complément scientifique École Doctorale MATISSE
IRISA et INRIA, salle Markov
jeudi 19 janvier 2012

Introduction au Filtrage Particulaire
Exemples en Navigation, Localisation et Poursuite

François Le Gland
INRIA Rennes et IRMAR

http://www.irisa.fr/aspi/legland/ed-matisse/

directeur de recherche INRIA, responsable de l'équipe

ASPI : Applications Statistiques des Systèmes de Particules en Interaction

http://www.irisa.fr/aspi/

thèmes

- filtrage particulaire, et applications en
  - localisation, navigation et poursuite
  - assimilation de données séquentielle
- inférence statistique des modèles de Markov cachés
- simulation d'évènements rares, et extensions en
  - simulation moléculaire
  - optimisation globale
- analyse mathématique des méthodes particulaires

contrats industriels

- avec France Télécom R&D, sur la localisation de terminaux mobiles
- avec Thalès, sur la navigation par corrélation de terrain
- avec DGA /Techniques Navales, sur l'optimisation du positionnement et de l'activation de capteurs

collaboration régulière avec l'ONERA (office nationale de recherche et d'études aérospatiales) : encadrement de doctorants

projets ANR

- FIL, sur la fusion de données pour la localisation
- PREVASSEMBLE, sur les méthodes d'ensemble pour l'assimilation de données et la prévision

projets européens

- HYBRIDGE et iFLY, sur les méthodes de Monte Carlo conditionnelles pour l'évaluation de risque en trafic aérien

Introduction to filtering

- Bayesian approach
- usefulness of a prior model
- some models
- Bayesian estimation (static case)
- exemple : Gaussian model (static case)

Bayesian approach

objective is to estimate, in a *recursive* manner as much as possible, the unknown state X_n of a system (typically, position and velocity of a mobile) in view of noisy observations $Y_{0:n} = (Y_0 \cdots Y_n)$ related to the hidden state by a relation such as

$$Y_k = h_k(X_k) + V_k$$

many applications in *localization, navigation and tracking* of a mobile

in such cases, objective is to estimate position and velocity of a mobile, using :

(i) *measurements* provided by sensors

(ii) possibly a *reference* database, available for instance as a digital map

often needed to take into account

(iii) a *prior model* (not necessarily accurate) for the mobile displacement or evolution, usually a Markov model

Bayesian approach, aka information fusion

	prior information (state model)
$+$	likelihood (sensor model, consistency between state and measurement)
$\Rightarrow$	posteriori information

general principle : using Bayes rule

$$p_{X,Y}(x,y) = p_{X \mid Y=y}(x) \; p_Y(y) = p_{Y \mid X=x}(y) \; p_X(x)$$

conditional distribution of $X_{0:n}$ given $Y_{0:n}$ (from which conditional distribution of X_n given $Y_{0:n}$ is easily obtained, by marginalization) can be expressed in terms of

- conditional distribution of $Y_{0:n}$ given $X_{0:n}$, often easy to evaluate for instance in the additive model

$$Y_k = h_k(X_k) + V_k$$

- distribution of $X_{0:n}$

recursive implementation possible thanks to Markov property however, conditional distribution of X_n given $Y_{0:n}$ does not have an explicit expression, except in a few special situations

- Markov chain with finite state space (forward Baum equation)
- Gaussian linear systems (Kalman filter)

in more general situations, growing interest (since beginning of 90's) for simulation–based methods of Monte Carlo type

particle filtering provides an implementation of Bayesian approach that is

- intuitive, easy to understand and implement
- flexible, many algorithmic variants available
- numerically efficient
- *easy* to analyze mathematically

questions

- why a prior model should be needed ? how to take it into account ?
- why (and how) should conditional distributions be computed ?

some notations

if X is a random variable taking values in E, then mapping

$$\phi \longmapsto \mathbb{E}[\phi(X)] \qquad \text{or equivalently} \qquad A \longmapsto \mathbb{P}[X \in A]$$

defines a probability distribution μ on E, denoted as

$$\mu(dx) = \mathbb{P}[X \in dx]$$

and such that

$$\mathbb{E}[\phi(X)] = \int_E \phi(x)\, \mu(dx) = \langle \mu, \phi \rangle \qquad \text{or} \qquad \mathbb{P}[X \in A] = \mu(A)$$

caracterizes uncertainty about X

usefulness of a prior model

without additional information, observing

$$Y_k = h_k(X_k) + V_k$$

poses a new estimation problem at each new time instant (observations are not accumulated)

in general, in each of these separated estimation problems, hidden state dimension is greater than observation dimension

estimating hidden state X_k in view of observation Y_k alone, and even if noise V_k would not be present, is an ill–posed problem

- relation has more unknown variables than equations
- even in the favorable case where dimensions are equal, relation can have several distinct solutions
- reconstructed sequence $X_0, X_1, \cdots, X_n$ can be poor as a trajectory, even though each of its component separately is fine (temporal inconsistency)

common *variational* approach, to overcome *underdetermination* and / or solve *temporal inconsistency* : introduction of additional information about the hidden state sequence, e.g. in the form of cost functions involving initial state or transitions between two successive states

for instance, minimize w.r.t. sequence $x_{0:n} = (x_0, x_1, \cdots, x_n)$, a criterion

$$J(x_{0:n}) = c_0(x_0) + \sum_{k=1}^{n} c_k(x_{k-1}, x_k) + \sum_{k=0}^{n} d_k(x_k)$$

that combines cost functions representing *prior* information about solution, and cost functions representing for instance *consistency* between state and measurement, of the form

$$d_k(x) = \tfrac{1}{2} \, |Y_k - h_k(x)|^2$$

interpretation : sought state x_k should also satisfy observation equation $Y_k = h_k(x_k)$ at least in some approximate sense

more generally, these cost functions could just represent a constraint (or a property) that a sequence solution should fulfill *prior* information is typically incorporated in terms of cost functions of the form

$$c_0(x) = \frac{1}{2} |x - \mu|^2 \qquad \text{and} \qquad c_k(x, x') = \frac{1}{2} |x' - f_k(x)|^2$$

interpretation

- sought initial state x_0 should be close to μ
- sought transition (x_{k-1}, x_k) should satisfy equation $x_k = f_k(x_{k-1})$ at least in some approximate sense

this regularization approach (from *optimisation* point of view) can be interpreted as incorporating prior information (from *statistical estimation* point of view) *prior* information can also be incorporated in terms of cost functions of the form

$$c_0(x) = -\log p_0(x) \qquad \text{and} \qquad c_k(x, x') = -\log p_k(x' \mid x)$$

in this case, minimizing criterion

$$J(x_{0:n}) = -\log p_0(x_0) - \sum_{k=1}^{n} \log p_k(x_k \mid x_{k-1}) + \sum_{k=0}^{n} d_k(x_k)$$

is equivalent to maximizing (MAP, maximum a posteriori)

$$\exp\{-J(x_{0:n})\} = \underbrace{p_0(x_0) \prod_{k=1}^{n} p_k(x_k \mid x_{k-1})}_{p_{0:n}(x_{0:n})} \exp\{-\sum_{k=0}^{n} d_k(x_k)\}$$

where $p_{0:n}(x_{0:n})$ denotes joint probability density of successive states (trajectory) $X_{0:n} = (X_0, X_1, \cdots, X_n)$ of a Markov chain characterized by

- initial probability density $p_0(x_0)$
- and transition probability densities $p_k(x' \mid x)$

pragmatical *statistical* approach : instead of looking for sequence minimizing (or maximizing) a criterion, produce an estimator (MMSE, minimum mean square error) and a quantitative measure of estimation error, i.e. evaluate path averages (or integrals) of the form

$$\int_E \cdots \int_E f(x_{0:n}) \; \exp\{-J(x_{0:n})\} \; dx_{0:n}$$

$$= \int_E \cdots \int_E f(x_{0:n}) \; \exp\{-\sum_{k=0}^{n} d_k(x_k)\} \; p_{0:n}(x_{0:n}) \; dx_{0:n}$$

$$= \mathbb{E}[f(X_{0:n}) \; \exp\{-\sum_{k=0}^{n} d_k(X_k)\}\,]$$

such evaluation problem will be addressed later in an approximate manner, by generating random sample paths approximately distributed according to abovementionned Gibbs–Boltzmann distribution

Summary usefulness of a *prior model*

- to complement information coming from observations
- to connect observations received at different time instants (needed, just because a mobile ... usually moves between these time instants $\neq$ fixed parameter)

this can be a *rough* not necessarily accurate model, and it is usually a noisy model (to acknowledge that a model is necessarily wrong, and in an attempt to quantify modelling error, in a statistical manner)

some models

modeles (simplest to most general)

state space model (continuous state space, usually $\mathbb{R}^m$)

- linear, Gaussian noise
- non–linear, Gaussian noise
- general state space model : non–linear, non Gaussian noise

hidden Markov model (HMM), and extensions

- hidden Markov model (hidden state sequence Markov, observations mutually independant / Markov conditionally w.r.t. hidden states)
- partially observed Markov chains (hidden state and observation sequence jointly Markov)

general *state space*

- discrete : finite, countable
- continuous : Euclidian space $\mathbb{R}^m$, differentiable manifold
- hybrid continuous / discrete
- constrained
- graphical (collection de connected edges)

Bayesian estimation (static case)

random vector $Z = (X, Y)$ taking values in $E \times F$ (for example $E = \mathbb{R}^m$, $F = \mathbb{R}^d$) with known joint probability distribution

for instance given by a probability density $p(x, y)$

objective : exploit observation of Y to improve knowledge about hidden component X

given an estimator $\psi(Y)$ (based on observation Y) of a statistics $\phi = \phi(X)$ taking values in $\mathbb{R}^p$, its mean square error is defined as

$$\mathbb{E}|\psi(Y) - \phi(X)|^2$$

trace of covariance matrix of estimation error

$$\mathbb{E}[\,(\psi(Y) - \phi(X))\,(\psi(Y) - \phi(X))^*\,]$$

symmetric $p \times p$ semi–definite positive matrix

MMSE (*minimum mean square error*) estimator : an estimator $\widehat{\phi}$ such that

$$\mathbb{E}[\,(\widehat{\phi}(Y) - \phi(X))\,(\widehat{\phi}(Y) - \phi(X))^*\,] \leq \mathbb{E}[\,(\psi(Y) - \phi(X))\,(\psi(Y) - \phi(X))^*\,]$$

in symmetric matrix sense, a fortiori such that

$$\mathbb{E}|\widehat{\phi}(Y) - \phi(X)|^2 \leq \mathbb{E}|\psi(Y) - \phi(X)|^2$$

for any other estimator ψ

implicit (non–constructive) definition $\longrightarrow$ any more explicit formulation ?

Proposition MMSE estimator of statistics $\phi = \phi(X)$ in view of observation Y coincides with conditionnal mean of $\phi(X)$ given Y

$$\widehat{\phi}(Y) = \mathbb{E}[\phi(X) \mid Y] = \int_E \phi(x)\, \mathbb{P}[X \in dx \mid Y]$$

for instance

$$\widehat{\phi}(y) = \int_E \phi(x)\, p(x \mid y)\, dx \qquad \text{with} \qquad p(x \mid y) = \frac{p(x,y)}{p(y)}$$

exemple : Gaussian model (static case)

explicit MMSE estimator in Gaussian static case $\neq$ dynamical case

Gaussian random vector $Z = (X, Y)$ taking values in $\mathbb{R}^m \times \mathbb{R}^d$ with mean vector $\bar{Z} = (\bar{X}, \bar{Y})$, and covariance matrix $Q_Z = \begin{pmatrix} Q_X & Q_{XY} \\ Q_{YX} & Q_Y \end{pmatrix}$

Proposition if covariance matrix Q_Y is *inversible*, then conditional distribution of X given Y is Gaussian, with mean vector

$$\widehat{X}(Y) = \bar{X} + Q_{XY}\, Q_Y^{-1}\, (Y - \bar{Y})$$

and covariance matrix (does not depend on observation Y)

$$0 \leq R = Q_X - Q_{XY}\, Q_Y^{-1}\, Q_{YX} \leq Q_X$$

Schur complement of matrix Q_Y in block–matrix Q_Z

repeatedly used in correction step of Kalman filter (MMSE estimator in linear Gaussian systems)

linear Gaussian systems : Kalman filter

hidden state sequence $\{X_k\}$ taking values in $\mathbb{R}^m$, such that

$$X_k = F_k \, X_{k-1} + f_k + W_k$$

and observation sequence $\{Y_k\}$ taking values in $\mathbb{R}^d$, such that

$$Y_k = H_k \, X_k + h_k + V_k$$

assumptions :

- initial state X_0 Gaussian vector with mean $\bar{X}_0$ and covariance matrix Q_0^X
- state noise $\{W_k\}$ Gaussian white noise with covariance matrix Q_k^W
- observation noise $\{V_k\}$ Gaussian white noise with covariance matrix Q_k^V
- noise sequences $\{W_k\}$ and $\{V_k\}$ and initial state X_0 mutually independant

interpretation of prior model

$$X_k = F_k\, X_{k-1} + f_k + W_k$$

in terms of uncertainty propagation

- even if state $X_{k-1} = x$ would be known exactly at time instant $(k-1)$, state X_k at time instant k is uncertain, distributed as a Gaussian vector with mean $F_k\, x + f_k$ and covariance matrix Q^W_k

- if state X_{k-1} is uncertain at time instant $(k-1)$, distributed as a Gaussian vector with mean $\bar{X}_{k-1}$ and covariance matrix Q^X_{k-1}, then this uncertainty is propagated to time instant k : even if model noise would not be present, state X_k at time instant k is uncertain, distributed as a Gaussian vector with mean $F_k\, \bar{X}_{k-1} + f_k$ and covariance matrix $F_k\, Q^X_{k-1}\, F^*_k$

Theorem [Kalman filter] if covariance matrix Q_k^V is *invertible*

then conditional distribution of hidden state X_k given observation sequence $Y_{0:k}$ is Gaussian, with mean $\widehat{X}_k$ and covariance matrix P_k, that satisfy following recurrent equations

$$\widehat{X}_k^- = F_k \, \widehat{X}_{k-1} + f_k$$

$$P_k^- = F_k \, P_{k-1} \, F_k^* + Q_k^W$$

and

$$\widehat{X}_k = \widehat{X}_k^- + K_k \, [Y_k - (H_k \, \widehat{X}_k^- + h_k)]$$

$$P_k = [I - K_k \, H_k] \, P_k^-$$

where matrix

$$K_k = P_k^- \, H_k^* \, [H_k \, P_k^- \, H_k^* + Q_k^V]^{-1}$$

is called *Kalman gain*, and with initialization

$$\widehat{X}_0^- = \bar{X}_0 = \mathbb{E}[X_0] \qquad \text{and} \qquad P_0^- = Q_0^X = \mathrm{cov}(X_0)$$

Non–linear and non Gaussian systems, with examples

- non–linear and non Gaussian systems
- hybridized inertial navigation (TAN, terrain aided navigation)
- visual tracking with color histogramme
- tracking a dim target (TBD, track–before–detect)
- indoor navigation

non–linear and non Gaussian systems

▶ evolution of *hidden* state

$$X_k = f_k(X_{k-1}, W_k) \qquad \text{with} \qquad W_k \sim p_k(dw)$$

interpretation as a model for propagating estimators and attached uncertainties

only requirement : easy to *simulate* model uncertainties

$$X_0 \sim \mu_0(dx) \qquad \text{and} \qquad W_k \sim p_k(dw)$$

more generally, for any $x \in E$

$$\mathbb{P}[X_k \in dx' \mid X_{k-1} = x] = Q_k(x, dx')$$

only requirement : easy to *simulate* uncertain transitions

$$X_k \sim Q_k(x, dx')$$

▶ relation between *observation* and *hidden* state

$$Y_k = h_k(X_k) + V_k \qquad \text{with} \qquad V_k \sim q_k(v)\, dv$$

only requirement : easy to *evaluate likelihood* function

$$g_k(x) = q_k(Y_k - h_k(x))$$

consistency between a *possible* state and *true* observation, e.g.

$$g_k(x) \propto \exp\{-\tfrac{1}{2}\, |Y_k - h_k(x)|^2\}$$

more generally, for any $x \in E$

$$\mathbb{P}[Y_k \in dy \mid X_k = x] = g_k(x, y)\, \lambda_k(dy)$$

only requirement : easy to *evaluate likelihood* function (with abuse of notation)

$$g_k(x) = g_k(x, Y_k)$$

objective : recursively estimate *hidden* state X_k

in view of *observations* $Y_{0:k} = (Y_0 \cdots Y_k)$

for instance, numerically approximate Bayesian filter

$$\mu_k(dx) = \mathbb{P}[X_k \in dx \mid Y_{0:k}]$$

non–linear / non Gaussian systems : particle filtering

numerical approximation of Bayesian filter by weighted empirical probability distribution

$$\mu_k^N = \sum_{i=1}^N w_k^i \, \delta_{\xi_k^i} \qquad \text{with} \qquad \sum_{i=1}^N w_k^i = 1$$

associated with a system of N particles, caracterised by

- their *positions* $(\xi_k^1 \cdots \xi_k^N)$
- and positive (nonnegative) normalized *weights* $(w_k^1 \cdots w_k^N)$

system of interacting particles, that

- *explore* independently the state space, by following *prior model*
- and *interact* under a *selection* mechanism, related with how each particle is consistent with current observation, by evaluating a *likelihood* function

SIS (*sequential importance sampling*) algorithm

- prediction : independently for any $i = 1 \cdots N$

$$\xi_k^i = f_k(\xi_{k-1}^i, W_k^i) \qquad \text{with} \qquad W_k^i \sim p_k(dw)$$

- correction : for any $i = 1 \cdots N$

$$w_k^i = \frac{w_{k-1}^i \; q_k(Y_k - h_k(\xi_k^i))}{\displaystyle\sum_{j=1}^N w_{k-1}^j \; q_k(Y_k - h_k(\xi_k^j))}$$

alternatively (and preferably), weights can be used for re–sampling

SIR (*sampling with importance resampling*) algorithm : basic (bootstrap) version

- resampling (selection) : independently for any $i = 1 \cdots N$

  within current population $(\xi_{k-1}^1, \cdots, \xi_{k-1}^N)$, pick an individual $\widehat{\xi}_{k-1}^i$
  in a manner that is related with weights $(w_{k-1}^1, \cdots, w_{k-1}^N)$

- prediction (mutation) : independently for any $i = 1 \cdots N$

$$\xi_k^i = f_k(\widehat{\xi}_{k-1}^i, W_k^i) \qquad \text{with} \qquad W_k^i \sim p_k(dw)$$

- correction (weighting) : for any $i = 1 \cdots N$

$$w_k^i = \frac{q_k(Y_k - h_k(\xi_k^i))}{\displaystyle\sum_{j=1}^N q_k(Y_k - h_k(\xi_k^j))}$$

some examples

twofold purpose

- modeling with non–linear and non (necessarily) Gaussian systems
- behaviour (more than actual performance) of basic bootstrap SIR algorithm

four examples

- non–linear and non Gaussian systems
- hybridized inertial navigation (TAN, terrain aided navigation)
- visual tracking with color histogramme
- tracking a dim target (TBD, track–before–detect)
- indoor navigation

1st example : hybridized inertial navigation (TAN)

funding : DGA, programme d'étude amont *NCT (nouveaux concepts pour la navigation par corrélation de terrain)*, coordination Thalès Communications

- inertial measurements
  - linear acceleration and angular velocity of an aircraft (INS, inertial navigation system)

→ double integration yields inertial estimation of aircraft position and velocity

- prior model for evolution of inertial estimation error
- noisy altimeter measurements of
  - height of aircraft above sea level (baro–altimeter)
  - height of aircraft above terrain (radio–altimeter)

→ noisy measurement of height of terrain below aircraft above sea level

- correlation with a digital map (DTED, digital terrain elevation data) providing height of terrain for any position in horizontal plane

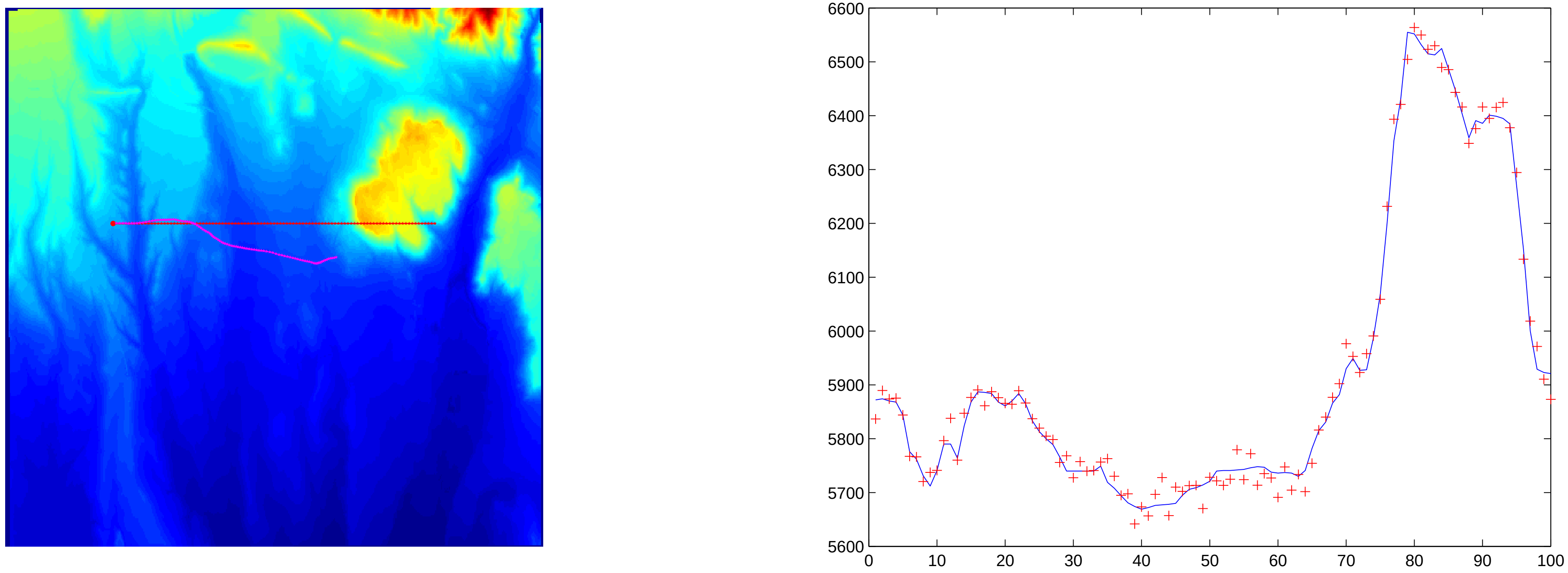


Figure 1: Actual trajectory and estimated trajectory by integration of noisy inertial measurements alone — Terrain profile and noisy altimeter measurements

inertial measurements : noisy aircraft acceleration in horizontal plane

$$a_k^{\rm INS} = a_k + w_k^{\rm INS}$$

estimated trajectory by integration of inertial measurements only

$$\begin{aligned} r_k^{\rm INS} &= r_{k-1}^{\rm INS} + \Delta_k \, v_{k-1}^{\rm INS} \\ v_k^{\rm INS} &= v_{k-1}^{\rm INS} + \Delta_k \, a_k^{\rm INS} \end{aligned}$$

state variable $X_k = (\delta r_k, \delta v_k)$: correction to inertial estimation of position and velocity in horizontal plane, i.e. $\delta r_k = r_k - r_k^{\rm INS}$ and $\delta v_k = v_k - v_k^{\rm INS}$

state model : by difference (or by linearization, in more realistic modeling)

$$\begin{aligned} \delta r_k &= r_{k-1} + \Delta_k \, v_{k-1} - (r_{k-1}^{\rm INS} + \Delta_k \, v_{k-1}^{\rm INS}) = \delta r_{k-1} + \Delta_k \, \delta v_{k-1} \\ \delta v_k &= v_{k-1} + \Delta_k \, a_k - (v_{k-1}^{\rm INS} + \Delta_k \, a_k^{\rm INS}) = \delta v_{k-1} - \Delta_k \, w_k^{\rm INS} \end{aligned}$$

i.e.

$$\begin{pmatrix} \delta r_k \\ \delta v_k \end{pmatrix} = \begin{pmatrix} \delta r_{k-1} + \Delta_k \, \delta v_{k-1} \\ \delta v_{k-1} \end{pmatrix} - \Delta_k \begin{pmatrix} 0 \\ w_k^{\rm INS} \end{pmatrix}$$

observation Y_k : terrain height obtained as difference between

- height of aircraft above see level

$$z_k^{\mathrm{BAR}} = z_k + w_k^{\mathrm{BAR}}$$

- height of aircraft about terrain

$$z_k^{\mathrm{ALT}} = (z_k - h_k(X_k)) + w_k^{\mathrm{ALT}}$$

i.e.

$$Y_k = z_k^{\mathrm{BAR}} - z_k^{\mathrm{ALT}} = h_k(X_k) + V_k \qquad \text{with} \qquad V_k = w_k^{\mathrm{BAR}} - w_k^{\mathrm{ALT}}$$

relation $h(x)$ with hidden state $x = (\delta r, \delta v)$: height of terrain (read of digital map) below position $r_k^{\mathrm{INS}} + \delta r$ in horizontal plane

weighting : likelihood function

$$g_k(x) \propto q_k(Y_k - h_k(x))$$

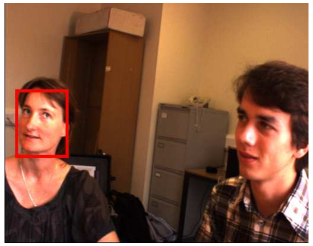
$\longrightarrow$ demo

2nd example : visual tracking by color histogramme

source : Patrick Pérez, Carine Hue, Jako Vermaak and Marc Gangnet, *Color-based probabilistic tracking*, European Conference on Computer Vision (ECCV'02)

user selects a zone within first image of a sequence

objective : automatically track this same zone along whole sequence



initialization

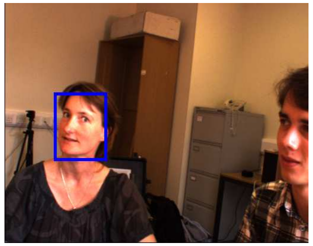


image 2

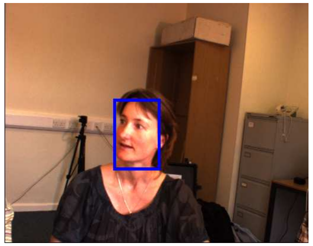


image 3

. . .

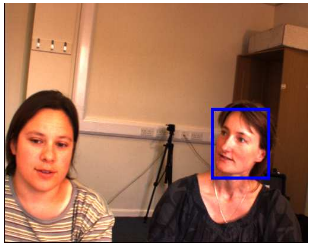
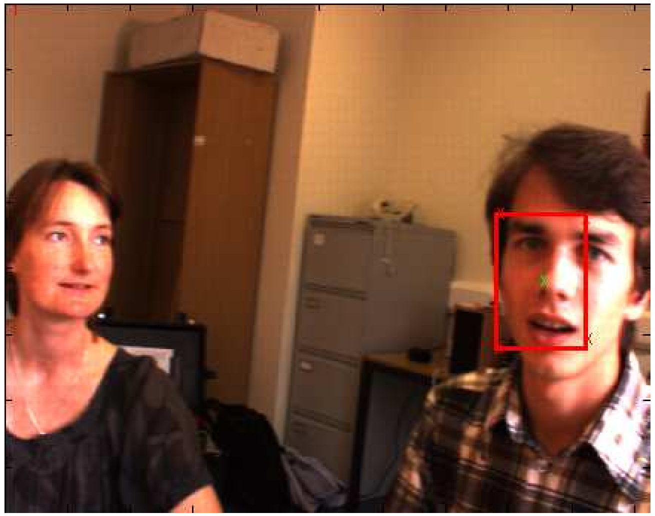


image 10

Figure 2: Tracking a face within a sequence of 10 images

assumption : color histogramme of the zone to be tracked remains constant along whole sequence initial zone is caracterized by a color histogramme (reference histogramme), build with Nb most frequent colors in this zone

normalized number $q^*(n)$ of pixels in initial zone, whose color is $n = 1 \cdots Nb$



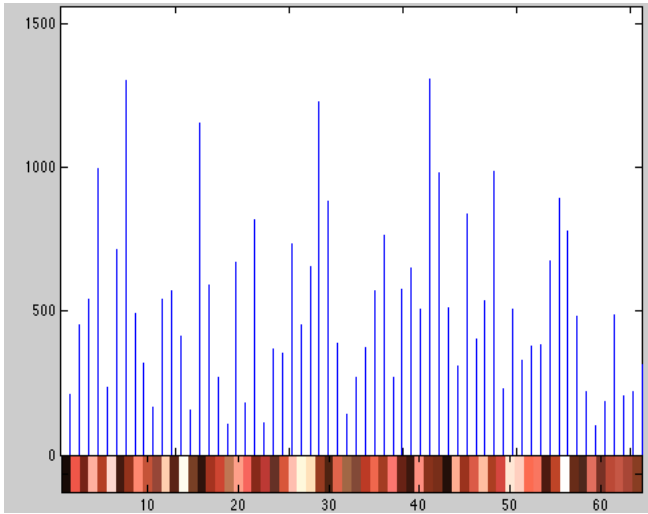


Figure 3: Initial zone and color histogramme associated with Nb=64

to evaluate consistency of a test zone in k–th image, characterized by position (expressed in pixels) of its center x, compute its color histogramme

> normalized number $q_k(x, n)$ of pixels in test zone centered in x in k–th image of sequence, whose closest color is color $n = 1 \cdots Nb$

and define a mesure of distance (Hellinger distance) between the two normalized color histogrammes

$$D(q^*, q_k(x)) = 1 - \sum_{n=1}^{Nb} \sqrt{q^*(n)\, q_k(x, n)} = \tfrac{1}{2} \sum_{n=1}^{Nb} (\sqrt{q^*(n)} - \sqrt{q_k(x, n)})^2$$

state variable $X_k = r_k$: position of center of tracked zone in k–th image

state model : simple random walk

$$r_k = r_{k-1} + W_k$$

observation : k–th image within sequence

weighting of test zone centered in x in k–th image within sequence

$$g_k(x) \propto \exp\{-\lambda\, D(q^*, q_k(x))\}$$

$\longrightarrow$ demo

3rd example : tracking a dim target (TBD)

source : David J. Salmond and H. Birch, *A particle filter for track-before-detect*, American Control Conference (ACC'01)

radar image : rectangular pixel array, where echo intensity received in a pixel is coded in terms of grey level, ranging from darkest (echo with low intensity) to brighter (echo with high intensity)

in principle, if a target is present in physical space, it will appear in image plane in the form of a pixel brighter than other pixels

to detect (and localize) target, it is therefore sufficient to search for brightest pixel, i.e. with highest intensity, or simply to use thresholding

repeating this procedure on each radar image successively, it is therefore possible first to detect, then to track, target in image sequence

high SNR : threshold detection on each image separately, then tracking

vs. low SNR : tracking before (or even without) detection

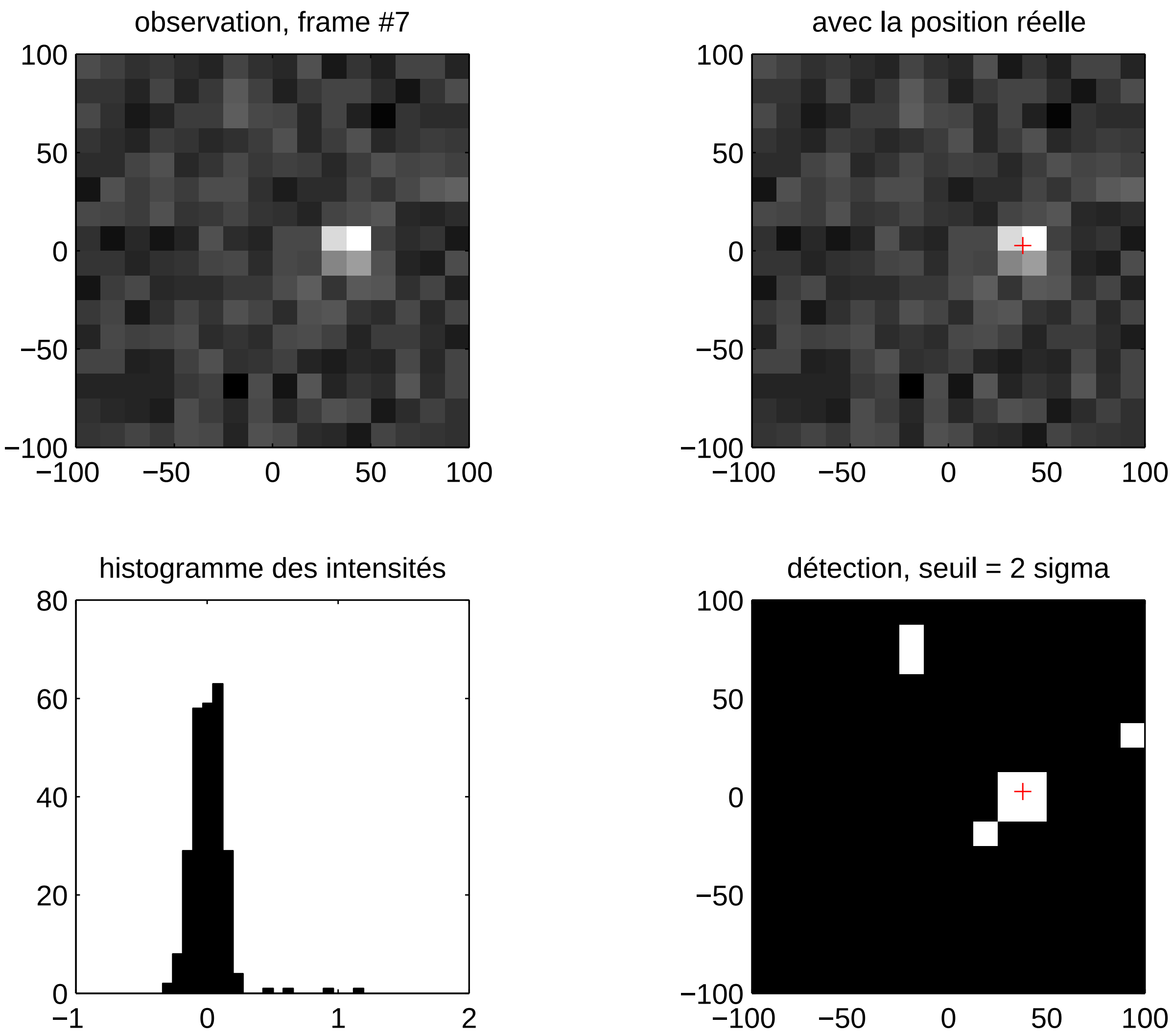


Figure 4: Observed image, actual position, histogramme, detection (visible target)

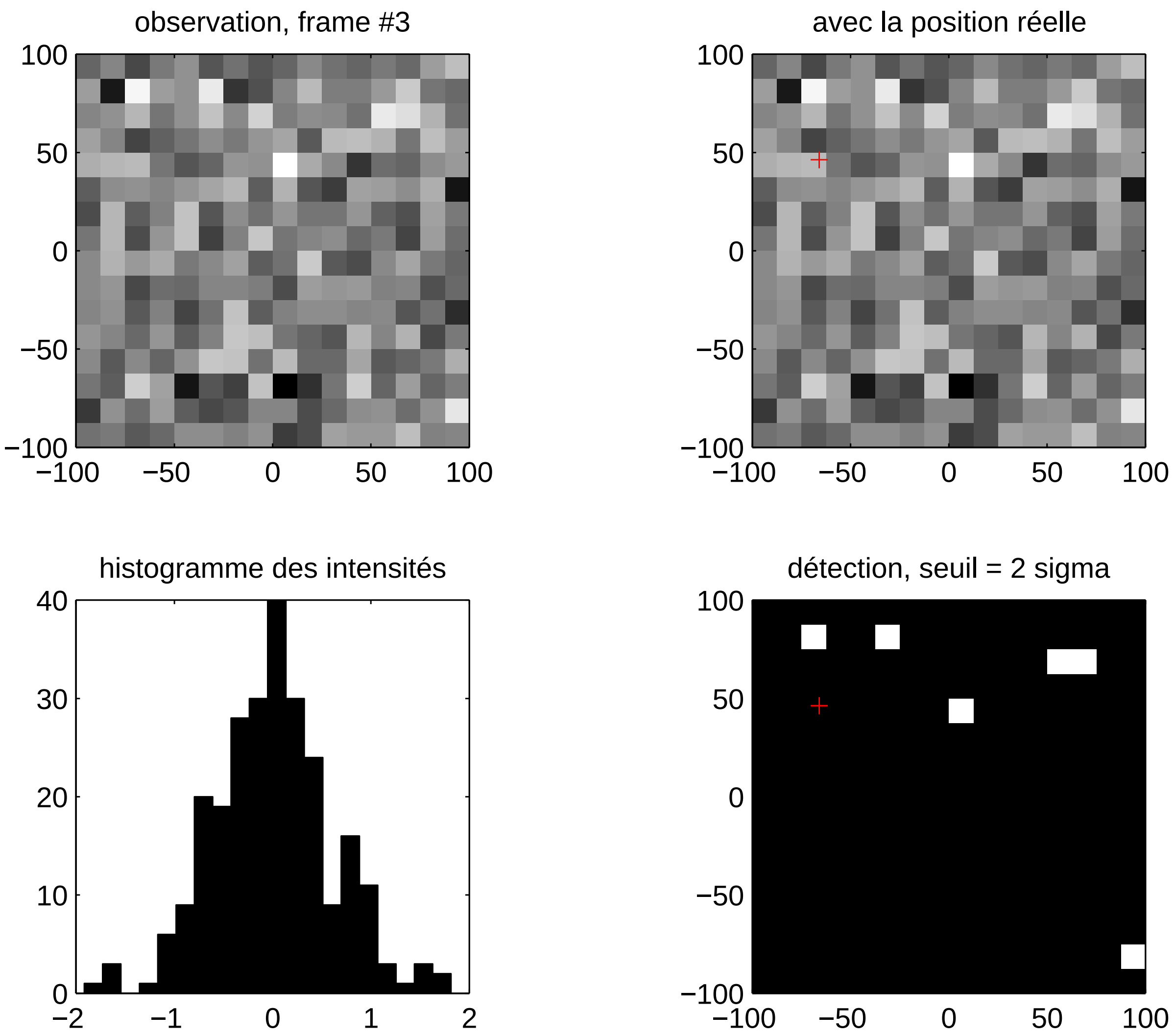


Figure 5: Observed image, actual position, histogramme, detection (dim target)

state variable $X_k = (r_k, v_k)$: target position and velocity

state model : constant velocity motion

allows to search for a consistent motion between successive images

$$r_k = r_{k-1} + \Delta_k \, v_{k-1} \qquad \text{and} \qquad v_k = v_{k-1} + W_k$$

hence

$$\begin{pmatrix} r_k \\ v_k \end{pmatrix} = \begin{pmatrix} r_{k-1} + \Delta_k \, v_{k-1} \\ v_{k-1} \end{pmatrix} + \begin{pmatrix} 0 \\ I \end{pmatrix} W_k$$

observation $Y_k \equiv (Y_k(s) \, , \, s \in S)$: intensity of each pixel in k–th image

$$Y_k(s) = I(X_k, s) + B_k(s)$$

caused either by

- noise only
- or noise and target, if target is present in region of physical space corresponding to considered pixel in image plane

relation $I(x,s)$ with hidden state $x=(r,v)$: intensity of pixel $s \in S$ caused by a target present in position r in physical space

$$I(x,s) = I_0 \, \frac{\delta^2}{2\,\pi\,\sigma^2_{\rm PSF}} \; \exp\{-\frac{|r(s)-r|^2}{2\,\sigma^2_{\rm PSF}}\} \, 1_{\{s \in C(r)\}}$$

where $r(s)$ denotes position in physical space corresponding to pixel center s, where $\delta > 0$ denotes pixel size in physical space, and where set $C(r)$ denotes 9–points neighbourhood in image plane around pixel corresponding to position r in physical space

weighting : likelihood function

$$\begin{aligned} g_k(x) \; &\propto \; \exp\{-\frac{1}{2\,\sigma_B^2} \sum_{s \in S} |Y_k(s) - I(x,s)|^2\} \\ &\propto \; \exp\{\frac{1}{\sigma_B^2} \sum_{s \in C(r)} I(x,s)\, Y_k(s) - \frac{1}{2\,\sigma_B^2} \sum_{s \in C(r)} |I(x,s)|^2\} \end{aligned}$$

$\longrightarrow$ demo

4th example : indoor navigation

funding : ANR, project *FIL (fusion d'information pour la localisation)*, programme Télécommunications, coordination Thalès Alénia Space

- navigation measurements (PNS, pedestrian navigation system)
  - walked distance and direction change between two time instants

$\rightarrow$ integration yields PNS estimation of user position and orientation

- prior model for user motion, i.e. position and orientation evolution, based on incremental PNS measurements
- noisy measurements of
  - distance between user and a ranging beacon with known position and with limited range
- user detection (or non–detection) by a beacon
- fingerprinting : knowledge of motion constraints due to obstacles, available from a map of building

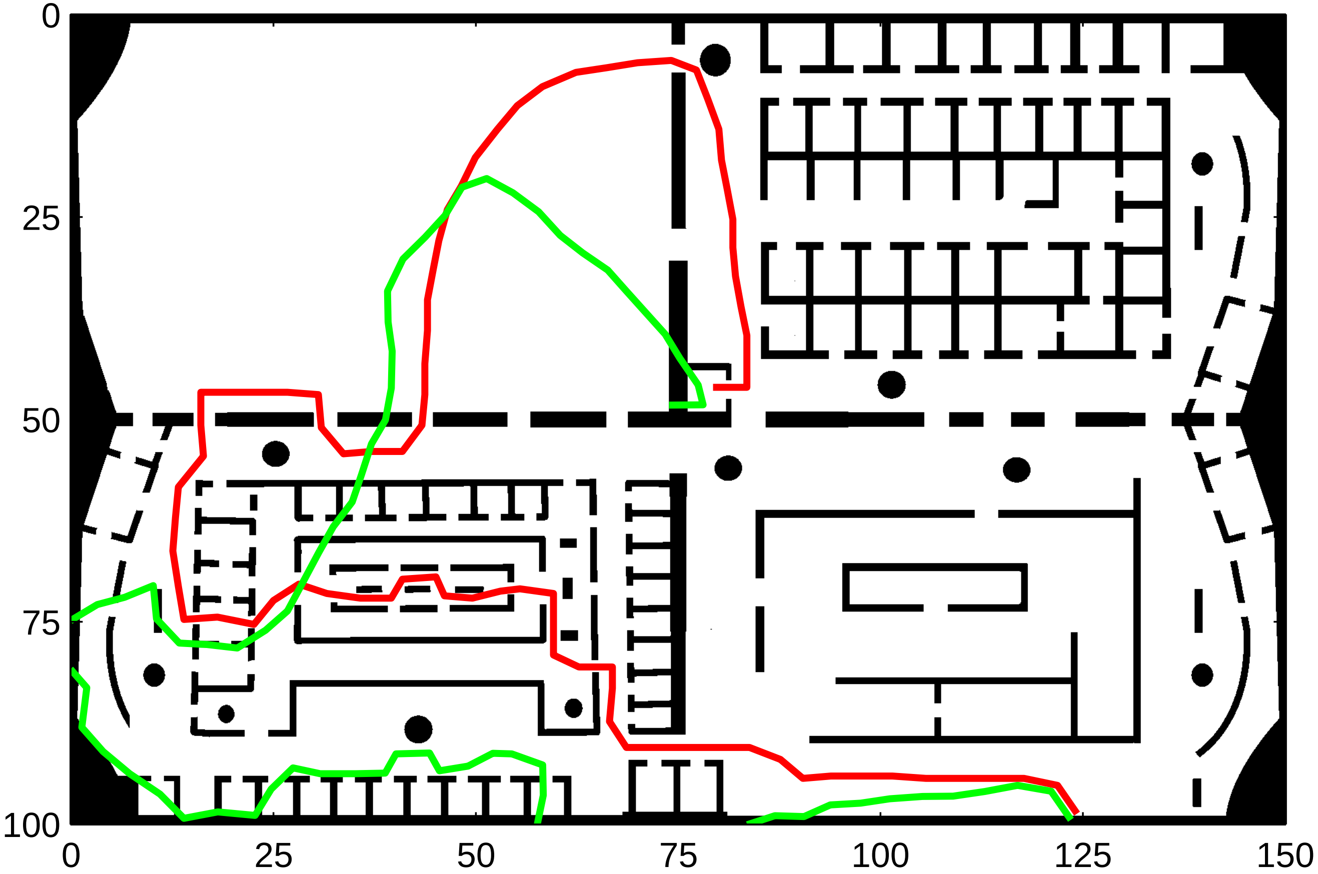


Figure 6: Admissible trajectory and estimated trajectory from PNS measurements alone

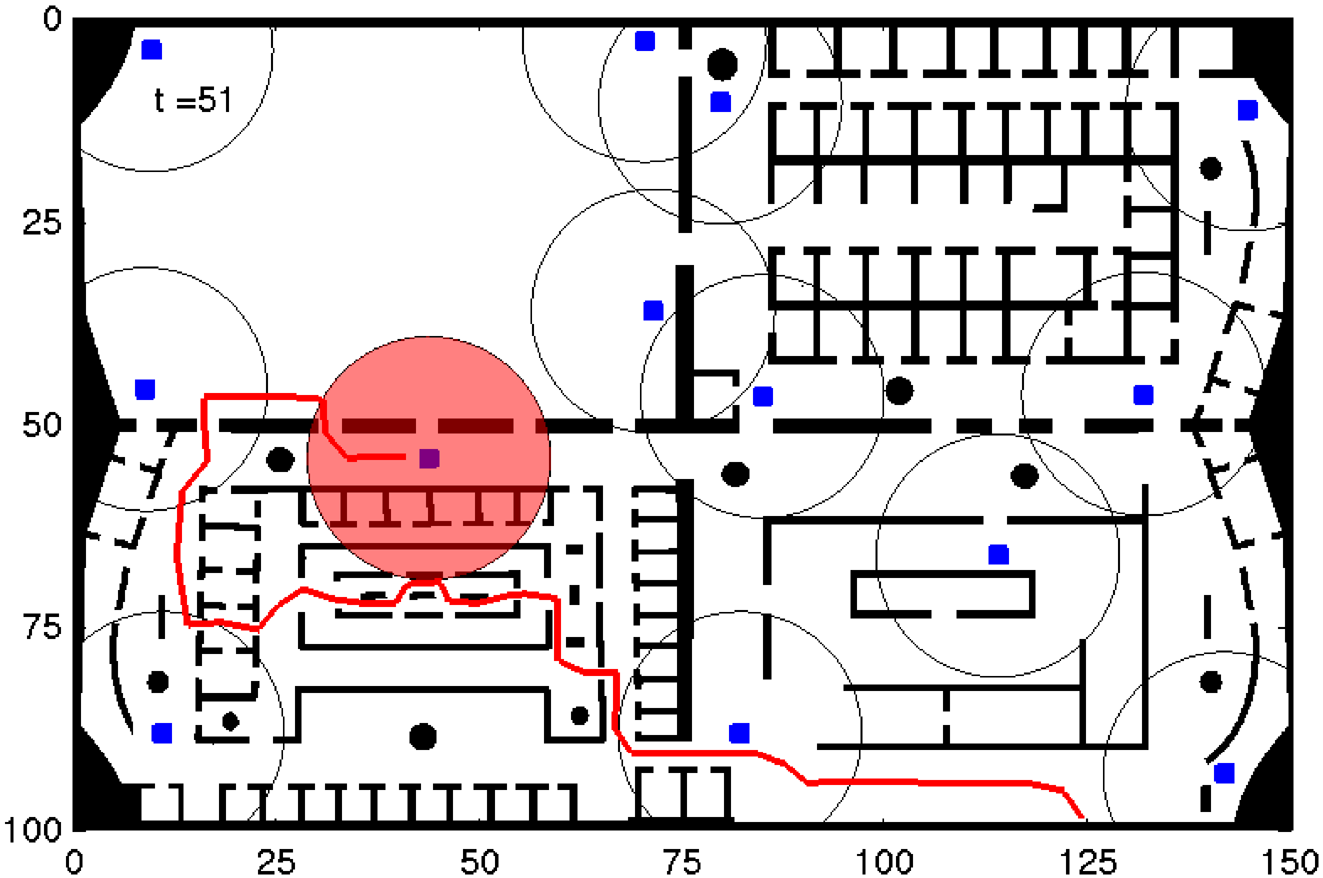


Figure 7: Detection by a ranging beacon

incremental PNS measurements : walked distance and direction change

$$d_k^{\text{PNS}} = d_k + w_k^{\text{walk}} \qquad \text{et} \qquad \alpha_k^{\text{PNS}} = \alpha_k + w_k^{\text{turn}}$$

estimated trajectory by integration of PNS measurements alone

$$r_k^{\text{PNS}} = r_{k-1}^{\text{PNS}} + u(\theta_{k-1}^{\text{PNS}})\, d_k^{\text{PNS}}$$

$$\theta_k^{\text{PNS}} = \theta_{k-1}^{\text{PNS}} + \alpha_k^{\text{PNS}}$$

state variable $X_k = (r_k, \theta_k)$: user position and orientation

state model : incremental update based on PNS measurements

$$r_k = r_{k-1} + u(\theta_{k-1})\, \underbrace{(d_k^{\text{PNS}} - w_k^{\text{walk}})}_{d_k} \qquad \text{and} \qquad \theta_k = \theta_{k-1} + \underbrace{(\alpha_k^{\text{PNS}} - w_k^{\text{turn}})}_{\alpha_k}$$

i.e.

$$\begin{pmatrix} r_k \\ \theta_k \end{pmatrix} = \begin{pmatrix} r_{k-1} + u(\theta_{k-1})\, d_k^{\text{PNS}} \\ \theta_{k-1} + \alpha_k^{\text{PNS}} \end{pmatrix} - \begin{pmatrix} u(\theta_{k-1}) & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} w_k^{\text{walk}} \\ w_k^{\text{turn}} \end{pmatrix}$$

observation Y_k : distance to an active beacon located at position a and with range R

$$Y_k = h(X_k) + V_k$$

relation $h(x)$ with hidden state $x = (r, \theta)$: distance $|r - a|$ to beacon

weighting : likelihood function

$$g_k(x) \propto q_k(Y_k - h(x))$$

if user is / is not detected by beacon, then

$$g_k(x) \propto 1_{\{h(x) \leq R\}} \qquad \text{or else} \qquad g_k(x) \propto 1_{\{h(x) \geq R\}}$$

$\longrightarrow$ demo

to be continued next time . . .

#1 more general models, from non–linear and non Gaussian systems to hidden Markov models and partially observed Markov chains, so as to handle e.g.

- regime / mode switching
- correlation between state noise and observation noise

#2 for each of these models (or just for most general model), representation of

$$\mathbb{P}[X_{0:n} \in dx_{0:n} \mid Y_{0:n}]$$

as a Gibbs–Boltzmann distribution, with recursive formulation, and idem for

$$\mathbb{P}[X_n \in dx_n \mid Y_{0:n}]$$

#3 particle approximation (SIS and SIR algorithms) from either representations

#4 asymptotic behaviour as sample size goes to infinity

#5 numerous algorithmic variants

complément scientifique École Doctorale MATISSE
IRISA et INRIA, salle Markov
jeudi 26 janvier 2012

Variantes Algorithmiques et Justifications Théoriques

François Le Gland
INRIA Rennes et IRMAR

http://www.irisa.fr/aspi/legland/ed-matisse/

Modèles généraux
au–delà des systèmes linéaires gaussiens

- systèmes non–linéaires / non–gaussiens
- modèles de Markov cachés

some notations (continued)

if X is a random variable taking values in E, then mapping

$$\phi \longmapsto \mathbb{E}[\phi(X)] \qquad \text{or equivalently} \qquad A \longmapsto \mathbb{P}[X \in A]$$

defines a probability distribution μ on E, denoted as

$$\mu(dx) = \mathbb{P}[X \in dx]$$

and such that

$$\mathbb{E}[\phi(X)] = \int_E \phi(x)\, \mu(dx) = \langle \mu, \phi \rangle \qquad \text{or} \qquad \mathbb{P}[X \in A] = \mu(A)$$

caracterizes uncertainty about X

transition probability kernel $M(x, dx')$ on E

collection of probability distributions on E indexed by $x \in E$

acts on functions according to

$$M\,\phi(x) = \int_E M(x, dx')\; \phi(x')$$

and acts on probability distributions according to

$$\mu\,M(dx') = \int_E \mu(dx)\; M(x, dx')$$

seen as a mixture distribution caracterized by

$$\begin{aligned}
\langle \mu\,M, \phi \rangle \;&=\; \int_E [\int_E \mu(dx)\; M(x, dx')]\; \phi(x') \\
&=\; \int_E \mu(dx)\; [\int_E M(x, dx')\; \phi(x')] \\
&=\; \langle \mu, M\,\phi \rangle
\end{aligned}$$

product of two transition probability kernels

systèmes non–linéaires / non–gaussiens

modèle a priori pour l'état caché à valeurs dans E

$$X_k = f_k(X_{k-1}, W_k) \qquad \text{avec} \qquad W_k \sim p_k^W(dw)$$

condition initiale $X_0 \sim \eta_0(dx)$

observation à valeurs dans $\mathbb{R}^d$ dans un bruit additif possédant une densité

$$Y_k = h_k(X_k) + V_k \qquad \text{avec} \qquad V_k \sim q_k^V(v)\, dv$$

les v.a. X_0, $W_1, \cdots, W_k, \cdots$ et $V_0, V_1, \cdots, V_k, \cdots$ sont

indépendantes mais pas nécessairement gaussiennes

pour la suite, il suffit de savoir

- *simuler* une v.a. selon $\eta_0(dx)$ ou selon $p_k^W(dw)$
- *évaluer* pour tout $v \in \mathbb{R}^d$, la fonction $q_k^V(v)$

Proposition la suite $\{X_k\}$ est une chaîne de Markov à valeurs dans E, c–à–d que conditionnellement aux états passés $X_{0:k-1}$, l'état présent X_k ne dépend que de X_{k-1} (statistique exhaustive)

$$\mathbb{P}[X_k \in dx \mid X_{0:k-1}] = \mathbb{P}[X_k \in dx \mid X_{k-1}]$$

et la loi de la suite $\{X_k\}$ est complètement déterminée par

- la loi initiale

$$\mathbb{P}[X_0 \in dx] = \eta_0(dx)$$

- et le *noyau de transition*

$$\mathbb{P}[X_k \in dx' \mid X_{k-1} = x] = Q_k(x, dx')$$

Preuve compte tenu que $X_{0:k-1}$ et W_k sont indépendants

$$\begin{aligned}\mathbb{E}[\phi(X_k) \mid X_{0:k-1}] &= \mathbb{E}[\phi(f_k(X_{k-1}, W_k)) \mid X_{0:k-1}] \\ &= \int_{\mathbb{R}^p} \phi(f_k(X_{k-1}, w))\, p_k^W(dw)\end{aligned}$$

ne dépend que de X_{k-1}, c–à–d que

$$\mathbb{E}[\,\phi(X_k) \mid X_{0:k-1}] = \mathbb{E}[\,\phi(X_k) \mid X_{k-1}]$$

d'où la *propriété de Markov*, et

$$\int_E Q_k(x, dx')\,\phi(x') = \mathbb{E}[\,\phi(X_k) \mid X_{k-1} = x] = \int_{\mathbb{R}^p} \phi(f_k(x, w))\, p_k^W(dw)$$

ce qui définit implicitement le noyau de transition □

Remarque si $f_k(x,w) = b_k(x) + w$ et si la distribution de probabilité $p_k^W(dw)$ de la v.a. W_k admet une densité encore notée $p_k^W(w)$, c'est–à–dire si

$$X_k = b_k(X_{k-1}) + W_k \qquad \text{avec} \qquad W_k \sim p_k^W(w)\, dw$$

alors

$$Q_k(x, dx') = p_k^W(x' - b_k(x))\, dx'$$

c'est–à–dire que le noyau $Q_k(x, dx')$ admet une densité

en effet, le changement de variable $x' = b_k(x) + w$ donne immédiatement

$$Q_k\, \phi(x) = \int_{\mathbb{R}^m} \phi(b_k(x) + w)\; p_k^W(w)\, dw = \int_{\mathbb{R}^m} \phi(x')\; p_k^W(x' - b_k(x))\, dx'$$

Remarque en général, le noyau $Q_k(x, dx')$ n'admet pas de densité

en effet, conditionnellement à $X_{k-1} = x$, le v.a. X_k appartient nécessairement au sous–ensemble

$$\mathcal{M}(x) = \{x' \in \mathbb{R}^m \; : \text{ il existe } w \in \mathbb{R}^p \text{ tel que } x' = f_k(x, w)\}$$

et dans le cas où $p < m$ ce sous ensemble $\mathcal{M}(x)$ est généralement, sous certaines hypothèses de régularité, une sous–variété différentielle de dimension p dans l'espace $\mathbb{R}^m$

il ne peut donc pas y avoir de densité par rapport à la mesure de Lebesgue de $\mathbb{R}^m$ pour la distribution de probabilité $Q_k(x, dx')$ du v.a. X_k

Proposition la suite $\{Y_k\}$ vérifie l'hypothèse de *canal sans mémoire*, c–à–d que pour tout instant n

- conditionnellement aux états cachés $X_{0:n}$ les observations $Y_{0:n}$ sont mutuellement indépendantes, ce qui se traduit par

$$\mathbb{P}[Y_{0:n} \in dy_{0:n} \mid X_{0:n}] = \prod_{k=0}^{n} \mathbb{P}[Y_k \in dy_k \mid X_{0:n}]$$

- pour tout $k = 0 \cdots n$, la loi conditionnelle de Y_k sachant $X_{0:n}$ ne dépend que de X_k, ce qui se traduit par

$$\mathbb{P}[Y_k \in dy_k \mid X_{0:n}] = \mathbb{P}[Y_k \in dy_k \mid X_k]$$

avec la *densité d'émission*

$$\mathbb{P}[Y_k \in dy \mid X_k = x] = q_k^V(y - h_k(x))\, dy$$

on définit la *fonction de vraisemblance*

$$g_k(x) = q_k^V(Y_k - h_k(x))$$

qui mesure l'adéquation de $x \in \mathbb{R}^m$ avec l'observation Y_k

la loi conditionnelle jointe des observations $Y_{0:n}$ sachant les états $X_{0:n}$ vérifie

$$\mathbb{P}[Y_{0:n} \in dy_{0:n} \mid X_{0:n} = x_{0:n}] = \prod_{k=0}^{n} q_k^V(y_k - h_k(x_k))\, dy_0 \cdots dy_n$$

Preuve compte tenu que $V_{0:n}$ et $X_{0:n}$ sont mutuellement indépendants

$$\mathbb{E}[\phi_0(Y_0) \cdots \phi_n(Y_n) \mid X_{0:n}]$$

$$= \mathbb{E}[\phi_0(h_0(X_0) + V_0) \cdots \phi_n(h_n(X_n) + V_n) \mid X_{0:n}]$$

$$= \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \phi_0(h_0(X_0) + v_0) \cdots \phi_n(h_n(X_n) + v_n)\, \mathbb{P}[V_{0:n} \in dv_{0:n}]$$

$$= \prod_{k=0}^{n} \int_{\mathbb{R}^d} \phi_k(h_k(X_k) + v_k)\, \mathbb{P}[V_k \in dv_k]$$

$$= \prod_{k=0}^{n} \int_{\mathbb{R}^d} \phi_k(h_k(X_k) + v_k)\, q_k^V(v_k)\, dv_k = \prod_{k=0}^{n} \mathbb{E}[\phi_k(Y_k) \mid X_k]$$

$$= \prod_{k=0}^{n} \int_{\mathbb{R}^d} \phi_k(y_k)\, q_k^V(y_k - h_k(X_k))\, dy_k \qquad \square$$

modèles de Markov cachés

plus généralement, on peut supposer que les états cachés $\{X_k\}$ forment une chaîne de Markov à valeurs dans un espace E qui peut être très général, par exemple

- un espace hybride continu / discret
- un sous–ensemble défini par des contraintes
- une variété différentielle
- un graphe, etc.

y compris trajectoriel, de *noyau de transition*

$$\mathbb{P}[X_k \in dx' \mid X_{k-1} = x] = Q_k(x, dx')$$

et de loi initiale

$$\mathbb{P}[X_0 \in dx] = \eta_0(dx)$$

la loi jointe des états $X_{0:n}$ vérifie

$$\mathbb{P}[X_{0:n} \in dx_{0:n}] = \eta_0(dx_0) \prod_{k=1}^{n} Q_k(x_{k-1}, dx_k)$$

on suppose également que les observations $\{Y_k\}$ vérifient l'hypothèse de *canal sans mémoire*, c–à–d que pour tout instant n

- conditionnellement aux états cachés $X_{0:n}$ les observations $Y_{0:n}$ sont mutuellement indépendantes, ce qui se traduit par

$$\mathbb{P}[Y_{0:n} \in dy_{0:n} \mid X_{0:n}] = \prod_{k=0}^{n} \mathbb{P}[Y_k \in dy_k \mid X_{0:n}]$$

- pour tout $k = 0 \cdots n$, la loi conditionnelle de Y_k sachant $X_{0:n}$ ne dépend que de X_k, ce qui se traduit par

$$\mathbb{P}[Y_k \in dy \mid X_{0:n}] = \mathbb{P}[Y_k \in dy \mid X_k]$$

avec la *densité d'émission*

$$\mathbb{P}[Y_k \in dy \mid X_k = x] = g_k(x,y)\, \lambda_k^F(dy)$$

où la mesure positive $\lambda_k^F(dy)$ définie sur F ne dépend pas de $x \in E$

par abus de notation on définit la *fonction de vraisemblance*

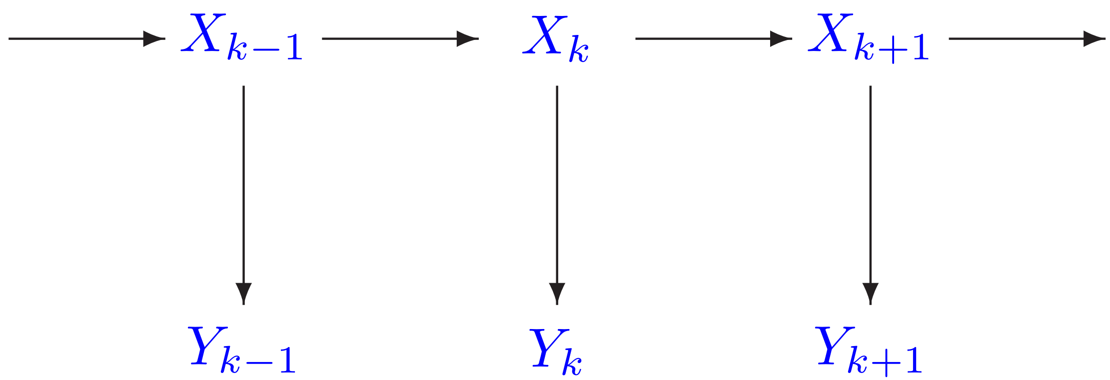
$$g_k(x) = g_k(x, Y_k)$$

qui mesure l'adéquation de $x \in E$ avec l'observation Y_k

la loi conditionnelle jointe des observations $Y_{0:n}$ sachant les états $X_{0:n}$ vérifie

$$\mathbb{P}[Y_{0:n} \in dy_{0:n} \mid X_{0:n} = x_{0:n}] = \prod_{k=0}^{n} g_k(x_k, y_k)\ \lambda_0^F(dy_0) \cdots \lambda_n^F(dy_n)$$

la situation est complètement décrite par



où les flèches représentent la dépendance entre variables aléatoires pour la suite il suffit de savoir

- *simuler* pour tout $x \in E$, une v.a. selon le noyau de transition $Q_k(x, dx')$
- *évaluer* pour tout $x' \in E$, la fonction de vraisemblance $g_k(x')$

pour tout instant $k = 1 \cdots n$

Filtre bayésien pour les modèles de Markov cachés

- représentation probabiliste
- équation récurrente

modèles de Markov cachés : représentation probabiliste

Theorem le filtre bayésien μ_n défini par

$$\langle \mu_n, \phi \rangle = \mathbb{E}[\phi(X_n) \mid Y_{0:n}]$$

admet la représentation probabiliste

$$\langle \mu_n, \phi \rangle = \frac{\langle \gamma_n, \phi \rangle}{\langle \gamma_n, 1 \rangle} \qquad \text{avec} \qquad \langle \gamma_n, \phi \rangle = \mathbb{E}[\phi(X_n) \prod_{k=0}^{n} g_k(X_k)]$$

Remarque l'espérance porte seulement sur les états cachés successifs $X_{0:n}$: les fonctions de vraisemblance $g_k(x)$ sont définies par abus de notation comme

$$g_k(x) = g_k(x, Y_k)$$

pour tout $k = 0, 1 \cdots n$, et dépendent implicitement des observations $Y_{0:n}$, mais celles–ci sont considérées comme *fixées*

Preuve d'après la formule de Bayes, et d'après la propriété de canal sans mémoire, la loi jointe des états cachés $X_{0:n}$ et des observations $Y_{0:n}$ vérifie

$$\mathbb{P}[X_{0:n} \in dx_{0:n}, Y_{0:n} \in dy_{0:n}] = \mathbb{P}[Y_{0:n} \in dy_{0:n} \mid X_{0:n} = x_{0:n}]\, \mathbb{P}[X_{0:n} \in dx_{0:n}]$$

$$= \mathbb{P}[X_{0:n} \in dx_{0:n}] \prod_{k=0}^{n} g_k(x_k, y_k)\, \lambda_0^F(dy_0) \cdots \lambda_n^F(dy_n)$$

en intégrant par rapport aux variables $x_{0:n}$, on obtient la loi jointe des observations $Y_{0:n}$, c–à–d

$$\mathbb{P}[Y_{0:n} \in dy_{0:n}] = \int_E \cdots \int_E \prod_{k=0}^{n} g_k(x_k, y_k)\, \mathbb{P}[X_{0:n} \in dx_{0:n}]\, \lambda_0^F(dy_0) \cdots \lambda_n^F(dy_n)$$

$$= \mathbb{E}[\prod_{k=0}^{n} g_k(X_k, y_k)]\, \lambda_0^F(dy_0) \cdots \lambda_n^F(dy_n)$$

d'après la formule de Bayes, il vient

$$\mathbb{P}[X_{0:n} \in dx_{0:n}, Y_{0:n} \in dy_{0:n}]$$

$$= \mathbb{P}[X_{0:n} \in dx_{0:n}] \prod_{k=0}^{n} g_k(x_k, y_k) \; \lambda_0^F(dy_0) \cdots \lambda_n^F(dy_n)$$

$$= \mathbb{P}[X_{0:n} \in dx_{0:n} \mid Y_{0:n} = y_{0:n}] \; \mathbb{P}[Y_{0:n} \in dy_{0:n}]$$

$$= \mathbb{P}[X_{0:n} \in dx_{0:n} \mid Y_{0:n} = y_{0:n}] \; \mathbb{E}[\prod_{k=0}^{n} g_k(X_k, y_k)] \; \lambda_0^F(dy_0) \cdots \lambda_n^F(dy_n)$$

et on obtient

$$\mathbb{P}[X_{0:n} \in dx_{0:n} \mid Y_{0:n} = y_{0:n}] = \frac{\displaystyle\prod_{k=0}^{n} g_k(x_k, y_k) \; \mathbb{P}[X_{0:n} \in dx_{0:n}]}{\displaystyle\mathbb{E}[\prod_{k=0}^{n} g_k(X_k, y_k)]}$$

pour toute suite $y_{0:n}$ d'observations pour toute fonction ϕ définie sur E

$$\mathbb{E}[\phi(X_n) \mid Y_{0:n} = y_{0:n}] = \frac{\displaystyle\int_E \cdots \int_E \phi(x_n) \prod_{k=0}^n g_k(x_k, y_k)\, \mathbb{P}[X_{0:n} \in dx_{0:n}]}{\displaystyle\mathbb{E}[\prod_{k=0}^n g_k(X_k, y_k)]}$$

$$= \frac{\displaystyle\mathbb{E}[\,\phi(X_n) \prod_{k=0}^n g_k(X_k, y_k)\,]}{\displaystyle\mathbb{E}[\prod_{k=0}^n g_k(X_k, y_k)\,]}$$

comme cette identité est vérifiée pour toute suite $y_{0:n}$ d'observations

$$\langle \mu_n, \phi \rangle = \mathbb{E}[\phi(X_n) \mid Y_{0:n}] = \frac{\displaystyle\mathbb{E}[\phi(X_n) \prod_{k=0}^n g_k(X_k)\,]}{\displaystyle\mathbb{E}[\prod_{k=0}^n g_k(X_k)\,]} = \frac{\langle \gamma_n, \phi \rangle}{\langle \gamma_n, 1 \rangle} \qquad \square$$

de la même manière, le prédicteur bayésien μ_n^- défini par

$$\langle \mu_n^-, \phi \rangle = \mathbb{E}[\phi(X_n) \mid Y_{0:n-1}]$$

admet la représentation probabiliste

$$\langle \mu_n^-, \phi \rangle = \frac{\langle \gamma_n^-, \phi \rangle}{\langle \gamma_n^-, 1 \rangle} \qquad \text{avec} \qquad \langle \gamma_n^-, \phi \rangle = \mathbb{E}[\phi(X_n) \prod_{k=0}^{n-1} g_k(X_k)]$$

filtre bayésien : modèles de Markov cachés : décomposition d'importance

décomposition (non unique)

$$\gamma_0(dx) = g_0(x)\, \eta_0(dx) = g_0^{\text{imp}}(x)\, \eta_0^{\text{imp}}(dx)$$

et

$$R_k(x, dx') = Q_k(x, dx')\, g_k(x') = g_k^{\text{imp}}(x, x')\, Q_k^{\text{imp}}(x, dx')$$

comme le produit de

- la *fonction de pondération* positive $g_0^{\text{imp}}(x)$ ou $g_k^{\text{imp}}(x, x')$
- et la distribution de probabilité $\eta_0^{\text{imp}}(dx)$ ou le *noyau markovien* $Q_k^{\text{imp}}(x, dx')$

respectivement, où en pratique la décomposition doit être telle qu'il est facile

- de *simuler* pour tout $x \in E$, une v.a. selon le noyau markovien $Q_k^{\text{imp}}(x, dx')$
- d'*évaluer* pour tout $x, x' \in E$, la fonction de pondération $g_k^{\text{imp}}(x, x')$

avec la convention $g_0^{\text{imp}}(x, x') = g_0^{\text{imp}}(x')$

attention : la fonction de pondération $g_k^{\text{imp}}(x, x')$ dépend de la transition (x, x') et pas seulement de l'état d'arrivée x'

pour une décomposition d'importance donnée

$$\begin{aligned}
\langle \gamma_n , \phi \rangle &= \mathbb{E}[\phi(X_n) \prod_{k=0}^n g_k(X_k)\,] \\
&= \int_E \cdots \int_E \phi(x_n)\, \eta_0(dx_0) \prod_{k=1}^n Q_k(x_{k-1}, dx_k) \prod_{k=0}^n g_k(x_k) \\
&= \int_E \cdots \int_E \phi(x_n)\, \eta_0^{\rm imp}(dx_0) \prod_{k=1}^n Q_k^{\rm imp}(x_{k-1}, dx_k) \prod_{k=0}^n g_k^{\rm imp}(x_{k-1}, x_k) \\
&= \mathbb{E}[\phi(X_n^{\rm imp}) \prod_{k=0}^n g_k^{\rm imp}(X_{k-1}^{\rm imp}, X_k^{\rm imp})\,]
\end{aligned}$$

où la suite $\{X_k^{\rm imp}\}$ est une chaîne de Markov, caractérisée par

- la loi initiale $\eta_0^{\rm imp}(dx)$
- et le *noyau de transition* $Q_k^{\rm imp}(x, dx')$

et avec une *fonction de pondération / sélection* $g_k^{\rm imp}(x, x')$ qui dépend de la transition (x, x') et pas seulement de l'état d'arrivée x'

modèles de Markov cachés : équation récurrente

Theorem la suite $\{\mu_k\}$ vérifie l'équation récurrente suivante

$$\mu_{k-1} \xrightarrow{\text{prédiction}} \mu_k^- = \mu_{k-1} \, Q_k \xrightarrow{\text{correction}} \mu_k = g_k \cdot \mu_k^-$$

avec la condition initiale $\mu_0^- = \eta_0$

Remarque dans l'énoncé du théorème, la notation

$$\mu_{k-1} \, Q_k(dx') = \int_E \mu_{k-1}(dx) \, Q_k(x, dx')$$

désigne la distribution de mélange obtenue par l'action du noyau markovien $Q_k(x, dx')$ sur la distribution de probabilité $\mu_{k-1}(dx)$, et la notation

$$g_k \cdot \mu_k^- = \frac{g_k \; \mu_k^-}{\langle \mu_k^- \, , g_k \rangle}$$

désigne le produit (projectif) de la distribution de probabilité a priori $\mu_k^-(dx')$ et de la fonction de vraisemblance $g_k(x')$

Preuve du Théorème on procède en deux étapes, correspondant respectivement aux étapes de prédiction et de correction, et en raisonnant d'abord sur les versions non normalisées, puis en normalisant

▶ étape de prédiction expression de μ_n^- en fonction de μ_{n-1}

on remarque immédiatement que

$$\langle \gamma_n^-, 1 \rangle = \mathbb{E}[\prod_{k=0}^{n-1} g_k(X_k)] = \langle \gamma_{n-1}, 1 \rangle$$

c–à–d que la constante de normalisation est conservée en utilisant la propriété de Markov, on a

$$
\begin{aligned}
\langle \gamma_n^-, \phi \rangle &= \mathbb{E}[\phi(X_n) \prod_{k=0}^{n-1} g_k(X_k)\,] \\
&= \mathbb{E}[\,\mathbb{E}[\phi(X_n) \mid X_{0:n-1}] \prod_{k=0}^{n-1} g_k(X_k)\,] \\
&= \mathbb{E}[\,\mathbb{E}[\phi(X_n) \mid X_{n-1}] \prod_{k=0}^{n-1} g_k(X_k)\,] \\
&= \mathbb{E}[Q_n\,\phi(X_{n-1}) \prod_{k=0}^{n-1} g_k(X_k)] = \langle \gamma_{n-1}, Q_n\,\phi \rangle = \langle \gamma_{n-1}\,Q_n, \phi \rangle
\end{aligned}
$$

pour toute fonction ϕ définie sur E

la dernière égalité exprime simplement que

$$\begin{aligned}\langle \gamma_{n-1}, Q_n\, \phi \rangle &= \int_E \gamma_{n-1}(dx)\, [\, \int_E Q_n(x, dx')\, \phi(x')\,] \\ &= \int_E [\, \int_E \gamma_{n-1}(dx)\, Q_n(x, dx')\,]\, \phi(x') \\ &= \int_E \gamma_{n-1}\, Q_n(dx')\, \phi(x') = \langle \gamma_{n-1}\, Q_n, \phi \rangle\end{aligned}$$

comme la fonction ϕ est quelconque, on en déduit que

$$\gamma_n^- = \gamma_{n-1}\, Q_n$$

et en normalisant, on obtient

$$\mu_n^- = \frac{\gamma_n^-}{\langle \gamma_n^-, 1 \rangle} = \frac{\gamma_{n-1}\, Q_n}{\langle \gamma_{n-1}, 1 \rangle} = \mu_{n-1}\, Q_n$$

▶ étape de correction expression de μ_n en fonction de μ_n^-

on a simplement

$$\langle \gamma_n, \phi \rangle = \mathbb{E}[\phi(X_n) \prod_{k=0}^{n} g_k(X_k)]$$

$$= \mathbb{E}[\phi(X_n)\, g_n(X_n) \prod_{k=0}^{n-1} g_k(X_k)] = \langle \gamma_n^-, g_n\, \phi \rangle = \langle g_n\, \gamma_n^-, \phi \rangle$$

pour toute fonction ϕ définie sur E, où la dernière égalité exprime simplement que

$$\langle \gamma_n^-, g_n\, \phi \rangle = \int_E [g_n(x)\, \phi(x)]\, \gamma_n^-(dx) = \int_E \phi(x)\, [g_n(x)\, \gamma_n^-(dx)] = \langle g_n\, \gamma_n^-, \phi \rangle$$

comme la fonction ϕ est quelconque, on en déduit que

$$\gamma_n = g_n\, \gamma_n^-$$

et en normalisant, on obtient

$$\mu_n = \frac{\gamma_n}{\langle \gamma_n, 1 \rangle} = \frac{g_n\, \gamma_n^-}{\langle \gamma_n^-, g_n \rangle} = \frac{g_n\, \mu_n^-}{\langle \mu_n^-, g_n \rangle} \qquad \square$$

à suivre : approximation particulaire du filtre bayésien

objectif : approximation des distributions de probabilité définies par la relation de récurrence

$$\mu_{k-1} \xrightarrow{\text{mutation}} \eta_k = \mu_{k-1}\, Q_k \xrightarrow{\text{pondération}} \mu_k = g_k \cdot \eta_k$$

avec la condition initiale $\mu_0 = g_0 \cdot \eta_0$

idée : rechercher une approximation sous la forme de distributions de probabilité empiriques (éventuellement pondérées)

$$\eta_k \approx \eta_k^N = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_k^i} \qquad \text{et} \qquad \mu_k \approx \mu_k^N = \sum_{i=1}^N w_k^i\, \delta_{\xi_k^i} \qquad \text{avec} \qquad \sum_{i=1}^N w_k^i = 1$$

associées à une population de N particules caractérisée par

- les positions $(\xi_k^1, \cdots, \xi_k^N)$ dans E
- et les poids positifs $(w_k^1, \cdots, w_k^N)$

étape de prédiction : à partir de la définition

$$\langle \mu_{k-1}^N \, Q_k, \phi \rangle \; = \; \int \mu_{k-1}^N(dx) \int Q_k(x, dx') \, \phi(x')$$

$$= \sum_{i=1}^N w_{k-1}^i \int Q_k(\xi_{k-1}^i, dx') \, \phi(x')$$

$$= \int [\sum_{i=1}^N w_{k-1}^i \; Q_k(\xi_{k-1}^i, dx') \,] \, \phi(x')$$

pour toute fonction ϕ, de sorte que

$$\mu_{k-1}^N \, Q_k = \sum_{i=1}^N w_{k-1}^i \; m_k^i$$

s'exprime comme un mélange fini, avec

$$m_k^i(dx') = Q_k(\xi_{k-1}^i, dx') \qquad \text{pour tout } i = 1 \cdots N$$

qu'il s'agit d'approximer / échantillonner (nombreuses méthodes possibles)