

# III - Processus markoviens

Romain HÉRAULT

INSA Rouen

Automne 2015



On suppose que l'on dispose d'un certain nombre de mesures (ou observations) bruitées  $\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots\}$  d'un signal aléatoire (ou état)  $\mathcal{X} = \{\mathbf{x}_0, \mathbf{x}_2, \dots\}$ .

- une mesure ou observation réside dans un espace réel de dimension  $m$ ,  $\mathbf{y}_k \in \mathbb{R}^m$  ;
- un état réside dans un espace réel de dimension  $n$ ,  $\mathbf{x}_k \in \mathbb{R}^n$  ;
- $\mathbf{x}_{i:j}$ , représente l'ensemble des états de l'instant  $i$  à l'instant  $j$  ;
- il existe un état  $\mathbf{x}_0$  avant la première observation  $\mathbf{y}_1$

Nous supposons que le signal est **markovien** et nous cherchons à réaliser les tâches suivantes :

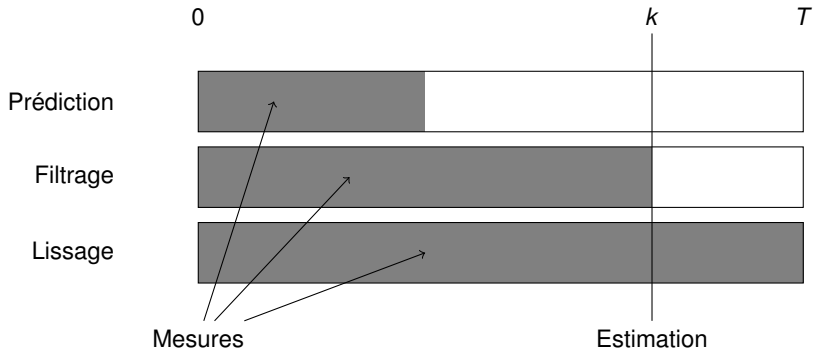
- **Prédiction** : trouver un état futur à partir d'observations passées ;
- **Filtrage** : trouver l'état présent à partir des observations passées et présente ;
- **Lissage ou Explication** : trouver les états passés des observations passées et présente ;
- **Signification statistique** : est-ce les observations que je mesure sont vraisemblable ?

Nous cherchons donc à réaliser les tâches suivantes :

- **Prédiction :**  
Connaissant  $\mathbf{y}_{1:T}$ , trouver une estimation de  $\mathbf{x}_{T+1}$  ;
- **Filtrage :**  
Connaissant  $\mathbf{y}_{1:T}$ , trouver une estimation de  $\mathbf{x}_T$  ;
- **Lissage ou Explication :**  
Connaissant  $\mathbf{y}_{1:T}$ , trouver une estimation de  $\mathbf{x}_{1:T}$  ;
- **Signification statistique :**  
Connaissant un modèle de notre système, calculer la vraisemblance de  $\mathbf{y}_{1:T}$ .

ou encore

- **Prédiction :**  
Connaissant  $\mathbf{y}_{1:t}$ , trouver une estimation de  $\mathbf{x}_k$  avec  $t < k$  ;
- **Filtrage :**  
Connaissant  $\mathbf{y}_{1:t'}$ , trouver une estimation de  $\mathbf{x}_k$  avec  $t' = k$  ;
- **Lissage ou Explication :**  
Connaissant  $\mathbf{y}_{1:T}$ , trouver une estimation de  $\mathbf{x}_k \quad \forall k \in 1 \dots T$  ;
- **Signification statistique :**  
Connaissant un modèle de notre système, calculer la vraisemblance de  $\mathbf{y}_{1:T}$ .



Dans un cadre probabiliste cela revient à chercher :

- **Prédiction :**

$$p(\mathbf{x}_{k+c} | \mathbf{y}_{1:k}) \text{ avec } c \geq 1$$

(distribution *prédictive a posteriori*)

- **Filtrage :**

$$p(\mathbf{x}_k | \mathbf{y}_{1:k})$$

(distribution *a posteriori*)

- **Lissage ou Explication :**

$$p(\mathbf{x}_{0:T} | \mathbf{y}_{1:T})$$

(distribution *a posteriori*)

- **Signification statistique :**

$$p(\mathbf{y}_{1:T})$$

(vraisemblance)

# Ce dont nous disposons

## Modèle dynamique

Le modèle dynamique décrit l'évolution temporelle des états  $\mathbf{x}$ . L'état à l'instant  $t$  ne dépend que de l'état à l'instant  $t - 1$  (processus markovien).

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}) + \mathbf{q} \text{ avec } \mathbf{q} \text{ variable aléatoire ou} \\ \mathbf{x}_k \sim p(\mathbf{x}_k | \mathbf{x}_{k-1}) \text{ sous forme probabiliste .}$$

## Modèle d'observation (ou de mesure)

Le modèle d'observation décrit le lien entre un état  $\mathbf{x}$  et une observation  $\mathbf{y}$ . L'observation à l'instant  $t$  ne dépend que de l'état à l'instant  $t$ .

$$\mathbf{y}_k = g(\mathbf{x}_k) + \mathbf{r} \text{ avec } \mathbf{r} \text{ variable aléatoire ou} \\ \mathbf{y}_k \sim p(\mathbf{y}_k | \mathbf{x}_k) \text{ sous forme probabiliste.}$$

## Distribution *a priori* de l'état initial

$$\mathbf{x}_0 \sim p(\mathbf{x}_0)$$

## Exemple : Voiture subissant une force constante (1)

On considère une voiture ayant une position initiale  $(x_{0,1}, x_{0,2})$  et une vitesse initiale  $(\dot{x}_{0,1}, \dot{x}_{0,2})$  sur lequel on applique une force constante  $(g_{k,1}, g_{k,2}) = cte \quad \forall k$ , on observe une position bruitée de la voiture  $(y_{k,1}, x_{k,2})$ .

### Loi de Newton

$$\mathbf{g}(t) = m \cdot \ddot{\mathbf{x}}(t)$$

où  $m$  est le poids du véhicule

### Problème

L'accélération ne peut pas être déterminée entre deux instants :  
Comment traduire ce problème en processus markovien ?

### Solution

Ne pas se contenter de la position du véhicule comme état du système mais prendre en compte aussi la vitesse et l'accélération !

# Exemple : Voiture subissant une force constante (2)

## Variables

$$\mathbf{x}_k = [x_{k,1}, x_{k,2}, \dot{x}_{k,1}, \dot{x}_{k,2}, \ddot{x}_{k,1}, \ddot{x}_{k,2}]^T$$

$$\mathbf{y}_k = [y_{k,1}, y_{k,2}]^T$$

Initialisation :  $p([x_{0,1}, x_{0,2}, \dot{x}_{0,1}, \dot{x}_{0,2}, g_{0,1}/m, g_{0,2}/m]^T)$

## Modèles

### Dynamique

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}) + \mathbf{q}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & \Delta t & 0 & 0 & 0 \\ 0 & 1 & 0 & \Delta t & 0 & 0 \\ 0 & 0 & 1 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 1 & 0 & \Delta t \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{x}_k = \mathbf{A} \cdot \mathbf{x}_{k-1} + \mathbf{q}_{k-1}$$

avec  $\mathbf{q}_{k-1} \sim \mathcal{N}(0, \mathbf{Q})$

$$\mathbf{x}_k \sim p(\mathbf{x}_k | \mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k | \mathbf{A} \cdot \mathbf{x}_{k-1}, \mathbf{Q})$$

### Observation

$$\mathbf{y}_k = g(\mathbf{x}_k) + \mathbf{r}$$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{y}_k = \mathbf{H} \cdot \mathbf{x}_k + \mathbf{r}_k$$

avec  $\mathbf{r}_k \sim \mathcal{N}(0, \mathbf{R})$

$$\mathbf{y}_k \sim p(\mathbf{y}_k | \mathbf{x}_k) = \mathcal{N}(\mathbf{y}_k | \mathbf{H} \cdot \mathbf{x}_k, \mathbf{R})$$



# Ce que nous cherchons

## La vraisemblance des observations

$$p(\mathbf{y}_{1:T}) = \int_{\mathbf{x}_{0:T}} p(\mathbf{y}_{1:T}|\mathbf{x}_{0:T})p(\mathbf{x}_{0:T})d\mathbf{x}_{0:T}$$

Signification statistique

## Distribution *a posteriori* des états

Inversion du conditionnement :

$$p(\mathbf{x}_{0:T}|\mathbf{y}_{1:T}) = \frac{p(\mathbf{x}_{0:T}, \mathbf{y}_{1:T})}{p(\mathbf{y}_{1:T})} = \frac{p(\mathbf{y}_{1:T}|\mathbf{x}_{0:T})p(\mathbf{x}_{0:T})}{p(\mathbf{y}_{1:T})}$$

Filtrage, Lissage

## Distribution *prédictive a posteriori* des états

$$p(\mathbf{x}_{0:T+1}|\mathbf{y}_{1:T}) = p(\mathbf{x}_{0:T+1}|\mathbf{x}_{0:T}, \mathbf{y}_{1:T})p(\mathbf{x}_{0:T}|\mathbf{y}_{1:T})$$

Prédiction

## Propriété des états

$\mathbf{x}$  suit un processus markovien alors :

- $\mathbf{x}_k$  est indépendant de tout ce qui c'est passé avant  $k - 1$ ,

$$p(\mathbf{x}_k | \mathbf{x}_{1:k-1}, \mathbf{y}_{1:k-1}) = p(\mathbf{x}_k | \mathbf{x}_{k-1}) .$$

- $\mathbf{x}_{k-1}$  est indépendant de tout ce qui ce passera dans le futur,

$$p(\mathbf{x}_{k-1} | \mathbf{x}_{k:T}, \mathbf{y}_{k:T}) = p(\mathbf{x}_{k-1} | \mathbf{x}_k) .$$

## Indépendances conditionnelles des mesures

L'observation courante  $\mathbf{y}_k$  connaissant  $\mathbf{x}_k$  est conditionnellement indépendante des états et mesures passées,

$$p(\mathbf{y}_k | \mathbf{x}_{1:k}, \mathbf{y}_{1:k-1}) = p(\mathbf{y}_k | \mathbf{x}_k) .$$

Connaissant la distribution initiale  $p(\mathbf{x}_0)$  on peut trouver la distribution d'une séquence d'état complète :

$$p(\mathbf{x}_{0:T}) = p(\mathbf{x}_T | \mathbf{x}_{0:T-1}) p(\mathbf{x}_{0:T-1})$$

$$p(\mathbf{x}_{0:T}) = p(\mathbf{x}_T | \mathbf{x}_{T-1}) p(\mathbf{x}_{0:T-1})$$

$$p(\mathbf{x}_{0:T}) = p(\mathbf{x}_T | \mathbf{x}_{T-1}) p(\mathbf{x}_{T-1} | \mathbf{x}_{T-2}) p(\mathbf{x}_{0:T-2})$$

$$p(\mathbf{x}_{0:T}) = p(\mathbf{x}_0) \prod_{k=1}^T p(\mathbf{x}_k | \mathbf{x}_{k-1})$$

De la même manière

$$p(\mathbf{y}_{1:T} | \mathbf{x}_{0:T}) = \prod_{k=1}^T p(\mathbf{y}_k | \mathbf{x}_k)$$

## Ce que nous cherchons (2)

### La vraisemblance des observations

$$p(\mathbf{y}_{1:T}) = \int \dots \int_{\mathbf{x}_{0:T}} p(\mathbf{x}_0) \left( \prod_{k=1}^T p(\mathbf{x}_k | \mathbf{x}_{k-1}) \right) \left( \prod_{k=1}^T p(\mathbf{y}_k | \mathbf{x}_k) \right) d\mathbf{x}_{0:T}$$

### Distribution *a posteriori* des états

$$p(\mathbf{x}_{0:T} | \mathbf{y}_{1:T}) = \frac{p(\mathbf{x}_0) \left( \prod_{k=1}^T p(\mathbf{x}_k | \mathbf{x}_{k-1}) \right) \left( \prod_{k=1}^T p(\mathbf{y}_k | \mathbf{x}_k) \right)}{p(\mathbf{y}_{1:T})}$$

Complexité  $\Rightarrow$  Intraitable !

On va calculer les distributions *a posteriori*  $p(\mathbf{x}_k | \mathbf{y}_{1:k})$  et *prédictive a posteriori*  $p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k})$  par récursion.

### Initialisation

On commence avec la distribution *a priori* de l'état initial  $p(\mathbf{x}_0)$ .

### Récursion

On répète deux étapes au cours de la récursion :

- Prédiction
- Mise à jour

### Étape de prédiction

A partir du modèle de dynamique  $p(\mathbf{x}_k | \mathbf{x}_{k-1})$  et de la distribution *a posteriori*  $p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1})$ , on peut calculer la distribution *prédictive a posteriori*  $p(\mathbf{x}_k | \mathbf{y}_{1:k-1})$ ,

$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1} .$$

C'est l'équation de Chapman-Kolmogorov.

### Étape de mise-à-jour

On prend en compte l'observation  $\mathbf{y}_k$  pour trouver la distribution *a posteriori*  $p(\mathbf{x}_k | \mathbf{y}_{1:k})$  à partir du modèle d'observation  $p(\mathbf{y}_k | \mathbf{x}_k)$  et de la distribution *prédictive a posteriori*  $p(\mathbf{x}_k | \mathbf{y}_{1:k-1})$ ,

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) = \frac{1}{Z_k} p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) ,$$

avec comme terme de normalisation

$$Z_k = \int p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) d\mathbf{x}_k .$$

C'est une application de la règle de Bayes.

On boucle ces deux étapes.

## Prédiction à plus 1 pas de temps

Grâce au modèle de dynamique  $p(\mathbf{x}_k|\mathbf{x}_{k-1})$  (ou  $\mathbf{x}_k = f(\mathbf{x}_{k-1}) + q$ ), on peut projeter à plus d'un pas dans le temps. On applique l'équation de Chapman-Kolmogorov autant de fois que nécessaire,

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})d\mathbf{x}_{k-1} ,$$

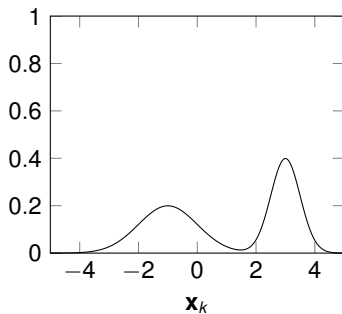
$$p(\mathbf{x}_{k+1}|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_{k+1}|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})d\mathbf{x}_k ,$$

$$p(\mathbf{x}_{k+2}|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_{k+2}|\mathbf{x}_{k+1})p(\mathbf{x}_{k+1}|\mathbf{y}_{1:k-1})d\mathbf{x}_{k+1} ,$$

...

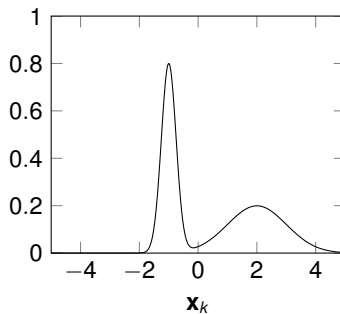
Distribution *a priori*

$$p(\mathbf{x}_k | \mathbf{y}_{1:k})$$



Modèle de dynamique

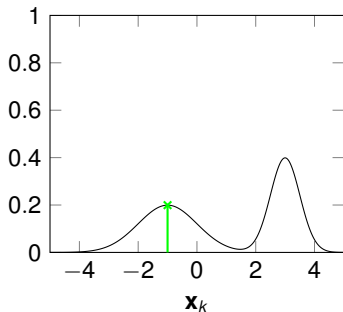
$$p(\mathbf{x}_k | \mathbf{x}_{k-1})$$





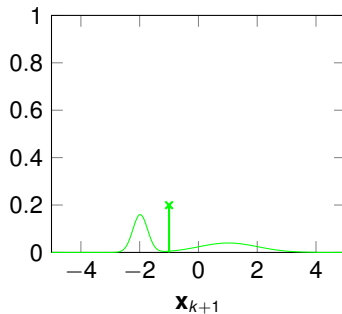
Distribution *a priori*

$$p(\mathbf{x}_k | \mathbf{y}_{1:k})$$



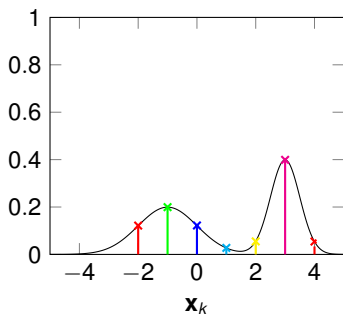
Application du modèle de dynamique

$$p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k})$$



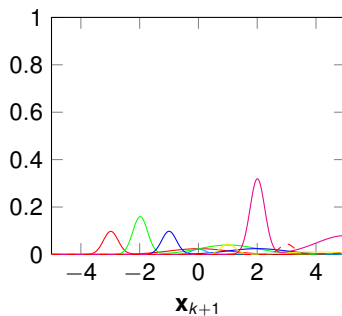
Distribution *a priori*

$$p(\mathbf{x}_k | \mathbf{y}_{1:k})$$



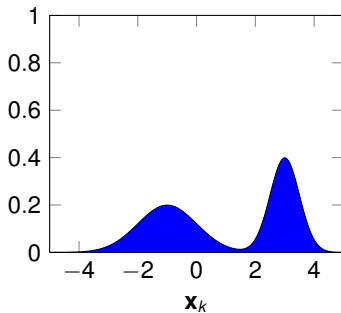
Application du modèle de dynamique

$$p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k})$$



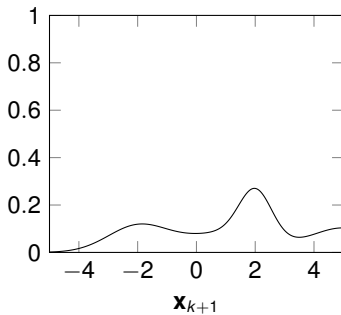
Distribution *a priori*

$$p(\mathbf{x}_k | \mathbf{y}_{1:k})$$

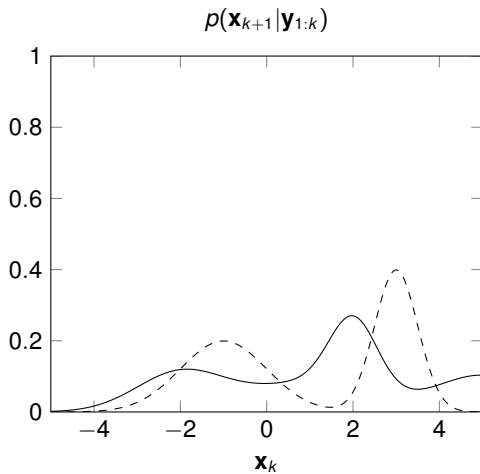


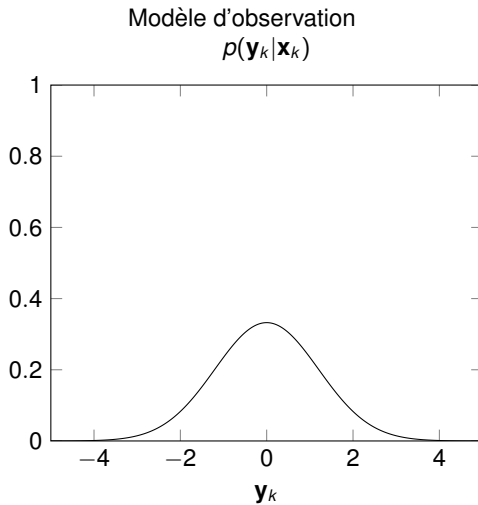
Distribution prédictive

$$p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k})$$



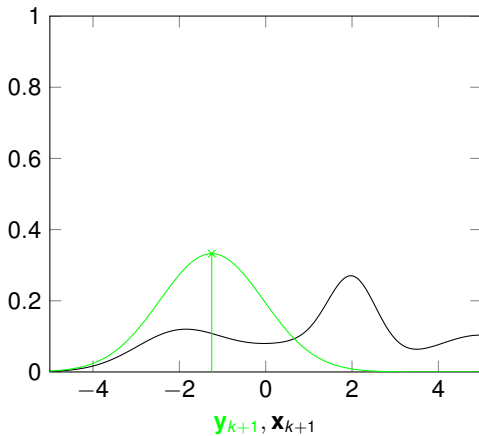
## Illustration : Prédiction (4)

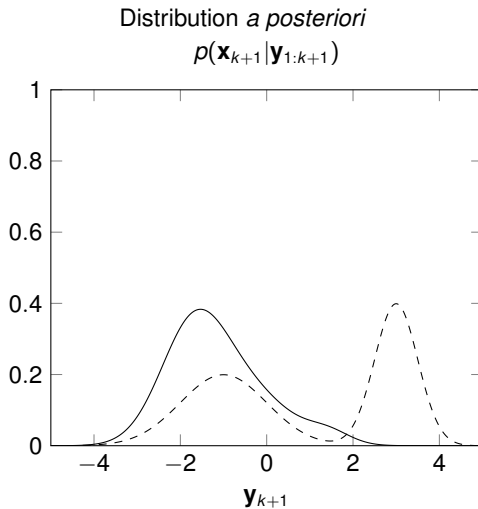




Une nouvelle observation arrive

$$p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k})$$





## Attention

$$p(\mathbf{x}_k | \mathbf{y}_{1:T}) \neq p(\mathbf{x}_k | \mathbf{y}_{1:k})$$



## Lissage (2)

Algorithme en deux passes.

### Passe 1 : Forward (Filtrage)

Initialisation :

$$p(\mathbf{x}_0)$$

Pour  $k$  allant de 1 à  $T$  :

$$\begin{aligned} p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) &= \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1} , \\ p(\mathbf{x}_k | \mathbf{y}_{1:k}) &= \frac{1}{Z_k} p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) , \end{aligned}$$

avec comme terme de normalisation

$$Z_k = \int p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) d\mathbf{x}_k .$$

## Passe 2 : Backward

Initialisation :

$p(\mathbf{x}_T | \mathbf{y}_{1:T}) \leftarrow$  Dernière itération de la première passe

Pour  $k$  allant de  $T - 1$  à 1 :

$$p(\mathbf{x}_k | \mathbf{y}_{1:T}) = p(\mathbf{x}_k | \mathbf{y}_{1:k}) \int \left[ \frac{p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_{k+1} | \mathbf{y}_{1:T})}{p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k})} \right] d\mathbf{x}_{k+1} ,$$

On peut simplifier, réduire la complexité des équations des filtrage/prédiction/lissage et signification statistique suivant les propriétés ou les approximations des modèles .

## Modèles

Dynamique	Observation
$\mathbf{x}_k = f(\mathbf{x}_{k-1}) + \mathbf{q}$	$\mathbf{y}_k = g(\mathbf{x}_k) + \mathbf{r}$
$\mathbf{x}_k \sim p(\mathbf{x}_k   \mathbf{x}_{k-1})$	$\mathbf{y}_k \sim p(\mathbf{y}_k   \mathbf{x}_k)$

- $\mathbf{x}$  discret,  $\mathbf{y}$  discret  $\rightarrow$  HMM ;
- $\mathbf{x}$  discret,  $\mathbf{y}$  continue  $\rightarrow$  GHMM, ... ;
- $\mathbf{x}$  continue,  $\mathbf{y}$  continue,  $f$  et  $g$  linéaire  $\rightarrow$  Filtre de Kalman, lissage de Rauch-Tung-Striebel ;
- $\mathbf{x}$  continue,  $\mathbf{y}$  continue,  $f$  et  $g$  non linéaire mais dérivable  $\rightarrow$  Filtre de Kalman étendu, ... ;
- $\mathbf{x}$  continue,  $\mathbf{y}$  continue,  $f$  et  $g$  quelconques  $\rightarrow$  Filtre particulière, ... ;

Espaces d'état et d'observation continus,  $f$  et  $g$  linéaire, bruits gaussien  $\Rightarrow$

## Filtre de Kalman

### Modèles

$$p(\mathbf{x}_k | \mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k | \mathbf{A} \cdot \mathbf{x}_{k-1}, \mathbf{Q}) \quad p(\mathbf{y}_k | \mathbf{x}_k) = \mathcal{N}(\mathbf{y}_k | \mathbf{H} \cdot \mathbf{x}_k, \mathbf{R})$$

### Notations

$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{x}_k | \mathbf{m}_k^-, \mathbf{P}_k^-) ,$$

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) = \mathcal{N}(\mathbf{x}_k | \mathbf{m}_k, \mathbf{P}_k) ,$$

$$p(\mathbf{y}_k | \mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{y}_k | \mathbf{H} \mathbf{m}_k^-, \mathbf{S}_k) .$$

### Prédiction

$$\begin{aligned}\mathbf{m}_k^- &= \mathbf{A}\mathbf{m}_{k-1}, \\ \mathbf{P}_k^- &= \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q}.\end{aligned}$$

### Mise à jour

$$\begin{aligned}\mathbf{v}_k &= \mathbf{y}_k - \mathbf{H}\mathbf{m}_k^-, \\ \mathbf{S}_k &= \mathbf{H}\mathbf{P}_k^-\mathbf{H}^T + \mathbf{R}, \\ \mathbf{K}_k &= \mathbf{P}_k^-\mathbf{H}^T\mathbf{S}_k^{-1}, \\ \mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{K}_k\mathbf{v}_k, \\ \mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^T.\end{aligned}$$

## Modèles

### Dynamique

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}) + \mathbf{q}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & \Delta t & 0 & 0 & 0 \\ 0 & 1 & 0 & \Delta t & 0 & 0 \\ 0 & 0 & 1 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 1 & 0 & \Delta t \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{x}_k = \mathbf{A} \cdot \mathbf{x}_{k-1} + \mathbf{q}_{k-1}$$

avec  $\mathbf{q}_{k-1} \sim \mathcal{N}(0, \mathbf{Q})$

$$\mathbf{x}_k \sim p(\mathbf{x}_k | \mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k | \mathbf{A} \cdot \mathbf{x}_{k-1}, \mathbf{Q})$$

### Observation

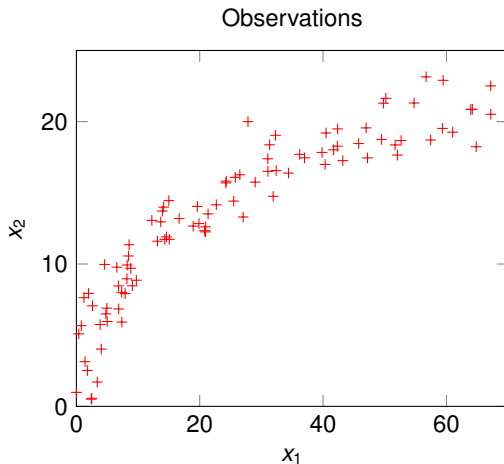
$$\mathbf{y}_k = g(\mathbf{x}_k) + \mathbf{r}$$

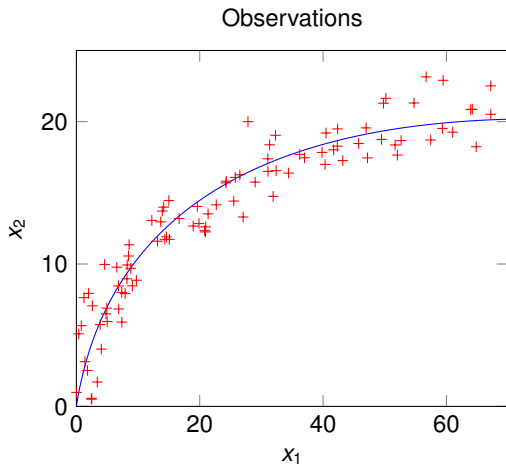
$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{y}_k = \mathbf{H} \cdot \mathbf{x}_k + \mathbf{r}_k$$

avec  $\mathbf{r}_k \sim \mathcal{N}(0, \mathbf{R})$

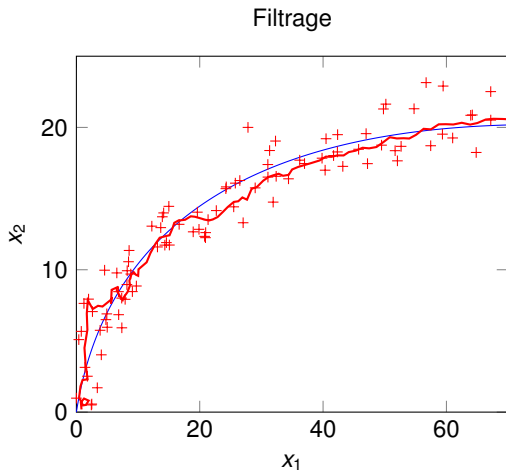
$$\mathbf{y}_k \sim p(\mathbf{y}_k | \mathbf{x}_k) = \mathcal{N}(\mathbf{y}_k | \mathbf{H} \cdot \mathbf{x}_k, \mathbf{R})$$

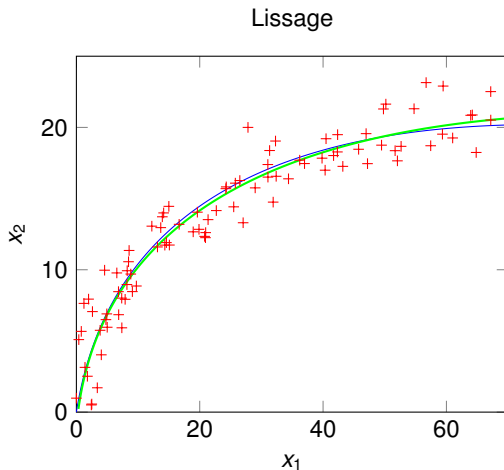






## Résolution : Voiture subissant une force constante (6)





## Résolution : Voiture subissant une force constante (8)

