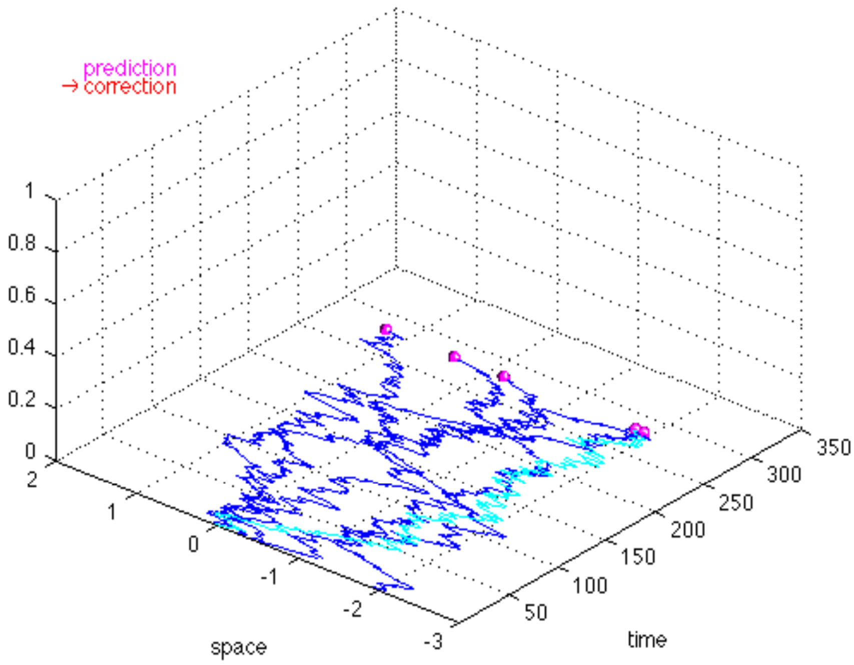
particle degeneracy



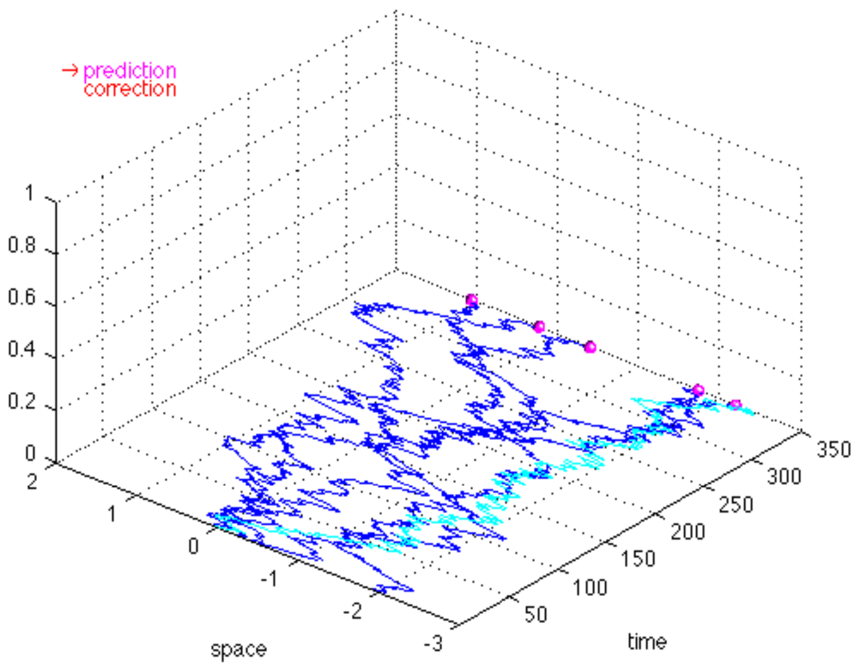
particle degeneracy



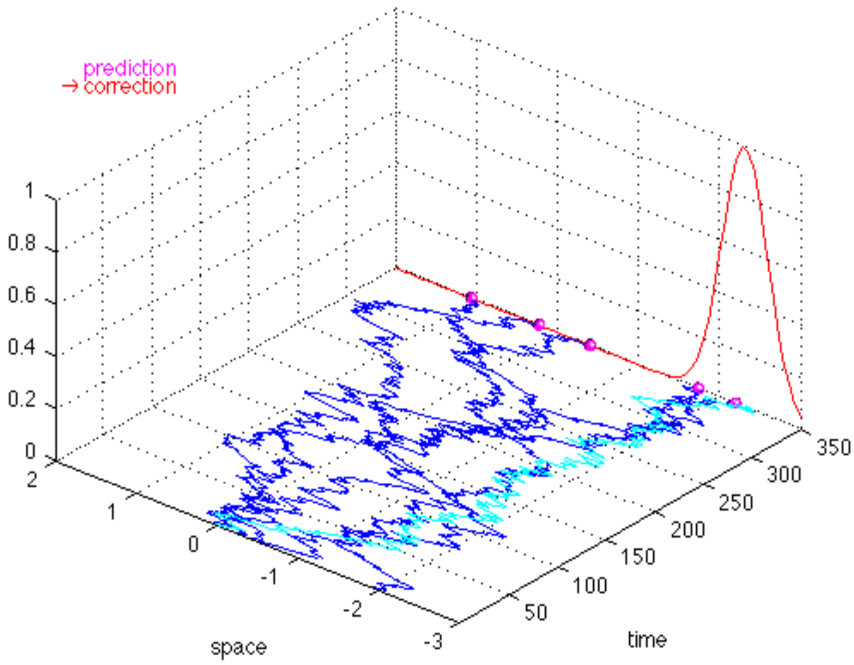
particle degeneracy



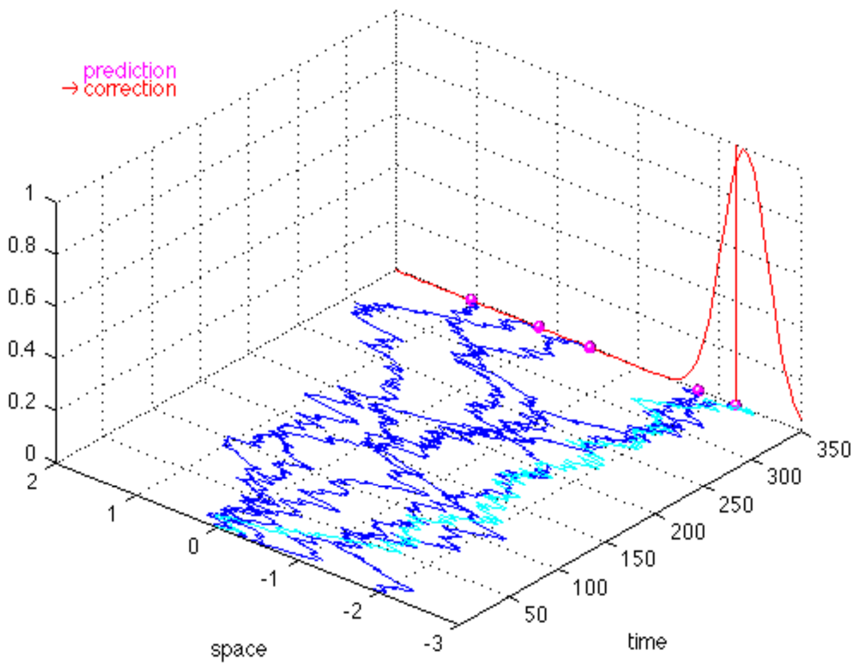
particle degeneracy



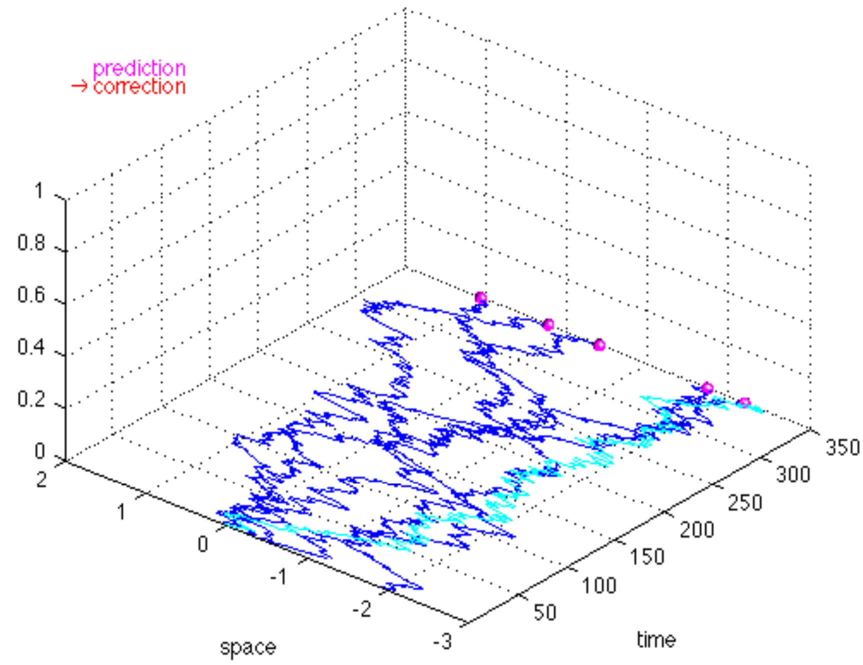
particle degeneracy



particle degeneracy



particle degeneracy



what do we miss ?

- particle degeneracy: in few time step all the weights ω_k^i are null except one
- when a particle $\xi_k^{i_0}$ has an almost null weight $\omega_k^{i_0} \simeq 0$:

$$\sum_{i=1}^{N} \omega_k^i \, \phi(\xi_k^i) \simeq \sum_{i \neq i_0} \omega_k^i \, \phi(\xi_k^i)$$

  this mean that this particle has a neglectable contribution in the approximation
- How to favor the particle with large weight ?
- How to do more with the same ?
- The solution is proposed by the Baron Münchhausen

bootstrap

- pull oneself out of a swamp by one's pigtail
- pull oneself over a fence by one's bootstraps

- Resampling: among the weighted particles $(\xi_k^i, \omega_k^i)_{i=1\cdots N}$ we duplicate those with significant weight to the detriment of those with low weight

the idea came in 1993

Novel approach to nonlinear/non-Gaussian Bayesian state estimation

N.J. Gordon
D.J. Salmond
A.F.M. Smith

Abstract: An algorithm, the bootstrap filter, is proposed for implementing recursive Bayesian filters. The required density of the state vector is represented as a set of random samples, which are updated and propagated by the algorithm. The method is not restricted by assumptions of linearity or Gaussian noise: it may be applied to any state transition or measurement model. A simulation example of the bearings only tracking problem is presented. This simulation includes schemes for improving the efficiency of the basic algorithm. For this example, the performance of the bootstrap filter is greatly superior to the standard extended Kalman filter.

Indexing terms: Kalman filter, Sequential estimation, Bayesian filter

Paper 9241F (E5), first received 27th April and in revised form 8th October 1992

N.J. Gordon and D.J. Salmond are with the Defence Research Agency, Farnborough, Hampshire, United Kingdom
A.F.M. Smith is with the Department of Statistics, Imperial College, London, United Kingdom

IEE PROCEEDINGS-F, Vol. 140, No. 2, APRIL 1993

bootstrap filter

bootstrap filter

$\xi_0^1, \dots, \xi_0^N \overset{\text{iid}}{\sim} p_{X_0}$ {initialization}

for $k = 1, 2, 3 \dots$ **do**

$\xi_{k-}^i \sim Q_k(\,\cdot\,|\xi_{k-1}^i)$, $i = 1 \cdots N$ {sampling}

$\omega_{k-}^i = L_k(y_k, \xi_{k-}^i)$, $i = 1 \cdots N$ {importance weights}

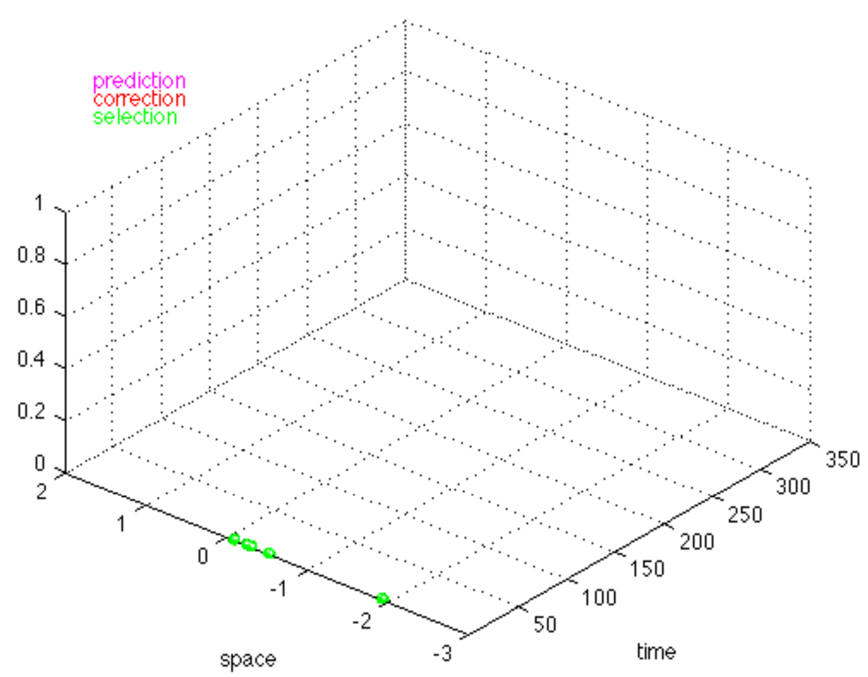
$\omega_{k-}^i = \omega_{k-}^i \,/\, \sum_{i'} \omega_{k-}^{i'}$ {normalizing}

$\xi_k^1, \dots, \xi_k^N \overset{\text{iid}}{\sim} \sum_{i=1}^N \omega_{k-}^i \, \delta_{\xi_{k-}^i}$ {re-sampling}

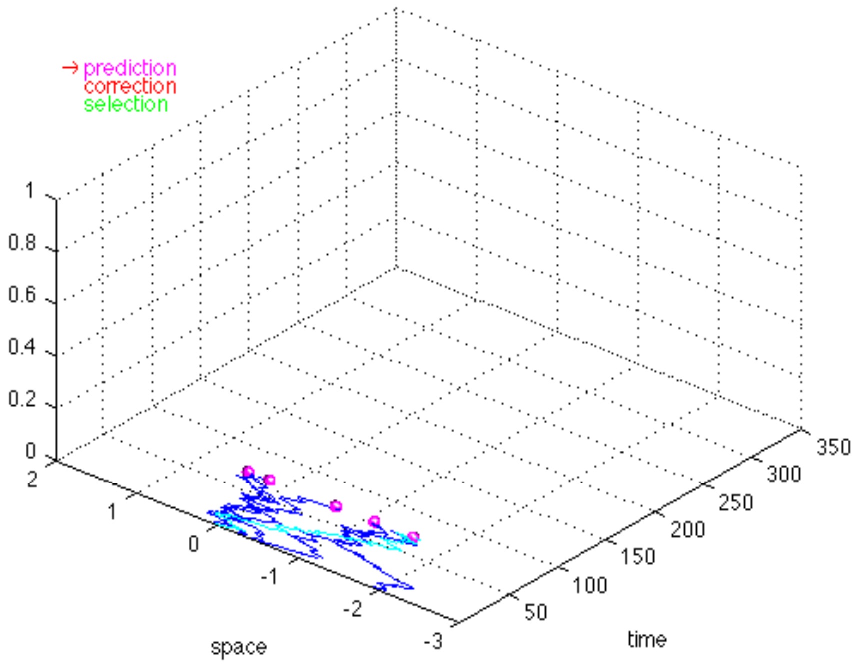
end for

- 3 ingredients
  - "mimic" the state model
  - computing a likelihood weight
  - resampling procedure
- bootstrap filter; sampling importance re-sampling (SIR) filter

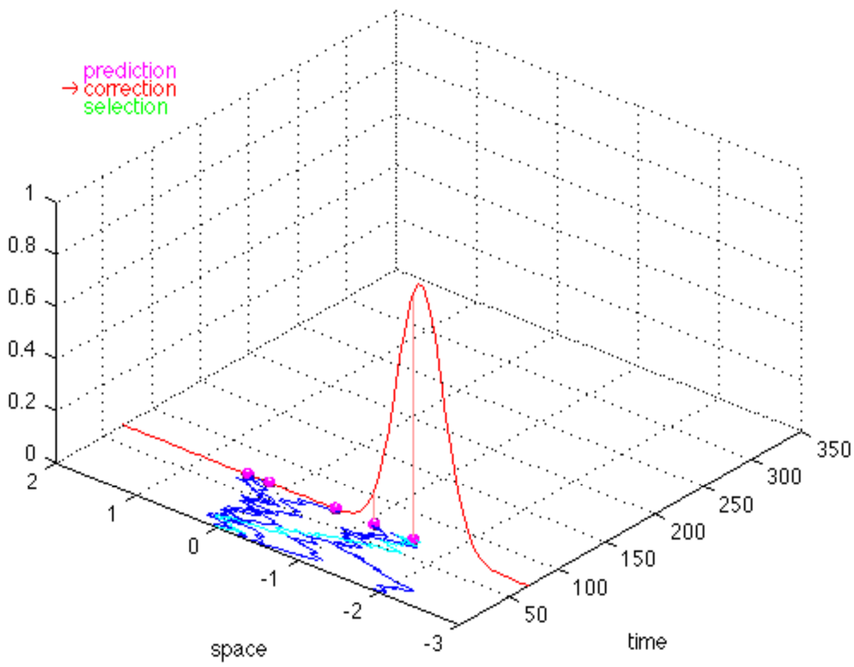
same example



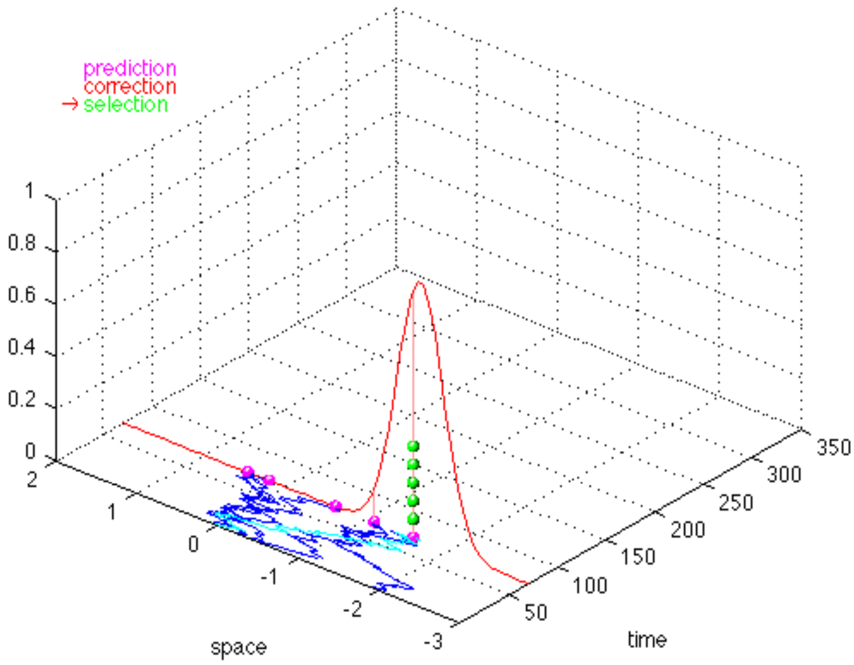
same example



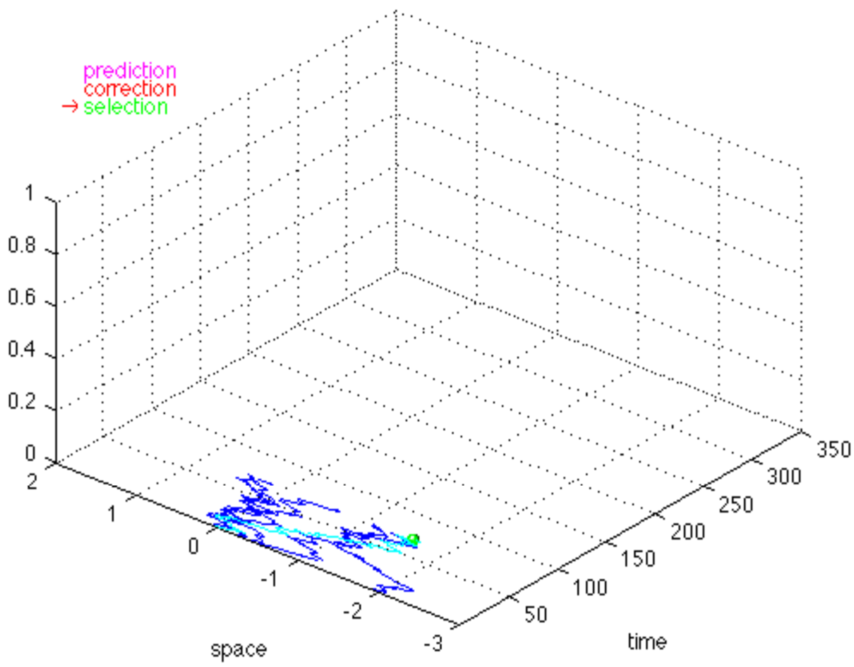
same example



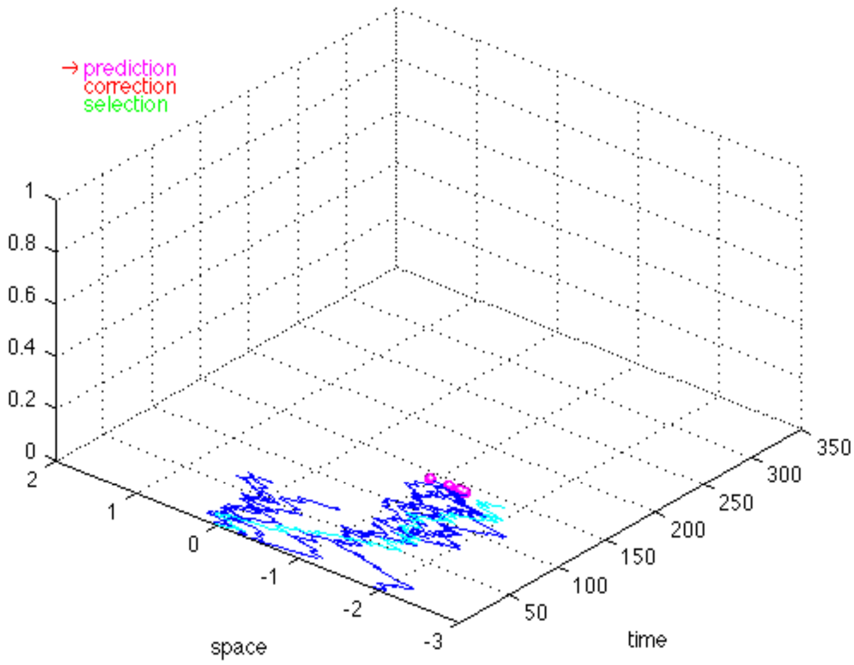
same example



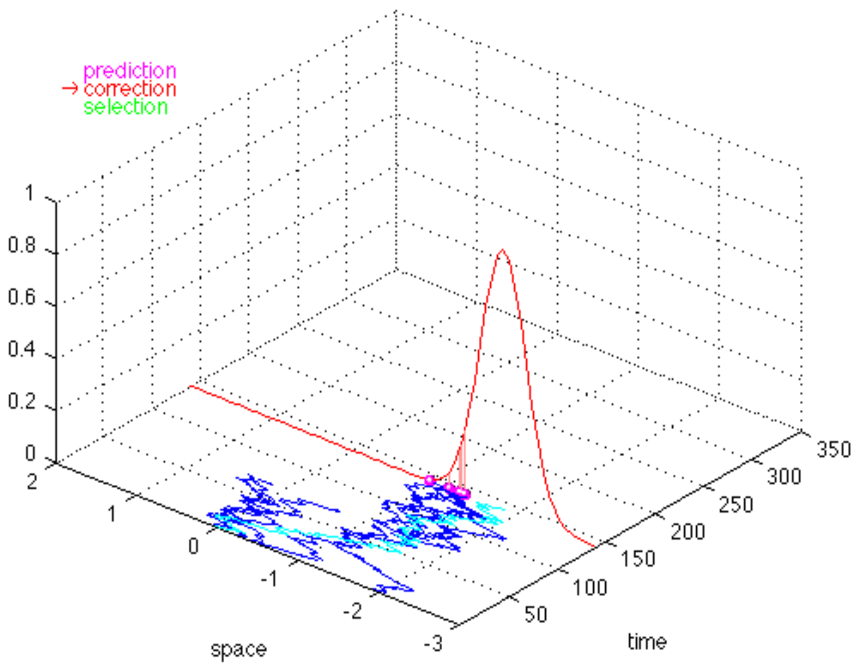
same example



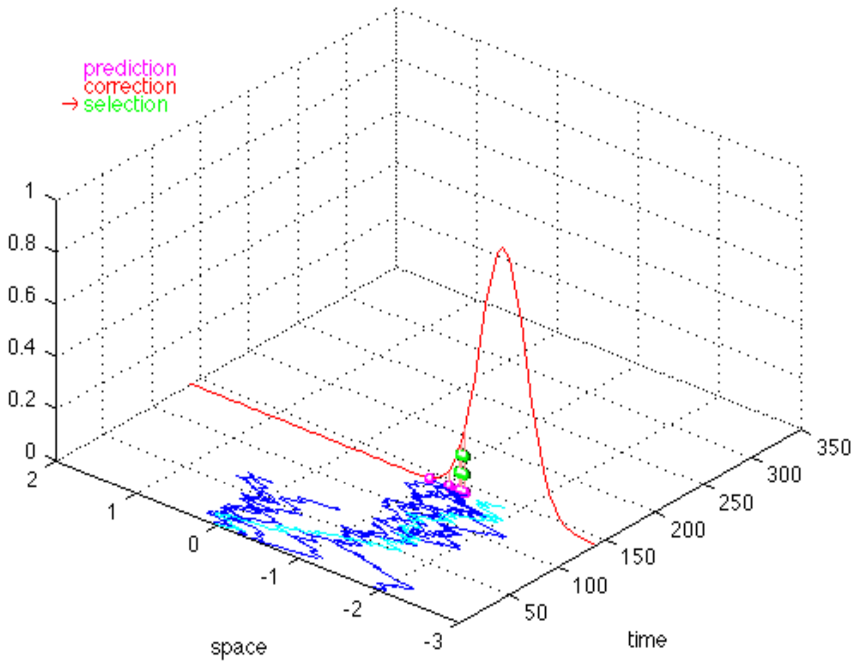
same example



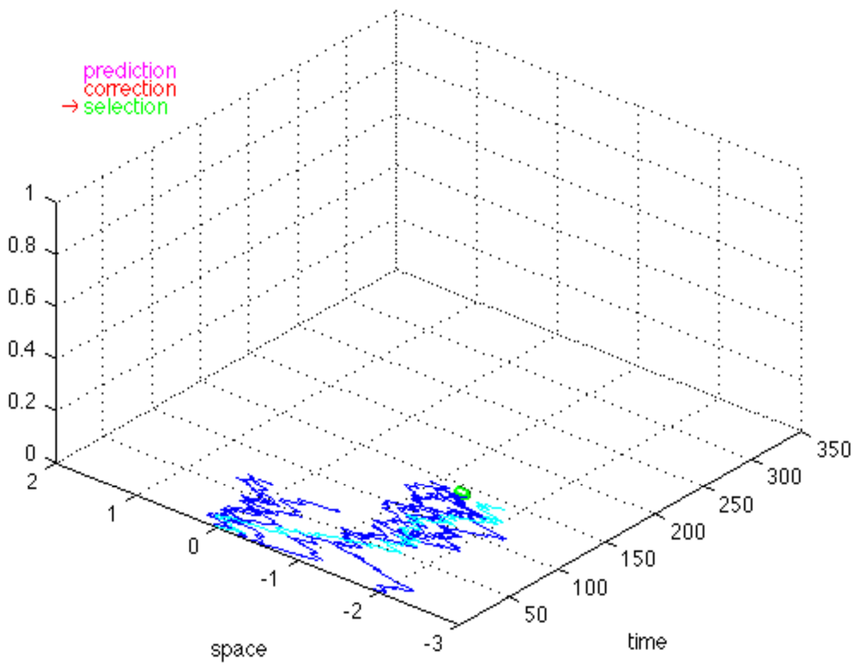
same example



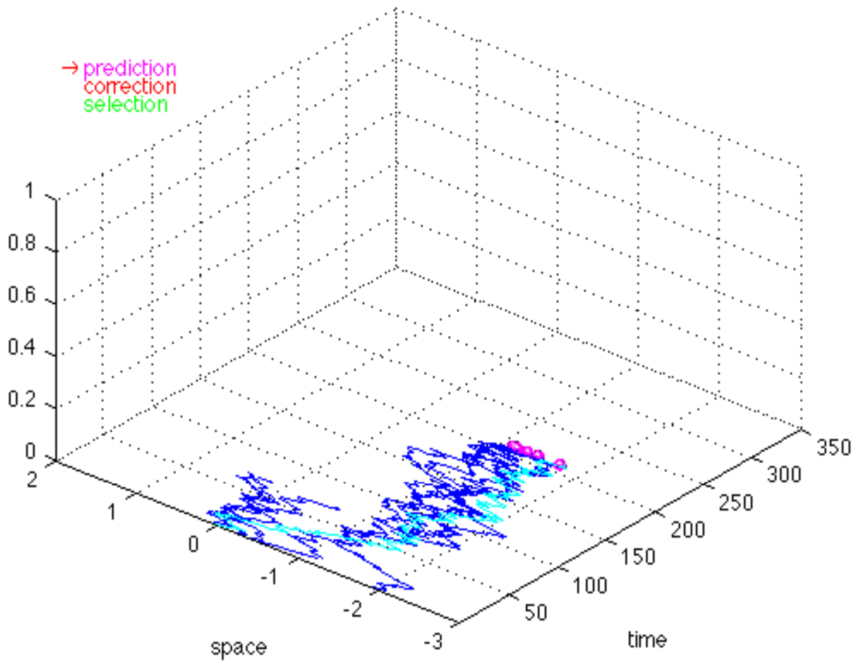
same example



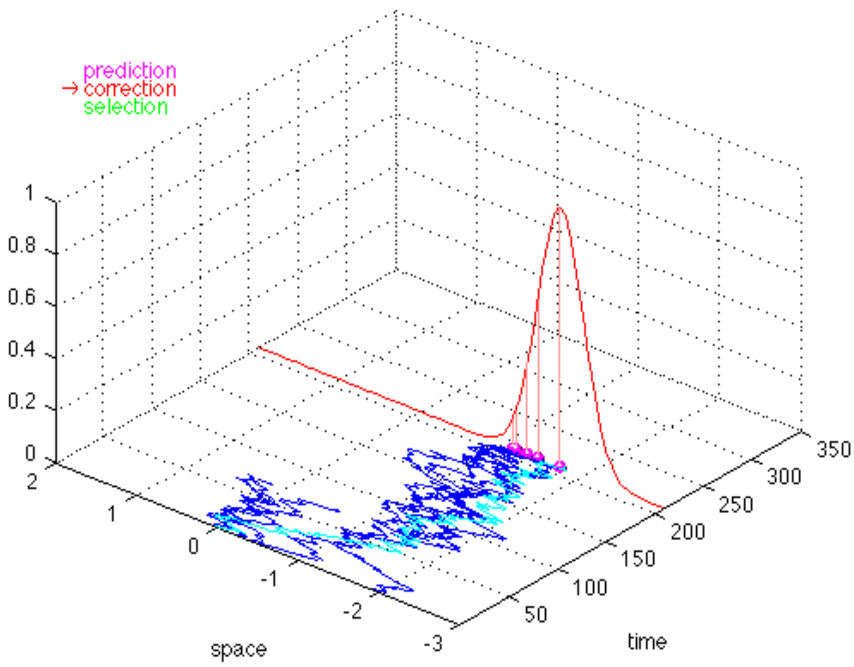
same example



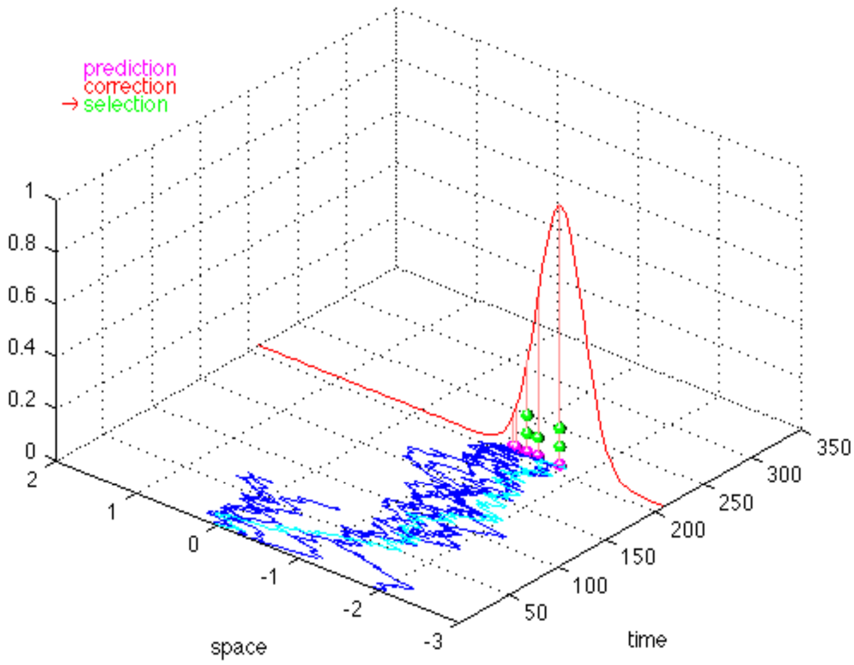
same example



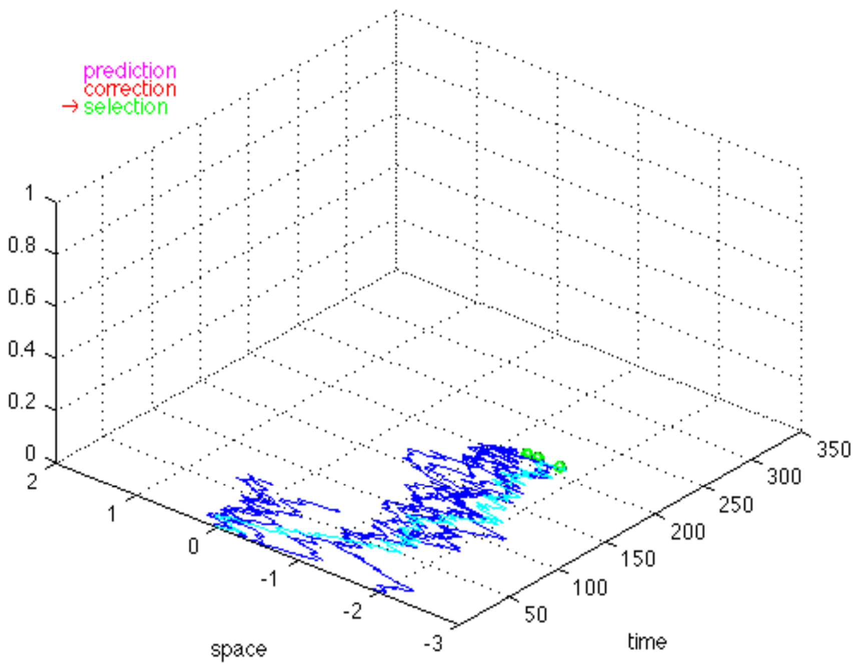
same example



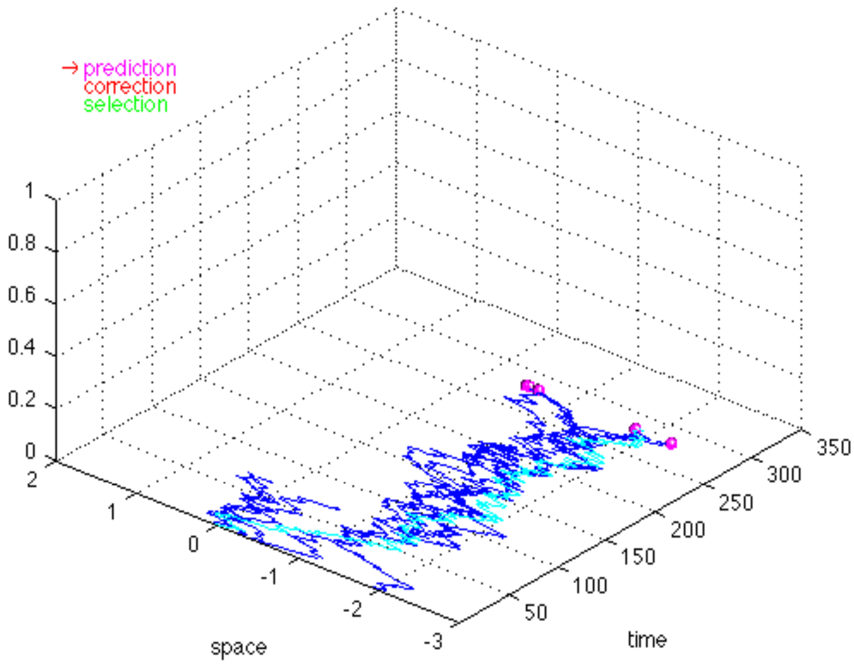
same example



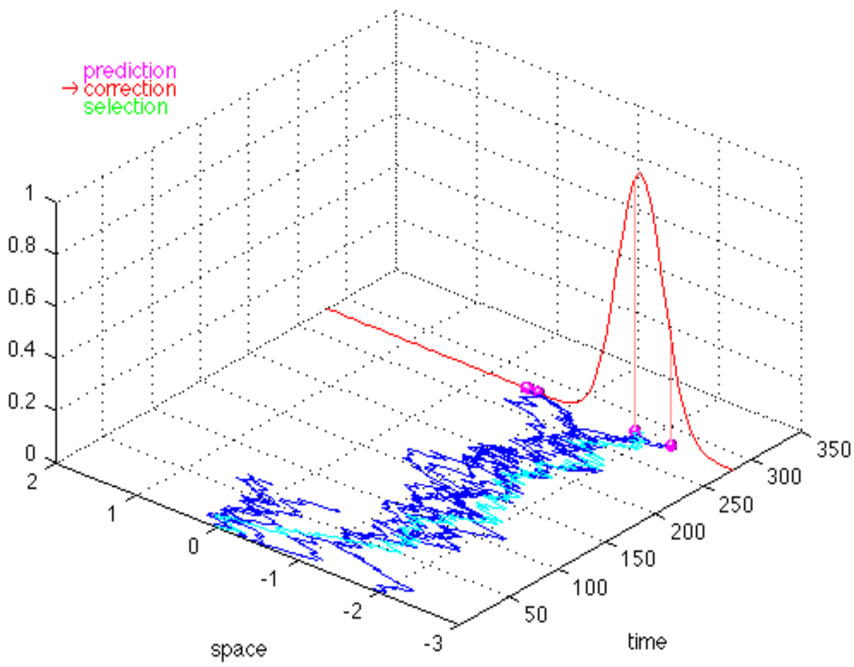
same example



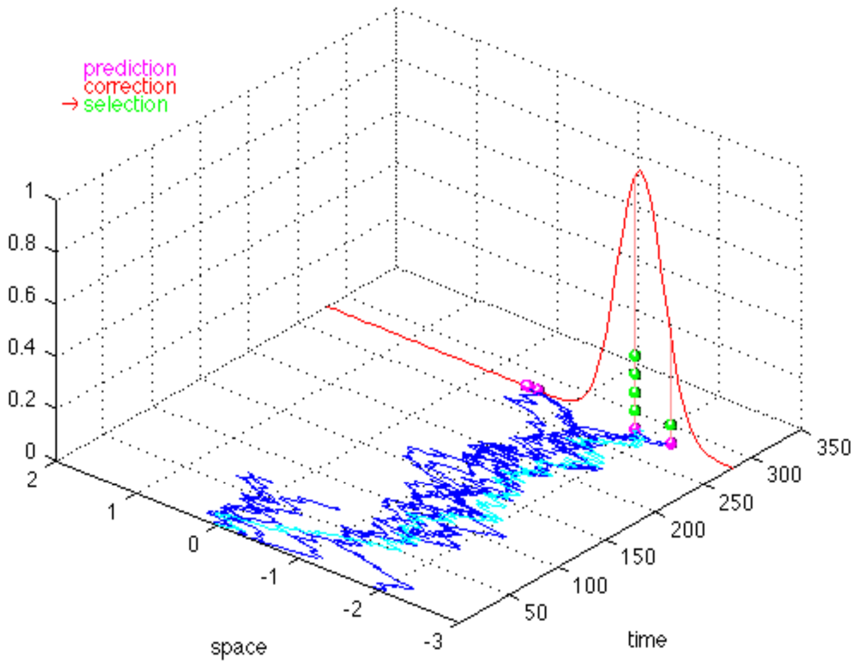
same example



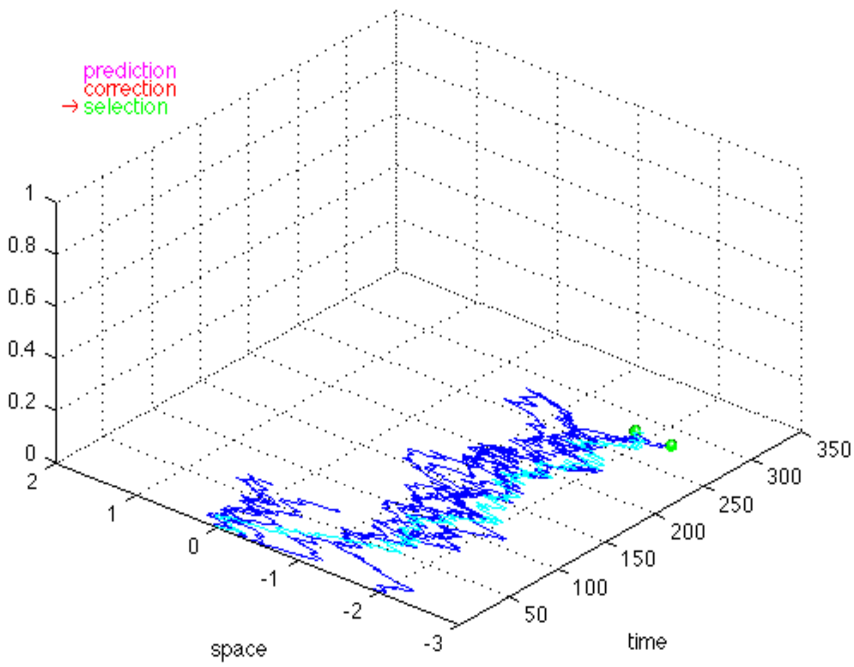
same example



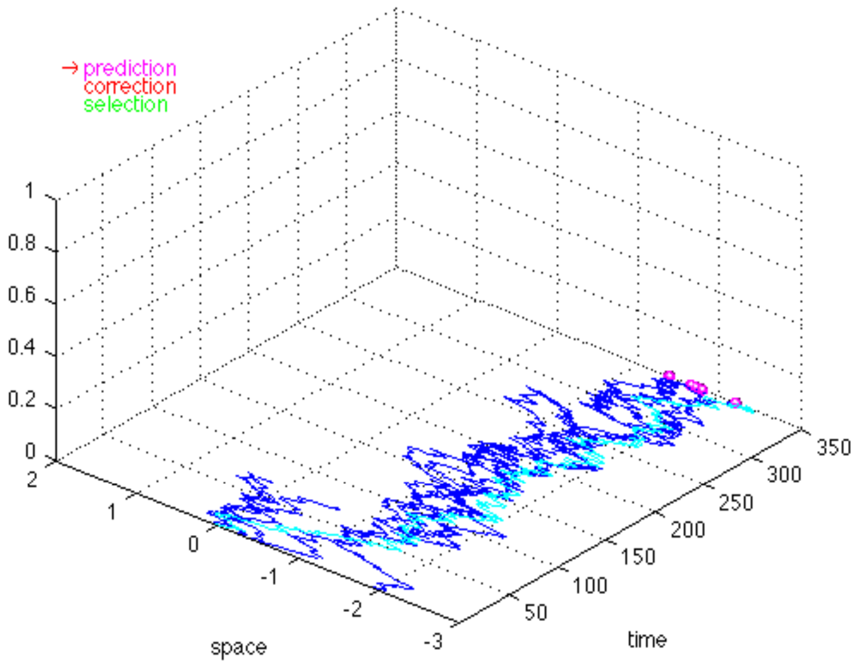
same example



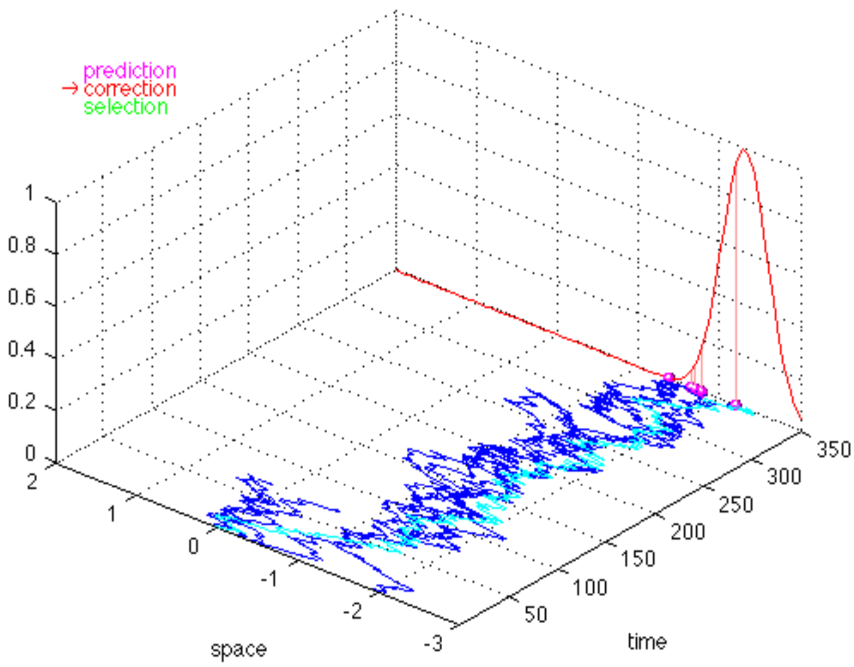
same example



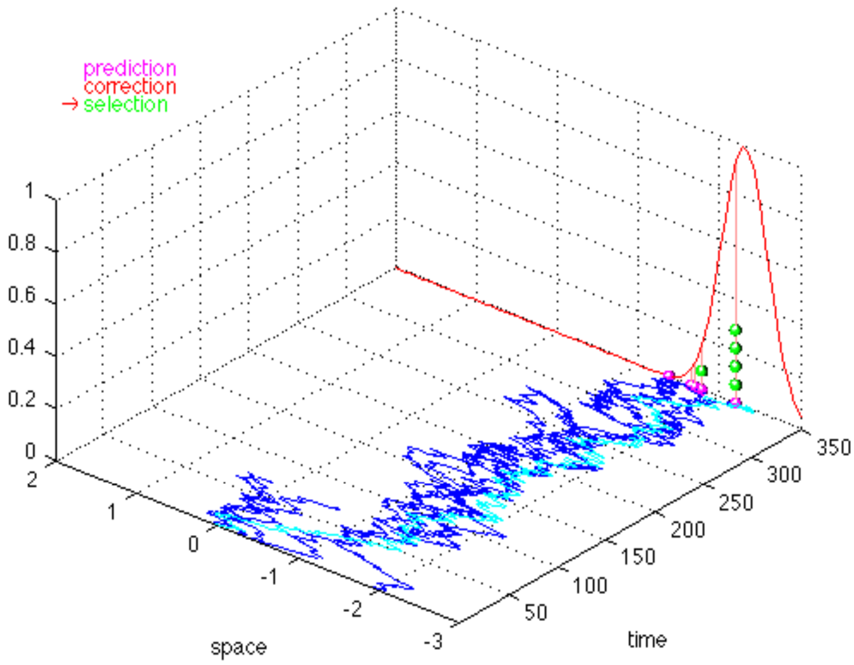
same example



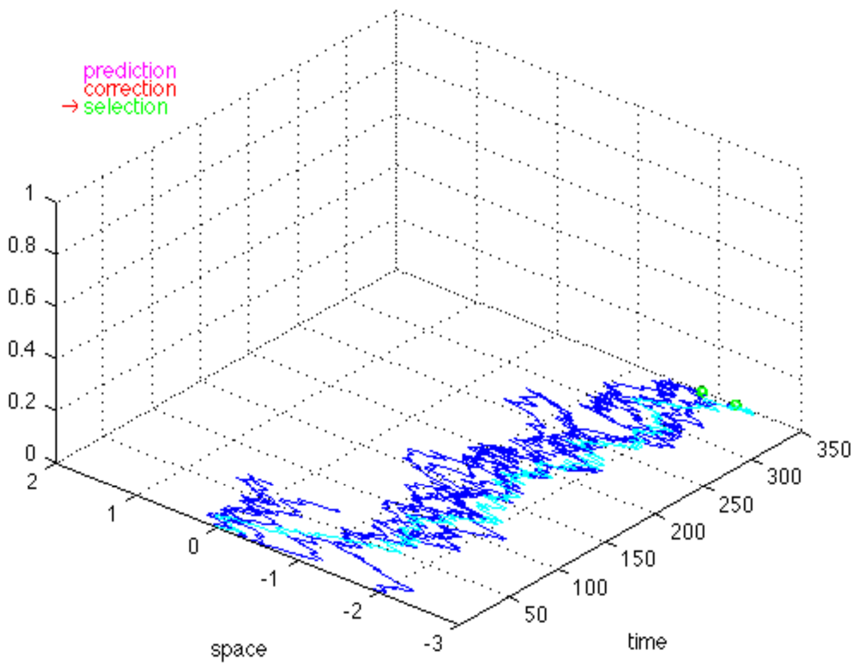
same example



same example

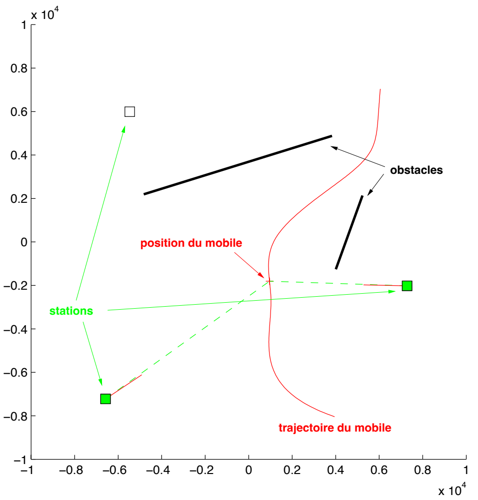


same example



content

bearing-only tracking with obstacles



- mobile on the plane
- S measuring stations (azimuth)
- L obstacles may occult the mobile
- state equation

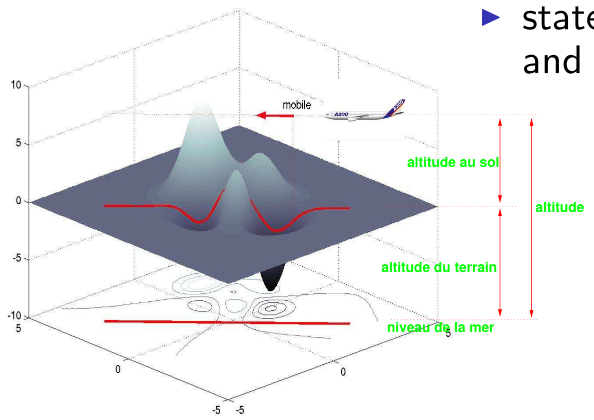
$$X^1_{k+1} = X^1_k + \sigma_W \, W^1_k$$
$$X^2_{k+1} = X^2_k + \sigma_W \, W^2_k$$

- observation equation: for each station $s = 1 \cdots S$
  - if $s \not\circledast \text{mobile} \to$ no measure
  - if $s \circledast \text{mobile} \to$ the measure is:

$$Y^s_k = h_s(X_k) + \sigma_V \, V^s_k \qquad \text{with} \qquad h_s(x) = \text{arctg}\big(\tfrac{x^1 - x^{1,s}}{x^2 - x^{2,s}}\big)$$

$(x^{1,s}, x^{2,s})$ is the location of the station s

digital elevation model



- state model X_k with constant heading and velocity

$$r_k^1 = r_{k-1}^1 + \Delta t \, v_{k-1} \, \cos(c_{k-1})$$
$$r_k^2 = r_{k-1}^2 + \Delta t \, v_{k-1} \, \sin(c_{k-1})$$
$$v_k = v_{k-1} + \Delta t \, \sigma_v \, W_{k-1}^1$$
$$c_k = c_{k-1} + \Delta t \, \sigma_c \, W_{k-1}^2$$

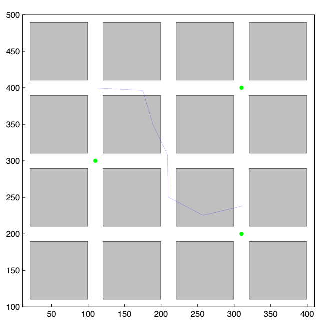
- observation model

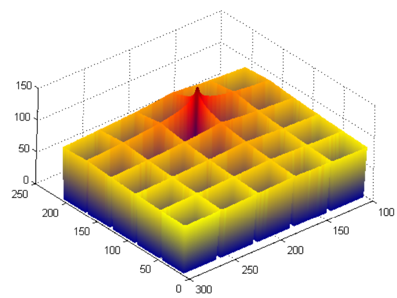
$$Y_k = h(X_k) + \text{measurement noise}$$

$h(X_k)$ is the ground elevation at location (r_k^1, r_k^2)

tracking a mobile phone in Manhattan

- Mobile phone in the streets and buildings of Manhattan and communicating with different stations $s = 1, \dots, S$ • one digital map (strength of the signal) per station s • the received signal strength (RSS) is measured $\to$ locate the mobile.





digital map $x \to h_s(x)$

- state equation: 2D Brownian particle

$$X_{k+1}^1 = X_k^1 + \sigma_W \sqrt{\Delta t} \, W_k^1 \qquad X_{k+1}^2 = X_k^2 + \sigma_W \sqrt{\Delta t} \, W_k^2$$

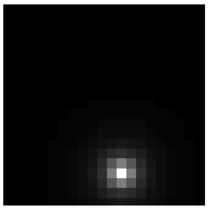
- observation equation: $Y_k^s = h_s(X_k) + \text{ noise}, \; s = 1 \cdots S$

track before detect

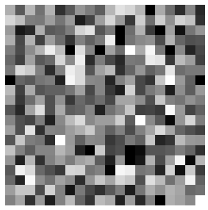
- observations: sequence of images $Y_1\, Y_2 \cdots$, each image is:

$Y_k(s) = \mathcal{I}_{r_k}(s) + B_k(s)$

$s \in \mathcal{S}$ pixel index



$\mathcal{I}_{r_k}$



noise B_k

$r_k = (r_k^1, r_k^2)$ unknown position of the mobile in the 2D plane

- state model: $X = ($position r^j, speed v^j, acceleration $a^j)$

$$\frac{d}{dt}\begin{pmatrix} r^j(t) \\ v^j(t) \\ a^j(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} r^j(t) \\ v^j(t) \\ a^j(t) \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \beta^j(t)$$

- very low signal to noise ratio • bootstrap filter fails • Rao-Blackwellization (hybrid particles/Kalman filter)

content

convergence I

$$\widehat{f(X_k)} \overset{\text{def}}{=} \mathbb{E}[f(X_k)|Y_{1:k}] \qquad \widehat{f(X_k)}^N \overset{\text{def}}{=} \frac{1}{N}\sum_{i=1}^N f(\xi_k^i)$$

- Monte Carlo type cv, i.e. central limit theorem results:

$$\left\{\mathbb{E}\left[|\widehat{f(X_k)} - \widehat{f(X_k)}^N|^p \middle| Y_{1:k} = y_{1:k}\right]\right\}^{1/p} \leq \frac{C_k}{\sqrt{N}} \sup_x |f(x)|$$

$$\sqrt{N}\,\left[\widehat{f(X_k)} - \widehat{f(X_k)}^N\right] \underset{N\to\infty}{\Longrightarrow} \mathcal{N}(0, \sigma_k^2)$$

  - SMC beats the curse of dimensionality as the rate of convergence is independent of the state variable dimension
  - quite limited: C_k and $\sigma_k^2 \to \infty$ as $k \to \infty$

convergence II

- under the hypothesis of exponential stability of the filter, i.e. the filter forgets its past:

$$\tfrac{1}{2}\int\Big|p_{X_k|Y_{1:k}=y_{1:k},X_0=x'}(x) - p_{X_k|Y_{1:k}=y_{1:k},X_0=x''}(x)\Big|\,\mathrm{d}x \leq \alpha^n$$

with $0 < \alpha < 1$ then $C_k \leq C$ and $\sigma_k^2 \leq D$

  - this his the reason of the success of SMC
  - C and D depend exponentially on the dimension n
  - many convergence results, see Del Moral, P. (2004). Feynman-Kac formulae – Genealogical and interacting particle approximations. Springer–Verlag, New York

content

improvements

- standard SMC is like shooting in the dark:

$$p_{X_k|Y_{1:k-1}=y_{1:k-1},Y_k=y_k}(x) \propto p_{Y_k|X_k=x}(y_k) \times p_{X_k|Y_{1:k-1}=y_{1:k-1}}(x)$$

  to avoid degeneracy and improve the algorithm, use the measurement y_k to perform a better sample at time $k-1$, before the particles propagate to time k (see auxiliary particle filter, Pitt and Shephard 1999).

- resample-move algorithms: to improve the diversity of particles, after resampling a lot of particles are similar so we can move the particles according to an artificial dynamics (that preserve the posterior target distribution)

content

perspectives/conclusions

- perspectives
  - better bounds in the convergence theorems
  - algorithm improvements
    - control and adaptation of the number of particles
    - smoothing and joint state filtering and parameter estimation
    - if the time interval between the observation is large or for off-line applications: joint SMC and MCMC procedures
    - from particle to density representation thru kernel smoothing
- conclusions
  - easy to understand and to implement (need of a proper resampling technique)
  - easy to adapt, to modify...
  - EKF, UKF ... should be preferred when they works well
  - SMC: low signal-to-noise ratio

content

references

- renewal: Bayesian numerics

  - Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993). Novel approach to nonlinear/non–Gaussian Bayesian state estimation. IEE Proceedings, Part F, 140(2):107–113
  - Doucet, A., de Freitas, N., and Gordon, N. J., editors (2001). Sequential Monte Carlo Methods in Practice. Springer–Verlag, New York
  - Crisan, D. and Doucet, A. (2002). A survey of convergence results on particle filtering for practitioners. IEEE Transactions on Signal Processing, 50(3):736–746
  - Ristic, B., Arulampalam, M. S., and Gordon, N. J. (2004). Beyond the Kalman Filter: Particle Filters for Tracking Applications. Artech House
  - Del Moral, P. (2004). Feynman-Kac formulae – Genealogical and interacting particle approximations. Springer–Verlag, New York
  - Cappé, O., Moulines, E., and Ryden, T. (2005). Inference in Hidden Markov Models. Springer–Verlag

- a big book!

  - Crisan, D. and Rozovskii, B. (2011). The Oxford Handbook of Nonlinear Filtering. Oxford University Press

references I

- Anderson, B. D. O. and Moore, J. B. (1978).
  Optimal Filtering.
  Prentice-Hall.
- Anderson, B. D. O. and Moore, J. B. (1992).
  Kalman filtering: Whence, what and whither ?
  Mathematical system theory The influence of R.E. Kalman, 19(4):42–54.
- Bain, A. and Crişan, D. (2008).
  Fundamentals of stochastic filtering.
  Springer Verlag.
- Baum, L. E. and Petrie, T. (1966).
  Statistical inference for probabilistic functions of finite state Markov chains.
  The Annals of Mathematical Statistics, 37(6):1554–1563.
- Beneš, V. E. (1981).
  Exact finite–dimensional filters for certain diffusions with nonlinear drift.
  Stochastics, 5(1+2):65–92.
- Cappé, O., Moulines, E., and Ryden, T. (2005).
  Inference in Hidden Markov Models.
  Springer–Verlag.
- Chaleyat-Maurel, M. and Michel, D. (1984).
  Des résultats de non existence de filtre de dimension finie.
  Stochastics, 13(1-2):83–102.
- Churchill, G. A. (1989).
  Stochastic models for heterogeneous DNA sequences.
  Bulletin of Mathematical Biology, 51(1):79–94.

references II

- Crisan, D. and Doucet, A. (2002).
  A survey of convergence results on particle filtering for practitioners.
  IEEE Transactions on Signal Processing, 50(3):736–746.
- Crisan, D. and Rozovskii, B. (2011).
  The Oxford Handbook of Nonlinear Filtering.
  Oxford University Press.
- Del Moral, P. (2004).
  Feynman-Kac formulae – Genealogical and interacting particle approximations.
  Springer–Verlag, New York.
- Doucet, A., de Freitas, N., and Gordon, N. J., editors (2001).
  Sequential Monte Carlo Methods in Practice.
  Springer–Verlag, New York.
- Duncan, T. E. (1967).
  Probability Densities for Diffusion Processes with Applications to Nonlinear Filtering Theory and Detection Theory.
  PhD thesis, Stanford University.
- Durbin, R., Eddy, S. R., Krogh, A., and Mitchison, G. (1998).
  Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids.
  Cambridge University Press, Cambridge.
- Evensen, G. (2009).
  Data assimilation: the ensemble Kalman filter.
  Springer Verlag.

references III

- Fujisaki, M., Kallianpur, G., and Kunita, H. (1972).
  Stochastic differential equations for the non–linear filtering problem.
  Osaka Journal of Mathematics, 9(1):19–40.
- Gelb, A., editor (1974).
  Applied Optimal Estimation.
  MIT Press.
- Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993).
  Novel approach to nonlinear/non–Gaussian Bayesian state estimation.
  IEE Proceedings, Part F, 140(2):107–113.
- Gustafsson, F., Gunnarsson, F., Bergman, N., Forssell, U., Jansson, J., Karlsson, R., and Nordlund, P.-J. (2002).
  Particle filters for positioning, navigation and tracking.
  IEEE Transactions on Signal Processing, 50(2).
- Hazewinkel, M., Marcus, S. I., and Sussmann, H. J. (1983).
  Nonexistence of finite dimensional filters for conditional statistics of the cubic sensor problem.
  Systems & Control Letters, 3(6):331–340.
- Isard, M. and Blake, A. (1998).
  CONDENSATION — Conditional density propagation for visual tracking.
  International Journal of Computer Vision, 29(1):5–28.
- Jazwinskii, A. H. (1970).
  Stochastic processes and filtering theory.
  Academic Press.

references IV

- Kailath, T. (1974).
  A view of three decades of linear filtering theory.
  IEEE Trans. on Inf. Th., IT-20(2):146–181.
- Kalman, R. E. (1960).
  A new approach to linear filtering and prediction problems.
  Transactions of the ASME–Journal of Basic Engineering, 82, Series D:35–45.
- Kolmogorov, A. N. (1939).
  Sur l'interpolation et extrapolation des suites stationnaires.
  Comptes Rendus à l'Académie des Sciences, 208:2043–2045.
- Kushner, H. J. (1964).
  On the differential equations satisfied by conditional probability densities of Markov processes.
  SIAM Journal on Control and Optimization, 2:106–119.
- Mortensen, R. E. (1966).
  Optimal control of continuous-time stochastic systems.
  PhD thesis, Univ. California, Berkeley.
- Pardoux, É. (1991).
  Filtrage non linéaire et équations aux dérivées partielles stochastiques associées.
  In Hennequin, P.-L., editor, École d'Eté de Probabilités de Saint–Flour XIX, 1989, volume 1464 of Lecture Notes in Mathematics, pages 67–163. Springer–Verlag, Berlin.
- Rabiner, L. R. and Juang, B.-H. (1986).
  An introduction to hidden Markov models.
  IEEE ASSP Magazine, pages 4–15.

references V

- Ristic, B., Arulampalam, M. S., and Gordon, N. J. (2004).
  Beyond the Kalman Filter: Particle Filters for Tracking Applications.
  Artech House.
- Rozovskii, B. L. (1990).
  Stochastic Evolution Systems.
  Kluwer.
- Stratonovich, R. L. (1959).
  On the theory of optimal non-linear filtration of random functions.
  Teor. Veroyatnost. i Primenen., 4:223–225.
- Stratonovich, R. L. (1960).
  Application of the theory of Markov processes for optimum filtration of signals.
  Radio Eng. Electron. Phys, 1:1–19.
- Thrun, S., Burgard, W., and Fox, D. (2005).
  Probabilistic Robotics.
  MIT-Press.
- Viterbi, A. (1967).
  Error bounds for convolutional codes and an asymptotically optimum decoding algorithm.
  IEEE Transactions on Information Theory, 13(2):260–269.
- Wiener, N. (1949).
  Extrapolation, Interpolation and Smoothing of Time Series, with Engineering Applications.
  John Wiley & Sons.
  Originally appears in 1942 as a classified National Defense Research Council Report; also published under the title Time Series Analysis by MIT Press.

references VI

- Zakai, M. (1969).
  On the optimal filtering of diffusion processes.
  Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 1(I):230–243.

content

content

some applications of SMC

- image tracking • Isard, M. and Blake, A. (1998). Condensation — Conditional density propagation for visual tracking. International Journal of Computer Vision, 29(1):5–28
- tracking • Ristic, B., Arulampalam, M. S., and Gordon, N. J. (2004). Beyond the Kalman Filter: Particle Filters for Tracking Applications. Artech House • Gustafsson, F., Gunnarsson, F., Bergman, N., Forssell, U., Jansson, J., Karlsson, R., and Nordlund, P.-J. (2002). Particle filters for positioning, navigation and tracking. IEEE Transactions on Signal Processing, 50(2)
- IA and robotics • Thrun, S., Burgard, W., and Fox, D. (2005). Probabilistic Robotics. MIT-Press
- meteorology, oceanography, weather prediction systems • Evensen, G. (2009). Data assimilation: the ensemble Kalman filter. Springer Verlag

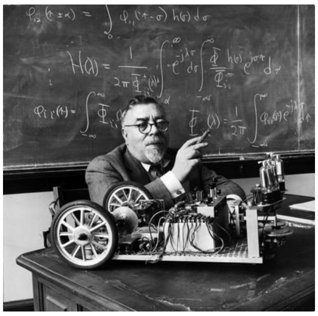
content

Historical references I

- Andrey Nikolaevich Kolmogorov (1903-1987)
  - discrete time

    ▷ Kolmogorov, A. N. (1939). Sur l'interpolation et extrapolation des suites stationnaires. Comptes Rendus à l'Académie des Sciences, 208:2043–2045

Historical references II



- Norbert Wiener (1894-1964)
  - Wiener filter
  - analog signal processing in continuous time: stationary signal X_t observed with additive $Y_t = X_t + v_t$

    ▷ Wiener, N. (1949). Extrapolation, Interpolation and Smoothing of Time Series, with Engineering Applications. John Wiley & Sons. Originally appears in 1942 as a classified National Defense Research Council Report; also published under the title Time Series Analysis by MIT Press

Historical references III

- Rudolf (Rudy) Emil Kálmán (1930-)
  - Kalman filter
  - linear/Gaussian case

    ▷ Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. Transactions of the ASME–Journal of Basic Engineering, 82, Series D:35–45

    ▷ Stratonovich, R. L. (1960). Application of the theory of Markov processes for optimum filtration of signals. Radio Eng. Electron. Phys, 1:1–19

Historical references IV

- Ruslan Leont'evich Stratonovich (1930-1997)
  - non linear optimal filter in continuous time

    ▷ Stratonovich, R. L. (1959). On the theory of optimal non-linear filtration of random functions. Teor. Veroyatnost. i Primenen., 4:223–225 [with a little problem with the definition of the stochastic integral]

    ▷ Kushner, H. J. (1964). On the differential equations satisfied by conditional probability densities of Markov processes. SIAM Journal on Control and Optimization, 2:106–119 [with Ito stochastic integral]
  - via the innovation process

    ▷ Fujisaki, M., Kallianpur, G., and Kunita, H. (1972). Stochastic differential equations for the non–linear filtering problem. Osaka Journal of Mathematics, 9(1):19–40

Historical references V

- Duncan–Mortensen–Zakai or Fujisaki-Kallianpur-Kunita equation
  - optimal nonlinear filter for the unnormalized distribution

    ▷ Mortensen, R. E. (1966). Optimal control of continuous-time stochastic systems. PhD thesis, Univ. California, Berkeley

    ▷ Duncan, T. E. (1967). Probability Densities for Diffusion Processes with Applications to Nonlinear Filtering Theory and Detection Theory. PhD thesis, Stanford University

    ▷ Zakai, M. (1969). On the optimal filtering of diffusion processes. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 1(I):230–243

    ▷ [link: Encyclopaedia of Mathematics Edited by Michiel Hazewinkel]

Historical references VI

- the finite dimensional filters
  - there are nonlinear filtre that – just like the Kalman filter – are in finite dimension (i.e. function of finite dimensional statistics)

    ▷ Beneš, V. E. (1981). Exact finite–dimensional filters for certain diffusions with nonlinear drift. Stochastics, 5(1+2):65–92
  - but most of the time it's not the case

    ▷ Hazewinkel, M., Marcus, S. I., and Sussmann, H. J. (1983). Nonexistence of finite dimensional filters for conditional statistics of the cubic sensor problem. Systems & Control Letters, 3(6):331–340

    ▷ Chaleyat-Maurel, M. and Michel, D. (1984). Des résultats de non existence de filtre de dimension finie. Stochastics, 13(1-2):83–102
  - nota: finite dimensional filter $\not\Rightarrow$ good filter

Historical references VII

- nonlinear filter: the "PDE/analysis" approaches

  ▷ Rozovskii, B. L. (1990). Stochastic Evolution Systems. Kluwer

  ▷ Pardoux, É. (1991). Filtrage non linéaire et équations aux dérivées partielles stochastiques associées. In Hennequin, P.-L., editor, École d'Eté de Probabilités de Saint–Flour XIX, 1989, volume 1464 of Lecture Notes in Mathematics, pages 67–163. Springer–Verlag, Berlin

  - analytical expression of the filter $\not\Rightarrow$ numerical filter

Historical references VIII

- the discrete approach: hidden Markov models
  - around Leonard E. Baum
  - discrete time, finite state space

    ▷ Baum, L. E. and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. The Annals of Mathematical Statistics, 37(6):1554–1563

    ▷ Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Transactions on Information Theory, 13(2):260–269
  - traitement de la parole ▷ Rabiner, L. R. and Juang, B.-H. (1986). An introduction to hidden Markov models. IEEE ASSP Magazine, pages 4–15
  - traitement des séquences d'ADN ▷ Churchill, G. A. (1989). Stochastic models for heterogeneous DNA sequences. Bulletin of Mathematical Biology, 51(1):79–94 ▷ Durbin, R., Eddy, S. R., Krogh, A., and Mitchison, G. (1998). Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press, Cambridge

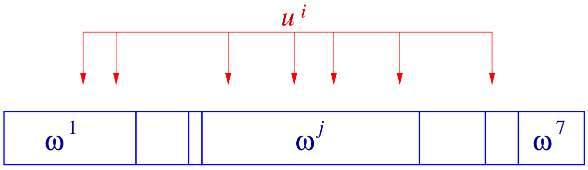
Historical references IX

- some extra references
  - Jazwinskii, A. H. (1970). Stochastic processes and filtering theory. Academic Press
  - Gelb, A., editor (1974). Applied Optimal Estimation. MIT Press
  - Kailath, T. (1974). A view of three decades of linear filtering theory. IEEE Trans. on Inf. Th., IT-20(2):146–181
  - Anderson, B. D. O. and Moore, J. B. (1978). Optimal Filtering. Prentice-Hall
  - Anderson, B. D. O. and Moore, J. B. (1992). Kalman filtering: Whence, what and whither ? Mathematical system theory The influence of R.E. Kalman, 19(4):42–54
  - Bain, A. and Crişan, D. (2008). Fundamentals of stochastic filtering. Springer Verlag

content

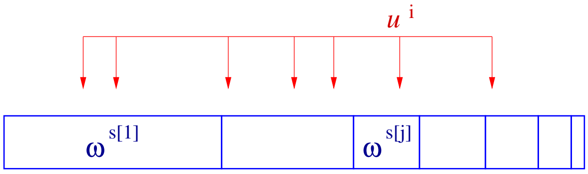
resampling I

- multinomial resampling: $u^1, \dots, u^N \sim U[0,1]$



too slow: $O(N\ \log(N))$

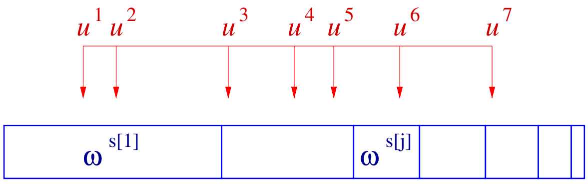
- 1st improvement: sort the $\omega^{1:N}$ by decreasing order: $\omega^{s[1]} > \omega^{s[2]} > \cdots > \omega^{s[N]}$



still $O(N\ \log(N))$ but much faster (less tests)

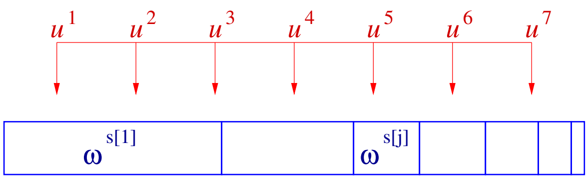
resampling II

- 2nd improvement: ordered uniform sample $u^1 < u^2 < \cdots < u^N \overset{\text{iid}}{\sim} \mathcal{U}[0,1]$



order $O(N)$ (with an efficient ordered uniform sample sampler)

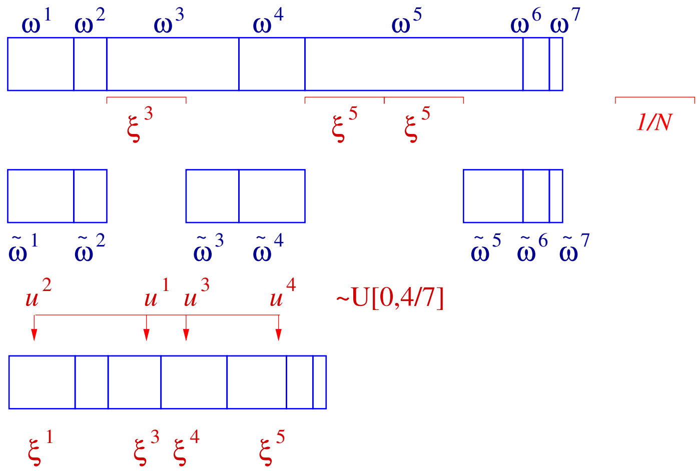
- Kitagawa resampling: we sample only $u_1 \sim \mathcal{U}[0, \frac{1}{N}]$ then $u^i = u^1 + \frac{i}{N}\,,\ i = 2:N$



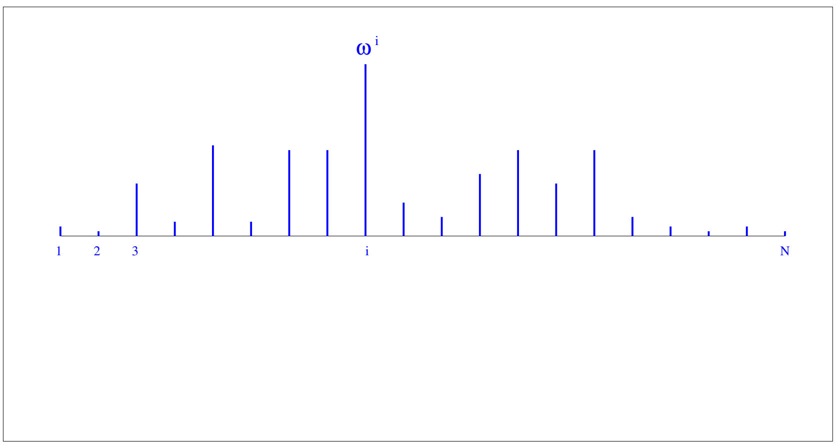
fast $O(N)$ but not very efficient

resampling III

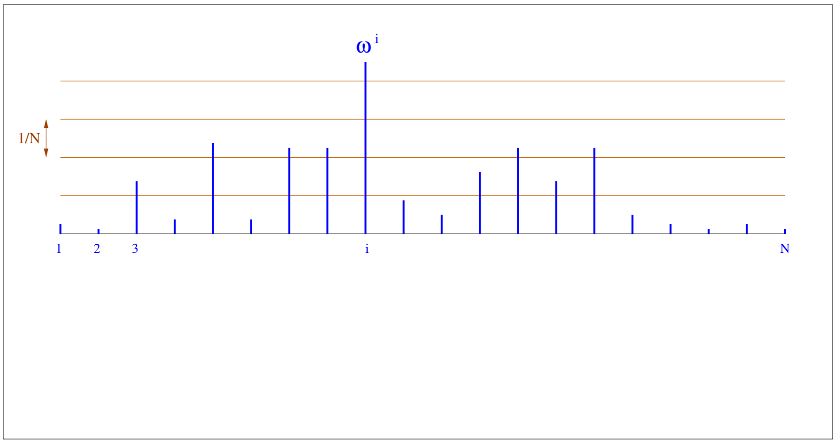
- residual sampling: Kitagawa for a certain number of particles and multinomial sampling for the other particles



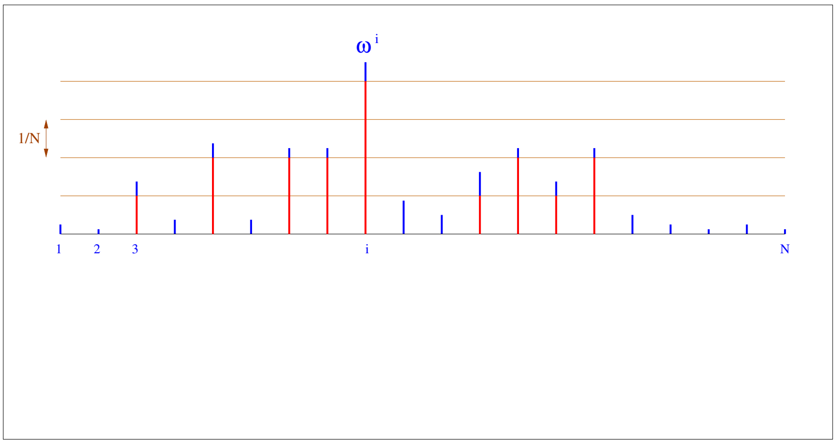
residual sampling (cont.)



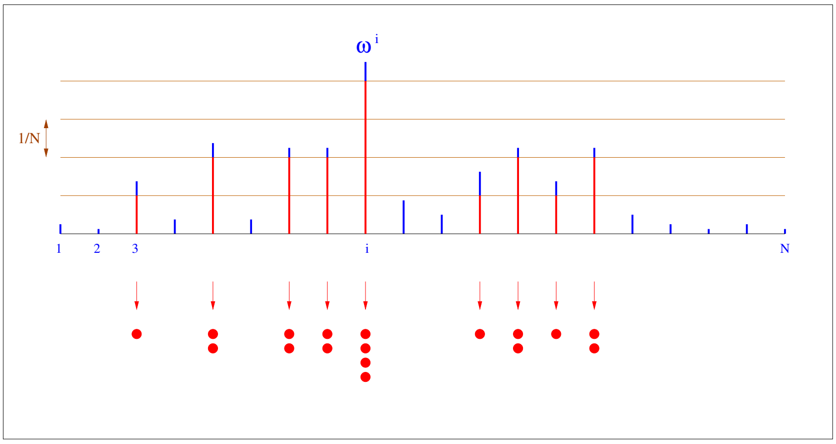
residual sampling (cont.)



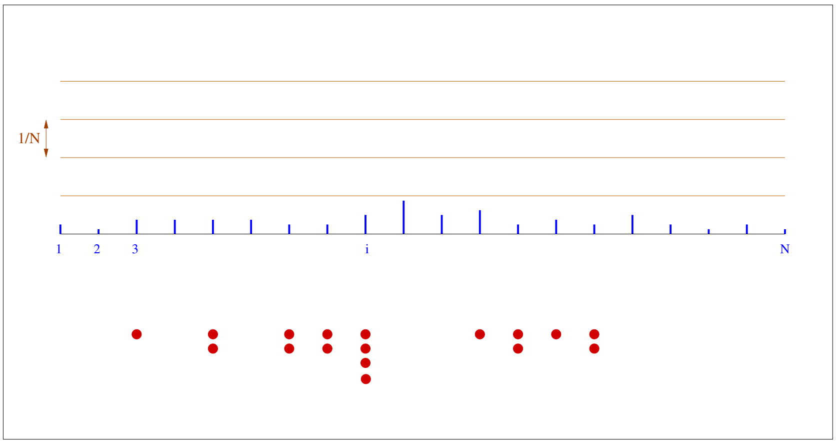
residual sampling (cont.)



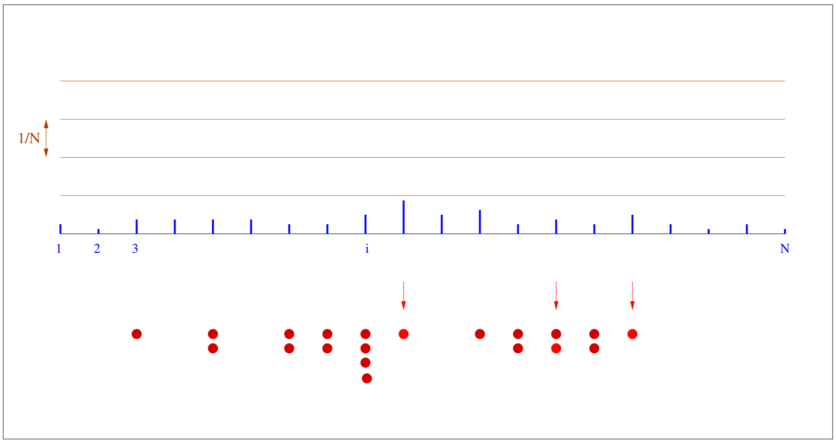
residual sampling (cont.)



residual sampling (cont.)



residual sampling (cont.)



content

method of least squares

- $X_k = Z_\theta(t_k)$ with

$$\dot{Z}_\theta(t) = f(\theta, Z_\theta(t)) \qquad \text{(state model)}$$

and

$$y_k^\theta = h(Z_\theta(t_k)) \qquad \text{(observation model)}$$

and determine the best fits to the data by findind θ which minimizes

$$J(\theta) = \sum_{\ell=1}^{k} |Y_k - y_k^\theta|^2$$

- basically non-recursive but can be adapted to make it recursive
- weighted least squares: favor more precise observations
- implicit model of noise that combine observation noise and model imprecision

content

Unscented transform

- quadrature formula: if $(\omega^i, x^i)_{i=1:N}$ is a quadrature for a random variable X i.e.:

$$\sum_{i=1}^{N} \omega^i \, x^i \simeq \mathbb{E}X$$

$$\sum_{i=1}^{N} \omega^i \, (x^i - \mathbb{E}X) \, (x^i - \mathbb{E}X)^* \simeq \text{cov}(X)$$

then $(\omega^i, y^i)_{i=1:N}$ with $y^i = f(x^i)$ is a quadrature for $Y = f(X)$, i.e.

$$\sum_{i=1}^{N} \omega^i \, y^i \simeq \mathbb{E}Y$$

$$\sum_{i=1}^{N} \omega^i \, (y^i - \mathbb{E}Y) \, (y^i - \mathbb{E}Y)^* \simeq \text{cov}(Y)$$

content

iteration $\pi_{k-1} \to \pi_k$

- prediction $\pi_{k-1} \to \pi_{k^-}$: π_{k-1} is "transported" by the transition kernel Q_k

$$\begin{aligned}\pi_{k^-}(x) &= p_{X_k|Y_{1:k-1}=y_{1:k-1}}(x) \\ &= \int p_{X_k|Y_{1:k-1}=y_{1:k-1},X_{k-1}=x'}(x)\, p_{X_{k-1}|Y_{1:k-1}=y_{1:k-1}}(x')\,\mathrm{d}x' \\ &= \int p_{X_k|X_{k-1}=x'}\, p_{X_{k-1}|Y_{1:k-1}=y_{1:k-1}}(x')\,\mathrm{d}x' \quad \text{(state space)} \\ &= \int Q_k(x'|x)\, \pi_{k-1}(x')\,\mathrm{d}x'\end{aligned}$$

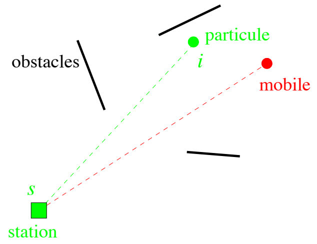
- correction $\pi_{k^-} \to \pi_k$:π_k is updated from π_{k^-} by integrating the new observation $Y_k = y_k$ via Bayes formula:

$$\begin{aligned}\pi_k(x) &= p_{X_k|Y_k=y_k,Y_{1:k-1}=y_{1:k-1}}(x) \\ &\propto p_{Y_k|X_k=x,Y_{1:k-1}=y_{1:k-1}}(y_k)\, p_{X_k|Y_{1:k-1}=y_{1:k-1}}(x) \quad \text{(Bayes)} \\ &\propto p_{Y_k|X_k=x}(y_k)\, p_{X_k|Y_{1:k-1}=y_{1:k-1}}(x) \quad \text{(state space)} \\ &\propto L_k(y_k|x)\, \pi_{k^-}(x)\end{aligned}$$

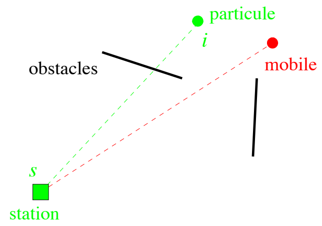
content

example 1: expression of the likelihood I

- likelihood: at time k, we compute the likelihood $\omega_k^{i,s}$ of the particle ξ_{k-}^i for station s
  - case 1 : $s \circledast \text{mobile}$



$s \circledast \xi_{k-}^i$

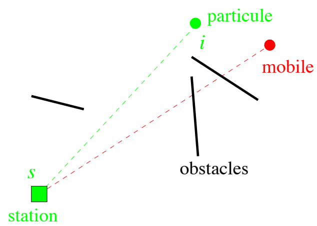


$s \not\circledast \xi_{k-}^i$

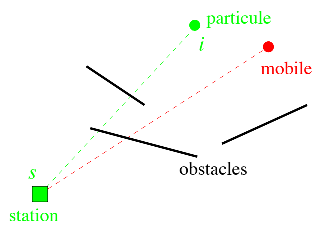
$$\omega_k^{i,s} = \begin{cases} \exp\big(-\frac{1}{2\,\sigma_V}|Y_k^s - h_s(\xi_{k-}^i)|^2\big) & \text{if } s \circledast \xi_{k-}^i \\ 0 & \text{if } s \not\circledast \xi_{k-}^i \end{cases}$$

example 1: expression of the likelihood II

- case 2: $s \not\circledast \text{mobile}$



$s \circledast \xi_{k-}^i$



$s \not\circledast \xi_{k-}^i$

$$\omega_k^{i,s} = \begin{cases} 0 & \text{if } s \circledast \xi_{k-}^i \\ 1 & \text{if } s \not\circledast \xi_{k-}^i \end{cases}$$

- the likelihood of the particle ξ_{k-}^i at time k:

$$\omega_k^i = \prod_{s=1}^{S} \omega_k^{i,s}$$

example 2: expression of the likelihood I

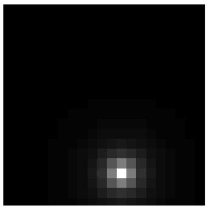
- PSF (point spread function)

$$\mathcal{I}_r(s) = \frac{\delta^2\, c}{2\pi\sigma^2} \exp\Big(-\frac{|s\,\delta - r|^2}{2\sigma^2}\Big)\, 1_{C(r)}(s)$$

where

$$1_{C(r)}(s) \stackrel{\text{def}}{=} 1_{(|s^1\,\delta - r^1|<3)}\, 1_{(|s^2\,\delta - r^2|<3)}$$
$$|s\,\delta - r|^2 \stackrel{\text{def}}{=} (s^1\,\delta - r^1)^2 + (s^2\,\delta - r^2)^2$$

and $s = (s^1, s^2) \in \mathcal{S}$, $r = (r^1, r^2) \in \mathbb{R}^2$, δ pixel size



example 2: expression of the likelihood II

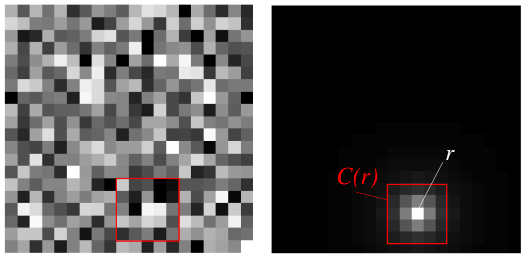
- likelihood function for $Y = \{Y(s)\}_{s\in\mathcal{S}}$ fixed, we compute

$$r \mapsto \mathcal{L}(Y|r) = \exp\Big\{-\frac{1}{2\,\sigma_B^2}\sum_{s\in\mathcal{S}}\big[\mathcal{I}_r(s) - Y(s)\big]^2\Big\}$$

$$\begin{aligned}
\mathcal{L}(Y|r) &= \exp\Big\{-\frac{1}{2\,\sigma_B^2}\sum_{s\in\mathcal{S}}\mathcal{I}_r(s)\,[\mathcal{I}_r(s) - Y(s)]\Big\}\exp\Big\{-\frac{1}{2\,\sigma_B^2}\sum_{s\in\mathcal{S}}[Y(s)]^2\Big\} \\
&= \exp\Big\{-\frac{1}{2\,\sigma_B^2}\sum_{s\in C(r)}\mathcal{I}_r(s)\,[\mathcal{I}_r(s) - Y(s)]\Big\}\exp\Big\{-\frac{1}{2\,\sigma_B^2}\sum_{s\in\mathcal{S}}[Y(s)]^2\Big\} \\
&\propto \exp\Big\{-\frac{1}{2\,\sigma_B^2}\underbrace{\underbrace{\sum_{s\in C(r)}}_{\text{local sum}}\mathcal{I}_r(s)\,[\mathcal{I}_r(s) - 2\,Y(s)]}_{=:M_{C(r)}(\mathcal{I}_r, Y)}\Big\}
\end{aligned}$$

example 2: expression of the likelihood III

$$\tilde{\mathcal{L}}(Y|r) = \exp\Big\{ -\frac{1}{2\,\sigma_B^2} M_{C(r)}(\mathcal{I}_r, Y) \Big\}$$



example 2: Rao-Blackwellized particle filter I

- the standard bootstrap filter fails at tracking the mobile
- improvement: Rao-Blackwellization
  - $X_k = (a_k, v_k, r_k)$, only r_k is in the observation and $\text{law}(a_k, v_k | r_k)$ is Gaussian
  - hybrid approach: r_k is represented as a particle and (a_k, v_k) is processes with a Kalman filter
  - more complex/more precise
- Rao-Blackwellized iteration $k \to k+1$
  - likelihood: $\omega^i \propto \tilde{\mathcal{L}}(Y_k | r^i_{k-})$
  - resampling: $r^i_k \leftarrow \texttt{resample}(\omega^{1:N}, r^{1:N}_{k-})$
  - prediction: $r^i_{k+1-} \sim \text{law}(r_{k+1} | r_k = r^i_{k-})$

  here $\text{law}(r_{k+1} | r_k = r^i)$ is Gaussian (Kalman filter)

example 2: Rao-Blackwellized particle filter II

décomposition du système décomposition: $\alpha = (a, v)$ $\beta = r$

$$\begin{pmatrix} \alpha_{k+1} \\ \beta_{k+1} \end{pmatrix} = \begin{pmatrix} F^{\alpha\alpha} & F^{\alpha\beta} \\ F^{\beta\alpha} & F^{\beta\beta} \end{pmatrix} \begin{pmatrix} \alpha_k \\ \beta_k \end{pmatrix} + \begin{pmatrix} W_k^\alpha \\ W_k^\beta \end{pmatrix}$$

avec

$$\mathrm{cov}(W_k) = \begin{pmatrix} Q^{\alpha\alpha} & Q^{\alpha\beta} \\ Q^{\beta\alpha} & Q^{\beta\beta} \end{pmatrix}$$

example 2: Rao-Blackwellized particle filter III

$\forall i$ simuler $r^i_{k+1} \sim \mathrm{law}(\beta_{k+1} | \beta_k = r^i_k)$, itération $k \to k+1$:

$\mathrm{law}(\alpha_{k+1} | \beta_k = r^i_k) = \mathcal{N}(\hat\alpha^i_{k+1^-}, R_{k+1^-})$

$$\begin{aligned} \hat\alpha^i_{k+1^-} &= F^{\alpha\alpha}\, \hat\alpha^i_k + F^{\alpha\beta}\, r^i_k \\ R_{k+1^-} &= F^{\alpha\alpha}\, R_k\, (F^{\alpha\alpha})^* + Q^{\alpha\alpha} \end{aligned}$$

$\mathrm{law}(\beta_{k+1} | \beta_k = r^i_k) = \mathcal{N}(\hat r^i_{k+1^-}, \Xi_{k+1})$

$$\begin{aligned} \hat r^i_{k+1^-} &= F^{\beta\alpha}\, \hat\alpha^i_k + F^{\beta\beta}\, r^i_k \\ \Xi_{k+1} &= F^{\beta\alpha}\, R_k\, (F^{\beta\alpha})^* + Q^{\beta\beta} \\ S_{k+1} &= F^{\alpha\alpha}\, R_k\, (F^{\beta\alpha})^* + Q^{\alpha\beta} \end{aligned}$$

$r^i_{k+1^-} \sim \mathrm{law}(\beta_{k+1} | \beta_k = r^i_k)$

example 2: Rao-Blackwellized particle filter IV

$\text{law}(\alpha_{k+1}|\beta_1 = r_1^i, \beta_{k+1} = r_{k+1^-}^i) = \mathcal{N}(\hat{\alpha}_{k+1}^i, R_{k+1})$

$$\hat{\alpha}_{k+1}^i = \hat{\alpha}_{k+1^-}^i + S_{k+1}\,\Xi_{k+1}^{-1}\,(r_{k+1}^i - \hat{r}_{k+1^-}^i)$$
$$R_{k+1} = R_{k+1^-} - S_{k+1}\,\Xi_{k+1}^{-1}\,S_{k+1}^*$$

les covariances ne dépendent pas de i