# ESTIMATION D'ÉTAT: DU FILTRE DE KALMAN AU FILTRE PARTICULAIRE

Qi Cheng

16 février 2010

- 1 Cadre de l'étude et motivations
- 2 Filtre optimal dans un modèle linéaire Gaussien
- 3 Certains filtres non linéaires sous-optimaux
- 4 Le filtre particulaire (PF)
- 5 Le filtre bootstrap modifié (MBF)
- 6 Le filtre particulaire "unscented" modifié (MUPF)
- Conclusion et perspectives

#### Modèle d'état

$$x_t = f(x_{t-1}) + u_t$$
$$y_t = h(x_t) + v_t$$

 $(x_t)$ : Variable d'état

 $(y_t)$ : Variable d'observation

 $(u_t)$ : Bruit d'état

 $(v_t)$ : Bruit d'observation

 $(u_t)$  et  $(v_t)$  sont indépendants

Deux hypothèses :

$$p(x_t|x_{0:t-1}) = p(x_t|x_{t-1})$$
  
$$p(y_t|x_{0:t}, y_{1:t-1}) = p(y_t|x_t)$$

#### Motivations

# Objectif: Estimation d'état Directions:

- Méthode des moindres carrés
- Méthode de Bayes
- Utilisation de la théorie de l'information

#### Contribution : Amélioration des performances du PF

- Quand la variance du bruit d'état est importante, nous proposons le MBF.
- Quand la variance du bruit d'observation est faible, nous proposons le MUPF.

# Bref historique

- Méthode des moindres carrés : Jazwinski (1970), Gelb (1974), Anderson and Moore (1979) ...
  Méthode de Bayes : Sorenson and Alspach (1971), Masreliez (1975)
- Méthode de Bayes: Sorenson and Alspach (1971), Masreliez (1975), Kitagawa (1987), Kramer and Sorenson (1988), Tanizaki (1991), Carlin et al. (1992), Gorden et al. (1993) ...
- Théorie de l'information: Weidemann and Stear (1970), Kalata and Priemer (1974), Tomita et al. (1976), Bucy (1979), Xu et al. (2005), Eyink and Kim (2006), Guo and Wang (2006) ...

# Filtre de Kalmam (KF)

$$x_t = Ax_{t-1} + u_t$$
$$y_t = Bx_t + v_t$$

 $u_t \sim N(0, Q)$  and  $v_t \sim N(0, R)$ .

Prédiction de l'état :  $\hat{x}_{t|t-1} = A\hat{x}_{t-1|t-1}$ .

Covariance de prédiction :  $\hat{P}_{t|t-1} = A\hat{P}_{t-1|t-1}A^T + Q$ .

Innovation :  $e_t = y_t - y_{t|t-1} = y_t - B\hat{x}_{t|t-1}$ .

Covariance d'innovation :  $P_{e_t} = B\hat{P}_{t|t-1}B^T + R$ .

Gain :  $K_t = \hat{P}_{t|t-1} B^T P_{e_t}^{-1}$ .

Correction :  $\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t e_t$ .

 $\hat{P}_{t|t} = [1 - K(t)B]\hat{P}_{t|t-1}.$ 

#### KF: Méthode des moindres carrés

Critère: Minimiser

$$J_t = \sum_{i=1}^t [(y_t - Bx_t)'R^{-1}(y_t - Bx_t) + (x_t - \hat{x}_{t|t-1})'P_{t|t-1}^{-1}(x_t - \hat{x}_{t|t-1})],$$

οù

$$\hat{x}_{t|t-1} = A\hat{x}_{t-1|t-1},$$

$$P_{t|t-1} = AP_{t-1|t-1}A' + Q.$$

Remarque : La méthode des moindres carrés ne donne aucune explication probabiliste.

# Filtrage Bayésien récursif

```
Prédiction : p(x_{t+1}|y_{1:t}) = \int p(x_t|y_{1:t}) p(x_{t+1}|x_t) dx_t Correction : p(x_{t+1}|y_{1:t+1}) = \frac{p(y_{t+1}|x_{t+1}) p(x_{t+1}|y_{1:t})}{p(y_{t+1}|y_{1:t})} A priori : p(x_{t+1}|y_{1:t}) Vraisemblance : p(y_{t+1}|y_{1:t}) = \int p(y_{t+1}|x_{t+1}) p(x_{t+1}|y_{1:t}) Évidence : p(y_{t+1}|y_{1:t}) = \int p(y_{t+1}|x_{t+1}) p(x_{t+1}|y_{1:t})
```

### KF: Maximum a posteriori

$$p(x_t|y_t) = C \exp\left[-\frac{1}{2}(y_t - Bx_t)'R^{-1}(y_t - Bx_t)\right]$$
$$\exp\left[-\frac{1}{2}(x_t - \hat{x}_{t|t-1})'P_{t|t-1}^{-1}(x_t - \hat{x}_{t|t-1})\right].$$

En prenant le logarithme, on obtient

$$\ln p(x_t|y_t) = \ln C - \frac{1}{2}(y_t - Bx_t)'R^{-1}(y_t - Bx_t) - \frac{1}{2}(x_t - \hat{x}_{t|t-1})'P_{t|t-1}^{-1}(x_t - \hat{x}_{t|t-1}).$$

L'estimation du maximum a posteriori est obtenue en résolvant l'équation :

$$\nabla_x \ln p(x_t|y_t) = B'R^{-1}(y_t - Bx_t) - P_{t|t-1}^{-1}(x_t - \hat{x}_{t|t-1}) = 0.$$

### KF: Minimum erreur quadratique moyenne (EQM)

Critère:

$$\hat{x}^{\text{EQM}} = \operatorname{argmin} E[(\hat{x} - x)^2 | Y].$$

$$\hat{x}^{\text{EQM}} = E[x|Y] = \int xp(x|y) dx$$

Pour le modèle linéaire Gaussien

$$p(x_t|y_t) = C \exp\left[-\frac{1}{2}(y_t - Bx_t)'R^{-1}(y_t - Bx_t)\right]$$
$$\exp\left[-\frac{1}{2}(x_t - \hat{x}_{t|t-1})'P_{t|t-1}^{-1}(x_t - \hat{x}_{t|t-1})\right],$$
$$= C \exp\left\{-\frac{1}{2}(x_t - \mu)'L^{-1}(x_t - \mu)\right\},$$

où 
$$\mu = \hat{x}_{t|t-1} + Ke_t$$
 et  $L = (I - K_t B)P_{t|t-1}$ .

#### KF: Utilisation de la théorie de l'information

Critère : Minimiser l'entropie d'erreur.

Définition de l'entropie de Shannon :

$$H(x) = E(-\ln p(x)) = -\int p(x) \ln p(x).$$

Pour une variable aléatoire Gaussienne :  $H(x) = \ln(\sigma\sqrt{2\pi e})$ .

Pour un système linéaire Gaussien, l'erreur d'estimation  $\tilde{x}_t = x_t - \hat{x}_t$  est Gaussienne.

L'entropie d'erreur d'estimation :  $H(\tilde{x}_t) = \ln(\sqrt{(2\pi e)^n \det P_t}$ .

Conclusion : l'estimation au sens du minimum d'entropie d'erreur est égale à l'estimation au sens de l'EQM minimale.

# Filtre de Kalman étendu (EKF)

Prédiction :

$$\hat{x}_{t|t-1} = f(\hat{x}_{t-1|t-1}),$$

$$P_{t|t-1} = F_t P_{t-1|t-1} F_t' + Q,$$

Innovation :

$$e_t = y_t - h(\hat{x}_{t|t-1}),$$
  
 $M_t = H_t P_{t|t-1} H'_t + R,$ 

Correction :

$$K_{t} = P_{t|t-1}H'_{t}M_{t}^{-1},$$

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_{t}e_{t},$$

$$P_{t|t} = (I - K_{t}H_{t})P_{t|t-1},$$

où 
$$F_t = \left. \frac{\partial f(x_{t-1})}{\partial x_{t-1}} \right|_{x_{t-1} = \hat{x}_{t-1|t-1}}$$
 et  $H_t = \left. \frac{\partial h(x_t)}{\partial x_t} \right|_{x_t = \hat{x}_{t|t-1}}$ .

#### Remarques

- L'EKF n'est valable que lorsque la loi a posteriori de l'état peut être bien approximée par une distribution gaussienne. Cela suppose que l'erreur d'estimation initiale et tous les bruits soient suffisamment petits et que la non-linéarité du modèle ne soit pas importante.
- L'EKF n'est pas nécessairement stable. L'estimation de l'état donnée par l'EKF peut diverger.
- Le calcul de la matrice jacobienne est compliqué pour certains modèles.

# Filtre de Kalman unscented (UKF)

- Choix des sigma points.
- Propagation des sigma points par le modèle non linéaire.
- Mise à jour selon les mêmes équations que le KF.

### Remarques

L'algorithme UKF est valable quand la distribution a posteriori de l'état est très proche d'une gaussienne.

- L'UKF est une méthode de quasi-Monte Carlo.
- L'UKF conserve la structure de mise à jour linéaire qui n'est optimale que pour un système linéaire et Gaussien.
- L'UKF utilise seulement les moments du second ordre ce qui n'est valable pour la distribution Gaussienne.
- Le nombre de point n'est pas suffisant pour représenter une distribution compliquée.

Quand la variance de bruit d'observation est faible, l'UKF fonctionne bien.

# Comparaison entre l'EKF et l'UKF

Algo.	Principe	Méthode
EKF	méthode de moindres carrés	développement en séries de Taylor
UKF	méthode de moindres carrés	quasi-Monte Carlo

Tab.: Comparaison entre l'EKF et l'UKF.

### PF: Principe

Implémenter un filtrage Bayésien récursif par une méthode de Monte Carlo. La densité a posteriori requise est représentée par un ensemble d'échantillonnage aléatoire avec des poids associés. [Arulampalam(2002)] Algorithme (Sequential Importance Sampling) : Pour i=1:N (N:N) Nombre de Particule)

- simuler  $x_t^i \sim q(x_t|x_{t-1}^i,y_t)$
- calculer les poids d'importance  $\omega_t^i = \omega_{t-1}^i \frac{p(y_t|x_t^i)p(x_t^i|x_{t-1}^i)}{q(x_t^i|x_{t-1}^i,y_t)}$
- normaliser les poids d'importance  $\overset{\sim}{\omega_t^i} = \frac{\omega_t^i}{\sum_{j=1}^N \omega_t^j}$

### Ré-échantillonnage

Motivation : résoudre le problème de dégénérescence

Méthode : Supprimer les particules de faible poids et dupliquer les particules de fort poids.

Sampling Importance Resampling (SIR):

- simuler  $x_t^i \sim q(x_t|x_{t-1}^i,y_t)$
- calculer les poids d'importance  $\omega_t^i = \omega_{t-1}^i \frac{p(y_t|x_t^i)p(x_t^i|x_{t-1}^i)}{q(x_t^i|x_{t-1}^i,y_t)}$
- normaliser les poids d'importance  $\tilde{\omega_t^i} = \frac{\omega_t^i}{\sum_{j=1}^N \omega_t^j}$
- ré-échantillonnage

### Choix de la Distribution d'importance

La performance du filtrage particulaire dépend fortement de la distribution d'importance :

- Le meilleur choix :  $q(x_t|x_{t-1}^i,y_{0:t}) = p(x_t|x_{t-1}^i,y_t)$ . Mais il est en général impossible d'échantillonner selon  $p(x_t|x_{t-1}^i,y_t)$ .
- Le choix le plus populaire :  $q(x_t|x_{t-1}^i,y_{0:t})=p(x_t|x_{t-1}^i)$ . Cependant,  $p(x_t|x_{t-1}^i)$  n'utilise pas la nouvelle observation  $y_t$  et donc le résultat peut ne pas être bon.
- Produire la distribution d'importance par un filtre auxiliaire e.g. l'EKF ou l'UKF. Ceci permet d'améliorer les performances sous certaines conditions.

# Comparaison entre le PF et l'UKF

Algorithme	Principe	Méthode
PF	filtre bayésien récursif	Monte Carlo
UKF	méthode des moindres carrés	quasi-Monte Carlo

Tab.: Comparaison entre le PF et l'UKF.

# Le filtre particulaire auxiliaire (APF)

- But : Utilisation de l'information contenue dans l'observation.
- Idée de la méthode : Le choix des particules est effectué en introduisant une variable auxiliaire.
- Remarque : Cette méthode est valable lorsque le bruit d'état est faible.

# Le filtre particulaire avec ré-échantillonnage au moyen d'une méthode de Monte Carlo par chaînes de Markov (PF-MCMC)

- But : Réduire la dégénérescence des particules.
- Idée : Ajouter une étape de MCMC après le ré-échantillonnage.
- Remarque : Cette méthode peut améliorer les performances du PF mais au prix d'une complexité calculatoire beaucoup plus grande.

#### Le filtre particulaire basé sur une linéarisation locale

- But : Utilisation de l'information contenue dans l'observation.
- Idée : Utiliser des filtres non linéaires sous-optimaux, e.g. l'EKF, l'UKF pour générer la densité d'importance.
- Problème : Quand les bruits sont importants ou la non-linéarité du système est importante, l'approximation linéaire est mauvaise.

### Remarque sur le PF

En général, On choisit la densité a priori comme distribution d'importance,  $q(x_t|x_{t-1}^i,y_{0:t})=p(x_t|x_{t-1}^i)$ .

- Quand la variance du bruit d'observation est faible, le PF fonctionne mal.
- Quand la variance du bruit d'état est grande, le PF fonctionne mal.

# Principe du MBF

Pour utiliser l'information contenue dans l'observation, nous proposons une nouvelle technique. Notre méthode utilise une technique d'échantillonnage en deux étapes :

- Pour  $j=1,\ldots,M$ , échantillonner  $x_t^{i,j}\sim p(x_t|x_{t-1}^i)$  et calculer la vraisemblance conditionnelle  $p(y_t|x_t^{i,j})$ .
- Choisir la particule  $x_t^{i,j^\star}$  dont la vraisemblance conditionnelle est maximale et prendre  $x_t^i=x_t^{i,j^\star}$ .

Dans la première étape, les particules se déplacent de manière aléatoire selon la densité a priori comme dans le BF, et dans la deuxième étape, l'information  $y_t$  est utilisée pour sélectionner les particules avec le maximum de la vraisemblance conditionnelle.

### Implémentation

- Échantillonnage d'importance : Pour  $j=1,\ldots,M$ ,
  - Échantillonner les particules  $x_t^{i,j} \sim p(x_t|x_{t-1}^i)$
  - ullet Calculer la vraisemblance conditionnelle  $p(y_t|x_t^{i,j})$
  - Choisir  $x_t^{i,j^\star}$  dont  $p(y_t|x_t^{i,j^\star})$  est maximum
  - Prendre  $x_t^i = x_t^{i,j^*}$
- Calculer les poids pour chaque particule
- Normaliser les poids d'importance
- Ré-échantillonnage

#### Remarques

- Dans le MBF, nous sélectionnons les particules échantillonnées selon la densité a priori et ayant une grande vraisemblance conditionnelle.
   Essentiellement, cette idée est la même que le ré-échantillonnage qui choisit les particules selon leur vraisemblance conditionnelle.
- Le MBF est différent de l'APF. Le MBF sera performant quand le bruit d'état est fort tandis que l'APF sera utilisé dans le cas d'un bruit d'état faible.
- Comment choisir M est le point clé du MBF. M est choisi selon le rapport entre la variance du bruit d'état et la variance du bruit d'observation. M augmente lorsque ce rapport augmente.

#### MBF: Simulation 1

$$x_{t} = 1 + \sin(w\pi(t-1)) + \frac{x_{t-1}}{2} + u_{t},$$

$$y_{t} = \begin{cases} \frac{x_{t}^{3}}{5} + v_{t}, & t \leq 30, \\ \frac{x_{t}}{2} - 2 + v_{t}, & t > 30, \end{cases}$$

w = 4e - 2,  $u_t \sim N(0, 100)$  et  $v_t \sim N(0, 5)$ .

Pour mesurer la performance de l'estimation des états  $x_t$  pour  $t=1,\ldots,T$ , nous introduisons la racine de l'erreur quadratique moyenne (RMSE)

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (x_t - \hat{x}_t)^2}.$$

#### MBF: Simulation 1

On avons N=2000 pour chaque filtre et M=3 dans le MBF.

Filtre	Moyenne RMSE	Variance RMSE	Temps
BF	3.60	1.15	1.43
PF-MCMC	3.53	1.11	4.52
PF-EKF	139.23	105812.03	19.28
UPF	3.41	0.68	67.42
APF	4.04	0.86	3.75
MBF	3.23	0.79	4.05

Tab.: Comparaison de méthodes d'estimation utilisant 2000 particules

#### MBF: Simulation 1

Filtre (N. de Particule)	Moyenne RMSE	Variance RMSE	Temps
BF (2000)	3.60	1.15	1.43
BF1 (5000)	3.27	0.82	4.33
PF-MCMC (2000)	3.53	1.11	4.52
UPF (2000)	3.41	0.68	67.42
MBF (2000)	3.23	0.79	4.05

Tab.: Comparaison avec la méthode BF utilisant 5000 particules (M=3 dans le MBF)

# Le filtre bootstrap modifié (MBF) : Simulation 1

Filtre ( $N$ Particules)	Moyenne RMSE	Variance RMSE	Temps
BF2 (8000)	3.18	0.72	8.01
PF-MCMC1 (4000)	3.34	0.84	9.45
MBF1 (2000)	3.15	0.57	7.65

Tab.: Résultats d'estimation avec M=6 dans le MBF

# Le filtre bootstrap modifié (MBF) : Simulation 2

$$x_{t} = \frac{x_{t-1}}{2} + 25 \frac{x_{t-1}}{1 + x_{t-1}^{2}} + 8\cos(1.2t) + u_{t},$$
$$y_{t} = \frac{x_{t}^{3}}{80} + v_{t},$$

où  $u_t \sim N(0,81)$ ,  $v_t \sim N(0,4)$  et  $x_0 \sim N(0,10)$ .

Filtre	Moyenne RMSE	Variance RMSE	Time
BF (3000)	2.26	0.74	2.32
BF2 (7000)	2.20	0.64	6.74
PF-MCMC (3000)	2.25	0.74	7.01
PF-EKF (3000)	23.81	597.00	29.79
UPF (3000)	2.31	0.48	103.81
APF (3000)	2.77	0.46	3.71
MBF (3000)	2.18	0.46	6.33

# Le filtre bootstrap modifié (MBF) : Conclusions

- Lorsque la variance du bruit d'état est grande, et quand le même nombre de particules est utilisé, le MBF surpasse le BF, le PF-MCMC, le PF-EKF et l'APF en termes de moyenne et variance de RMSE, au prix d'un temps de calcul plus élevé.
- Lorsque le même temps de calcul est utilisé, le MBF surpasse le BF, le PF-MCMC, le PF-EKF, l'UPF et l'APF. Plus on utilise de particules dans le BF, meilleure est la performance, mais pour un temps de calcul fixé, le MBF a toujours de meilleures performances.
- ullet En pratique, une valeur de M comprise entre 3 et 10 est suffisante.
- Lorsque la variance du bruit d'état est faible, le filtre MBF fonctionne aussi bien que le BF.

### Le filtre particulaire "unscented" modifié (MUPF) : Motivation

Pour exploiter l'information contenue dans l'observation  $y_t$ , nous pouvons utiliser itérativement l'UKF pour produire la densité d'importance dans le filtre particulaire. Nous développons ici un modèle auxiliaire simple permettant de suivre l'état en utilisant l'UKF. Lorsque la variance du bruit de l'observation est faible, l'observation  $y_t$  est très informative et nous pouvons l'utiliser pour affiner l'estimation de l'état au moyen de l'UKF.

#### MUPF: Modèle auxiliaire

#### Le modèle principal :

$$x_t = f(x_{t-1}) + u_t,$$
  
$$y_t = h(x_t) + v_t.$$

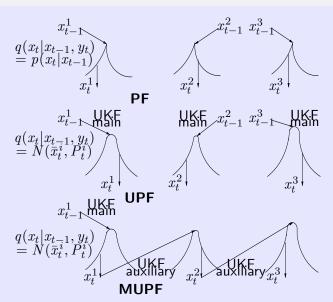
Le **modèle auxiliaire** est conçu pour suivre l'état caché  $r_t$  observées par  $z_t$  :

$$r_t = r_{t-1} + m_t,$$
  
$$z_t = h(r_t) + n_t.$$

### MUPF: Implémentation

- Produire la densité d'importance
  - pour i = 1, la méthode est identique à celle du UPF,
  - pour i=2...N, produire  $N(\bar{x_t^i},P_t^i)$  de  $x_t^{i-1}$  avec modèle auxiliaire par UKF.
- Simuler des particules selon la densité d'importance.
- Calculer les poids pour chaque particule.
- Normaliser les poids d'importance.
- Ré-échantillonner.

### MUPF : Schéma d'échantillonnage d'importance



### MUPF: Remarques

- Dans le PF, le particule  $x_t^i \sim q(x_t|x_{t-1}^i,y_t)$ ; dans le MUPF, la première particule  $x_t^1 \sim q(x_t|x_{t-1}^1,y_t)$ , les autres particules  $x_t^i \sim q(x_t|x_t^{i-1},y_t)$  pour i>1. La densité d'importance de  $x_t^i$  dépend de l'échantillon précédente  $x_t^{i-1}$  (pour i>1).
- Comme dans le PF, les particules échantillonnées selon ces densités d'importance sont indépendantes.
- L'UKF utilisé dans cet algorithme peut être remplacé par autres filtres de quasi-Monte Carlo.
- La complexité de calcul du MUPF est la même que celle de l'UPF.

#### MUPF: Le modèle de simulation

Nous considérons le modèle utilisé par Van der Merwe et el. (2000),

$$x_{t} = 1 + \sin(4\pi 10^{-2}(t-1)) + \frac{x_{t-1}}{2} + u_{t},$$

$$y_{t} = \begin{cases} \frac{x_{t}^{2}}{5} + v_{t}, & t \leq 30, \\ \frac{x_{t}}{2} - 2 + v_{t}, & t > 30, \end{cases}$$

où 
$$u_t \sim \Gamma(3,2)$$
 et  $v_t \sim N(0,10^{-4})$ .

### MUPF: Simulations en utilisant 200 particules

Condition initiale  $x_1 \sim N(0, \sigma_1^2)$ , où  $\sigma_1^2 = 1e - 5$ .

Filtre	Moyenne RMSE	Variance RMSE	Temps
PF	0.27	0.06	0.85
PF-MCMC	0.27	0.06	1.95
PF-EKF	0.43	0.02	2.24
UPF	0.22	0.01	6.92
MUPF	0.02	0.000003	6.97

Tab.: Comparaison des résultats d'estimation avec  $\sigma_1^2=1e-5$  et N=200 particules

### MUPF: Simulations en utilisant 20 particules

Filtre	Moyenne RMSE	Variance RMSE	Temps
PF	0.62	0.05	0.09
PF-MCMC	0.61	0.04	0.19
PF-EKF	0.40	0.02	0.22
UPF	0.26	0.01	0.69
MUPF	0.02	0.000003	0.69

Tab.: Comparaison des résultats d'estimation avec  $\sigma_1^2=1e-5$  et  ${\cal N}=20$  particules

### MUPF: Simulation en utilisant 2000 particules

Filtre	Moyenne RMSE	Variance RMSE	Temps
PF	0.08	0.03	8.30
PF-MCMC	0.10	0.04	18.96
PF-EKF	0.41	0.03	22.50
UPF	0.22	0.02	68.85
MUPF	0.01	0.000004	68.05

Tab.: Comparaison des résultats d'estimation avec  $\sigma_1^2=1e-5$  et N=2000 particules

#### MUPF: Influence de la condition initiale

On suppose maintenant que  $x_1 \sim N(0,\sigma_2^2)$ , où  $\sigma_2^2 = 1$ .

Filtre	Moyenne RMSE	Variance RMSE	Time
PF	1.10	1.56	0.81
PF-MCMC	1.12	1.59	1.85
PF-EKF	0.90	0.82	2.19
UPF	6.09	296.83	6.90
MUPF	0.46	0.54	6.68

Tab.: Comparaison des résultats d'estimation avec  $\sigma_2^2=1$  et N=200 particules

#### Conclusions

- Jusqu'à maintenant, le filtre particulaire classique est la solution la plus simple et souvent la meilleure parmi tous les autre filtres non linéaires.
- Quand la variance de bruit d'état est grande, le MBF surpasse les filtres particulaires classiques.
- Quand la variance de bruit d'observation est faible, le MUPF offre une solution plus performante que les filtres non linéaires testés.

### Perspectives

- Développer une méthode pour choisir M adaptativement.
- Développer des filtres non linéaires en utilisant la théorie de l'information (filtres basés sur l'entropie).

#### Merci

Merci!!

