



UNIVERSITÉ D'ANTANANARIVO

ÉCOLE SUPÉRIEURE POLYTECHNIQUE

ÉCOLE DOCTORALE EN SCIENCES ET TECHNIQUES
DE L'INGÉNIERIE ET DE L'INNOVATION

Équipe d'Accueil doctorale : SCIENCES COGNITIVES et APPLICATIONS

THÈSE de DOCTORAT

Pour l'obtention du titre de

DOCTEUR DE L'UNIVERSITÉ D'ANTANANARIVO

Spécialité : Sciences Cognitives

Par :ROBINSON Matio Hobihery

Titre

**ESTIMATION ET STABILITÉ EN FILTRAGE NON
LINÉAIRE : APPROXIMATIONS PARTICULAIRES**

Soutenue le 16 Décembre 2014 devant la commission de jury composée de

Examineurs :

Philippe ANDRIANARY,	Professeur Titulaire
Paul Auguste RANDRIAMITANTSOA,	Professeur Titulaire
Adolphe RATIARISON,	Professeur Titulaire
Elisée. RASTEFANO,	Professeur
Hery Zo ANDRIAMANOHISOA,	Professeur

Co-directeur de thèse

Yvon ANDRIANA HARISON, Professeur Titulaire

Directeur de thèse

Falimanana RADIMBINDRAINIBE, Professeur

Remerciements

Mes premiers remerciements reviennent à Philippe ANDRIANARY, Directeur de l'ESPA, pour m'avoir accueilli à son établissement. Mon intégration dans cet établissement me permettrait de continuer dans la voie des recherches.

Je remercie également, Paul Auguste RANDRIAMITANTSOA, Directeur de l'Ecole Doctorale Sciences et Techniques de l'Ingénierie et de l'Information, de m'avoir accepté dans son école. La continuité dans une bonne collaboration nous fait avancer les recherches.

Des remerciements particuliers sont destinés à Hery Zo ANDRIAMANHOA, Responsable de l'EAD-SCA, d'avoir cru à mes compétences et m'a pris dans l'équipe de Sciences cognitives et applications.

Mes Directeurs de thèse, Yvon ANDRIANAHARISON et Falimanana RANDIMBINDRAINIBE sont des professeurs disposés à m'écouter et à offrir leurs contributions; Ils m'ont accueilli dans leur laboratoire d'Électricité, Signaux et Automatique (LESA) et je leur exprime ma profonde gratitude.

Je remercie tous les membres de jury qui ont accepté de valider cette thèse. Leurs intérêts sur cette soutenance m'ont fait penser à les donner une sincère reconnaissance. Leurs critiques et remarques ne seraient que bénéfiques pour les recherches.

Je tiens à remercier Adolphe RATIARISON, professeur titulaire à la faculté des Sciences, d'avoir accepté d'être rapporteur externe de mes travaux de thèse. Son appréciation et ses remarques me permettent d'avancer davantage.

Mes vives reconnaissances à Etienne PARDOUX, Professeur à l'Université de Provence qui m'a initié à la résolution de problème de filtrage. Ses travaux sur les filtres non linéaires et équations différentielles stochastiques rétrogrades m'ont inspiré et m'ont servi d'articles de référence.

Merci infiniment à Fabien CAMPILLO, Directeur de Recherches à l'INRIA Montpellier qui m'a encadré en DEA et m'a initié dans les recherches. Son enthousiasme sur l'écologie m'a donné une source d'inspiration dans les applications des mathématiques dans divers domaines.

Et merci à ceux qui m'ont cru et m'ont encouragé dans mes travaux de recherche. Le soutien de ma famille et de mes amis m'a beaucoup marqué.

*À Nivo Hary Dorette,
à Mialitiana et Holitiana
à toute ma famille et
à tous mes amis...*

Notations et symboles

Abreviations

$v.a$: variable aléatoire.
$i.i.d$: indépendantes et identiquement distribuées.
δ_a	: mesure de Dirac au point a .
$\text{Arg max} f(x)$: argument du maximum de la fonction f .
$\lim \sup f(x)$: limite supérieure de f .

Ensembles

$(\Omega, \mathcal{A}, \mathbb{P})$: espace probabilisé, Ω univers, \mathcal{A} tribu sur Ω .
$\mathbb{P}(A)$: probabilité de l'événement $A \in \mathcal{A}$.
$\mathcal{B}(\mathbb{R}^d)$: tribu borélienne de \mathbb{R}^d .
I_n	: matrice unité d'ordre n .
A^t	: la transposée de la matrice A
$\mathcal{C}_b(\mathbb{R}^d)$: ensemble des fonctions continues et bornées de \mathbb{R}^d dans \mathbb{R} .
$\mathcal{B}(a, r) \subset \mathbb{R}^d$: la boule ouverte de centre a et de rayon r .
\mathcal{F}^X	$:= \{\mathcal{F}_n^X\} = \{\sigma(X_k, 0 \leq k \leq n)$ filtration naturelle de X .
\mathcal{Y}_n	$:= \sigma(Y_k, 0 \leq k \leq n)$ tribu des observations antérieures à l'instant n .

Probabilités

$\mathbb{E}(X)$: espérance mathématique de la variable aléatoire X .
$\text{var}(X)$: variance de la variable aléatoire X .
$\text{cov}(X, Y)$: covariance des variables aléatoires X et Y .
$\mathbb{P}_X(dx)$	$:= \mathbb{P}(X \in dx)$, loi de probabilité de X .
$\pi_n(x, dx')$	$:= \mathbb{P}(X_{n+1} \in dx' X_n = x)$ noyau de transition de X_{n+1} sachant $X_n = x$.
$\mathcal{N}(\mu, \sigma^2)$,	: loi gaussienne de moyenne μ et de variance σ^2 .
$\mathcal{U}[a; b]$: loi uniforme sur $[a; b]$.

Filtres

- n : instant courant de temps.
 X_n : signal inconnu à estimer.
 Y_n : signal observé.
 V_n : bruit blanc dans l'équation d'observation Y_n .
 W_n : bruit blanc perturbateur dans l'équation d'état.
 s_W : puissance du bruit perturbateur, $s_W = \mathbb{E}W_n^2$.
 s_V : variance du bruit générateur $s_V = \mathbb{E}V_n^2$.
 $\mu_n^-(dx)$: le filtre prédit à l'horizon n .
 $\mu_n(dx)$: le filtre optimal à l'horizon n .
 r_X : fonction de corrélation de X .
 ε_n : erreur d'estimation.
 $Y_{n:0}$: vecteur signal observé de 0 à n .
 $X_{n:0}$: vecteur signal inconnu.
 $r_{0:n}$: vecteur (r_0, \dots, r_n) .
 $r_{1:n}$: vecteur (r_1, \dots, r_n) .
 \tilde{h}_n : vecteur filtre optimal à l'instant n .
 Y' : vecteur transposé (vecteur ligne) de Y .
 R_n^X : matrice de covariance de $X_{n:0}$, $R_n^X = \mathbb{E}[X_{n:0}X_{n:0}']$.
 R_n^Y : matrice de covariance de $Y_{n:0}$, $R_n^Y = \mathbb{E}[Y_{n:0}Y_{n:0}']$.

Table des matières

Remerciements	i
Notations et symboles	v
Table des figures	xii
Liste des tableaux	xiii
Introduction générale	1
Contexte et historique	1
Problématique et motivations	3
Organisation du mémoire	4
I État de l’art des méthodes de filtrages non linéaires	5
Introduction de la première partie	7
1 Méthodes analytiques de filtrages non linéaires	9
1.1 Introduction	9
1.2 Modélisation du problème de filtrage	10
1.2.1 Modélisation de la mesure d’état	10
1.2.2 Modèle de mesure d’observations	11
1.2.3 Modèle à temps discret	12
1.3 Filtrage dans des modèles particuliers	13
1.3.1 Filtrage dans un modèle de diffusion	13
1.3.2 Filtrage dans un modèle général markovien	15
1.3.3 Filtrage avec une loi initiale à densité	17
1.4 Filtrage non linéaire de système paramétré	19
1.4.1 Modélisation en temps discret	19
1.4.2 Modèle d’état linéaire paramétré	20
1.5 Filtrage non linéaire de système contrôlé	22
1.5.1 Modélisation en temps discret	23
1.5.2 Modèle d’état linéaire contrôlé	24
1.6 Conclusion	26
2 Méthodes numériques de filtrages non linéaires	27
2.1 Introduction	27
2.2 Filtre de Kalman	28
2.2.1 Filtre linéaire gaussien	29
2.2.2 Filtre de Kalman étendu	33
2.3 Filtrage particulière	36
2.3.1 Formules de Bayes	36

2.3.2	Echantillonnage d'importance	37
2.3.3	Filtre particulaire ou bootstrap	38
2.3.4	Résolution par filtrage particulaire	40
2.4	Filtrage par noyau de convolution	41
2.4.1	Noyau de convolution	41
2.4.2	Estimation par noyau de convolution	44
2.4.3	Filtrage par noyau de convolution	47
2.4.4	Resolution par filtrage par noyau	49
2.5	Conclusion	51
Conclusion de la première partie		53
II Études de cas de Filtrage non linéaire et Études de l'oubli de la condition initiale		55
Introduction de la deuxième partie		57
3	Filtrage non linéaire en temps discret	59
3.1	Introduction	59
3.2	Filtrage non linéaire d'un modèle à bruits gaussiens	60
3.3	Filtrage particulaire avec de bruit d'observation uniforme	69
3.4	Filtrage par noyau de convolution d'un système commandé	71
3.4.1	Application du filtrage particulaire	72
3.4.2	Noyau de Parzen-Rosenblatt	72
3.4.3	Estimation de la densité conditionnelle	73
3.4.4	Résolution par filtrage particulaire d'un système paramétré	73
3.5	Filtrage de Kalman inodore d'un modèle à bruit d'observation quelconque .	76
3.5.1	Approximation de la loi gaussienne reduite	79
3.5.2	Approximation des paramètres des filtres	79
3.6	Approximation du bruit d'observation par un melange gaussien	81
3.7	Conclusion	82
4	Études de stabilité en filtrage	83
4.1	Introduction	83
4.2	Stabilité du filtrage de chaîne de Markov	84
4.3	Oubli dans le cas de bruits gaussiens	90
4.4	Oubli dans le cas de bruits d'observation uniformes	91
4.5	Stabilité de filtres de Kalman	94
4.6	Stabilité de filtres particuliers	98
4.7	Stabilité du filtrage par noyau de convolution	103
4.8	Conclusion	105
Conclusion de la deuxième partie partie		107
III Applications et simulations		109
Introduction de la troisième partie		111

5	Poursuite de trajectoire	113
5.1	Introduction	113
5.2	Modelisation	114
5.2.1	Modèle d'état	114
5.2.2	Modèle d'observation	117
5.3	Filtrage non linéaire	118
5.3.1	Algorithme	119
5.3.2	Oubli de la condition initiale	122
5.4	Filtrage de Kalman étendu	125
5.4.1	Algorithmes	125
5.4.2	Oubli de la condition initiale	128
5.5	Filtrage de Uhlmann et Julier	130
5.5.1	Modélisation	130
5.5.2	Algorithme	131
5.5.3	Oubli de la condition initiale	132
5.6	Conclusion	134
6	Commande optimale et estimation paramétrique	135
6.1	Introduction	135
6.2	Filtrage particulière	136
6.2.1	Modelisation de système commandé	136
6.2.2	Algorithme	137
6.2.3	Oubli de la condition initiale	139
6.3	Filtrage par noyau de convolution	143
6.3.1	Modelisation de système paramétré	143
6.3.2	Estimation paramétrique	143
6.3.3	Estimation non-paramétrique	144
6.3.4	Estimation par histogramme mobile	144
6.3.5	Estimation par noyau de convolution	146
6.3.6	Estimation des densités conditionnelles	147
6.3.7	Algorithme	150
6.3.8	Oubli de la condition initiale	153
6.4	Conclusion	158
	Conclusion de la troisième partie	159
	Conclusion générale	161
	Résultats	161
	Discussions et perspectives	162
IV	Annexes	163
A	Méthode de Monte-Carlo de chaînes de Markov	165
A.1	Bayes et Simulation	165
A.1.1	Exemple : loi Gamma	165
A.1.2	Modèles de mélange	166
A.1.3	Méthodes bayésiennes	166
A.1.4	Algorithme d'acceptation-rejet	167

A.1.5	Echantillonnage d'importance	168
A.2	Metropolis-Hastings	168
B	Chaînes de Markov cachées	173
B.1	Introduction	173
B.2	Algorithmes	174
B.2.1	Algorithme forward de Baum	174
B.2.2	Algorithme backward de Baum	175
B.2.3	Algorithme de Viterbi	175
B.2.4	Algorithme de Baum-Welch	176
C	Filtrage Statistique en Traitement de Signal	177
C.1	Processus aléatoire	177
C.1.1	Processus stationnaire	177
C.1.2	Bruit blanc	178
C.1.3	Processus stationnaire d'ordre 2	178
C.1.4	Processus à moyenne mobile	179
C.1.5	Processus autoregressif	179
C.1.6	Processus ARMA	180
C.1.7	Filtre de moyenne mobile	180
C.2	Filtrage optimal	181
D	Articles publiés	185
D.1	Oubli de la condition initiale et filtre de Kalman	185
D.2	Filtre particulaire avec un bruit d'observation uniforme	200
D.3	Estimation d'une commande par noyau de convolution	213
Bibliographie		xi
	Fiche de renseignement	xix
	Résumé	xx
	Abstract	xxi

Table des figures

5.1	Trajectoire nominale	114
5.2	Trajectoire réelle	116
5.3	Trajectoire observée	118
5.4	Signaux mesurés et observés	121
5.5	Densité et fonction de répartition de la loi initiale	122
5.6	Densité du filtre prédit et du filtre corrigé	122
5.7	Filtres prédits et corrigés	123
5.8	Conditions initiales différentes	123
5.9	Erreur linéaire à N=10	124
5.10	Filtrage de Kalman	127
5.11	Trajectoires avec $X_0 = (0; 0)$ et $X_0 = (100; 1000)$	128
5.12	Trajectoires avec $\sigma_0 = 1$ et $\sigma_0 = 2$	129
5.13	Approximation linéaire de la densité gaussienne	132
5.14	Trajectoires estimées pour X_0 différentes	133
5.15	Trajectoires estimées pour Q_0 différentes	134
6.1	Filtre particulaire	138
6.2	Mutation des particules	138
6.3	Selection des particules	139
6.4	Particules de la loi initiale	139
6.5	Trajectoire nominale et observée	140
6.6	Évolution des particules avec $\hat{X}_0 = (10; 90)$ et $s_0 = 0.8$	140
6.7	Évolution des particules avec $\hat{X}_0 = (10; 110)$ et $s_0 = 0.8$	141
6.8	Évolution des particules avec $\hat{X}_0 = (10; 110)$ et $s_0 = 2$	142
6.9	Estimations de densités par histogramme	144
6.10	Estimations par histogramme mobile pour h=1	145
6.11	Estimations par histogramme mobile pour h=10	146
6.12	Estimations par noyau pour h=0.3 et h=0.8	146
6.13	Estimations par noyau pour h=1	147
6.14	Estimations par noyau de la densité exponentielle	147
6.15	Estimation par noyau de convolution gaussien	148
6.16	Estimations par noyau d'Epanechnikov de la densité conditionnelle	149
6.17	Estimations par noyau de Picard de la densité conditionnelle	149
6.18	Distribution des particules initiales	151
6.19	Distribution des particules en mutation	151
6.20	Densités des filtres prédit et corrigé	152
6.21	Echantillonnage du filtre optimal	152
6.22	Densités du filtre et des paramètres	153

6.23	Différences des trajectoires 1 et 2	156
6.24	Différences des trajectoires 1 et 3	157
6.25	Différences des trajectoires 2 et 3	158

Liste des tableaux

5.1	Algorithme filtre optimal	120
5.2	Algorithme de calcul de l'erreur	124
5.3	Algorithme filtre de Kalman	126
5.4	Valeurs observées et estimées avec $X_0 = (0; 0)$ et $X_0 = (100; 1000)$	128
5.5	Valeurs observées et estimées avec $\sigma_0 = 1$ et $\sigma_0 = 2$	129
5.6	Algorithme filtrage de Uhlmann et Julier	131
5.7	Valeurs observées et estimés pour X_0 différentes	133
5.8	Valeurs observées et estimés pour Q_0 différentes	134
6.1	Algorithme filtre particulière	137
6.2	Vitesse et angle pour $\hat{X}_0 = (10; 90)$ et $s_0 = 0.8$	141
6.3	Vitesse et angle pour $\hat{X}_0 = (10; 110)$ et $s_0 = 0.8$	142
6.4	Vitesse et angle pour $\hat{X}_0 = (10; 110)$ et $s_0 = 2$	142
6.5	Algorithme filtrage par noyau de convolution	150
6.6	Estimations des paramètres avec $\sigma_0 = 0.8, \sigma_0 = 1.2, \sigma_0 = 2 \sigma_W = 0.2$ et $\sigma_a = 0.2$	154
6.7	Estimations des paramètres avec $\sigma_0 = 0.8, \sigma_0 = 1.2, \sigma_0 = 2 \sigma_W = 0.1$ et $\sigma_a = 0.1$	155
6.8	Evaluation de l'écart avec $s_a = 0.2, s_W = 0.2$	156
6.9	Evaluation de l'écart avec $s_a = 0.1, s_W = 0.1$	157

Introduction générale

Sommaire

Contexte et historique	1
Problématiques et motivations	3
Organisation du memoire	4

Contexte et historique

Depuis quelques decennies, la plupart des recherches en probabilités appliquées est axée sur les équations différentielles stochastiques, les contrôles optimaux stochastiques ou encore sur les filtrages linéaires et non linéaires de signaux. Parfois, on mélange ces thèmes principaux pour résoudre des problématiques particulières. Entre autres, dans les problèmes financiers, la résolution du contrôle optimal nécessite une résolution d'équations aux dérivées partielles et donc par approximation, une résolution d'équations différentielles stochastique [BENSOUSSAN,1984], [FUJISAKI,1980], [FUJISAKI,1981].

Dans cette thèse, on étudie particulièrement le problème de filtrage d'un signal. Le signal à étudier est un processus non observable. On dispose d'un signal observé mesuré périodiquement par un ou plusieurs capteurs. Dans la modélisation, un système d'équations différentielles stochastiques représente l'équation d'état et l'équation d'observation comme dans [PARDOUX,1981], [PARDOUX,1991].

Étant donnée un processus d'observations $\{Y_t, t \in \mathbb{R}^+\}$, le problème de filtrage consiste à extraire des informations à propos d'un processus d'états $\{X_t, t \in \mathbb{R}^+\}$ cachés sous ja-cents. Il s'agit donc d'extraire des informations utiles d'un ensemble de données entachées de bruits indésirables.

Historiquement à l'origine, les filtres en traitement de signal, reposaient sur des tech-niques fréquentielles et sur la notion de coupure en détruisant à la fois bruit et signal utile. Les filtres de Weiner ont été les premiers travaux dans ce domaine[ROSSI, 2004] ; il s'agit des filtres linéaires qui minimisent une intégrale de l'erreur entre le signal utile et le signal estimé, en 1949.

En 1958, apparaissaient les filtres de Kalman [ALAZARD,2006], [ENZO et al.,2010], [PETIT,2011]. Des filtres basés sur des modèles où tout est gaussien. Ces travaux de Kalman traitaient des processus en temps discret à dynamique linéaire de loi gaussienne entachés de bruits gaussiens. Bucy en 1961, les a développé pour les processus en temps

continu [ROTELLA, 2003]. Dans ces modèles gaussiens, le filtre de Kalman-Bucy était exactement le filtre optimal. Dans la construction, le filtre est récursif et peut être appliqué à des processus non stationnaires.

Stratonovich commençait les travaux sur les filtrages non linéaires dans les années 60 qui étaient complétés par Kallianpur et Strebel [PARDOUX, 1991] en 1967, Zakai en 1969 [ROSSI et VILA, 2003], Fujisaki et al. en 1972 [FUJISAKI, 1980], [FUJISAKI, 1981], Yor en 1977 et encore Pardoux en 1991 [PARDOUX, 1991]. Ils commençaient par des modèles non linéaires gaussiens.

Kallianpur montraient l'existence d'une mesure absolument continue au filtre avec laquelle on construit récursivement le filtre en introduisant la formule de Bayes. Zakai a établi une équation aux dérivées partielles en association au filtre optimal.

Fujisaki travaillaient sur les filtrages en temps continus et contrôles optimaux. Pardoux avec les équations différentielles stochastiques retrogrades résolvait le problème de filtrage avec l'équation de Zakai.

En 1990, Crisan et al. introduisait les filtrages particuliers [CAMPILLO, 2006], [DEL MORAL et al., 2006], [DUFLOS, 2011] pour les processus non-linéaires à bruits quelconques. Del Moral et al [DEL MORAL et al., 2006] développaient ces filtrages particuliers. Campillo [CAMPILLO, 1986] dans sa thèse, résolvait le filtrage non linéaire par approximation de Gauss Karlekin dans la résolution numérique des équations aux dérivées partielles.

En théorie une solution existe analytiquement mais n'est pas très explicitée pour pouvoir l'exploiter convenablement. Avec l'avancée technologique en informatique, la résolution approchée par des méthodes de simulation, entre autres les méthodes de Monte Carlo [LAPEYRE, 1993], [TOURNERET, 2007], [DEL MORAL, 2008] joue un rôle principal dans les validations des théories.

Le filtrage se présente sous plusieurs aspects :

- En traitement de signal [ANNEXE C] [JUTTEN, 2010], la transmission d'informations n'est jamais parfaite. Le filtre est déterminé par une fonction, en général, linéaire qui, appliquée aux observations, donne une estimation du signal voulu.
- En mécanique classique, dans la poursuite d'une trajectoire [BREHARD, 2006], [GUECHI, 2010], le filtrage est appliqué pour contrôler la trajectoire d'un mobile dans son évolution. Avec du bruit d'observation, le filtre est déterminé par une loi de distribution à chaque instant de l'estimateur de la position du mobile. Cette loi peut être caractérisée par ses paramètres statistiques, la moyenne et la variance, comme dans le cas des bruits suivant une loi gaussienne, ou par une fonction densité dans le cas plus général.

Le filtre obtenu est effectivement estimé compte tenu des perturbations dans les mesures. On a donc intérêt à considérer un filtre optimal dans le sens où l'erreur quadratique moyenne entre sa loi exacte et sa loi approchée soit le moindre possible : c'est le principe

de MMSE¹ [Le GLAND, 2011].

L'erreur d'estimation est évaluée par sa variance ; le critère d'optimalité du filtre est d'avoir la variance minimale.

Le filtrage présente plusieurs formes comme linéaire ou non, à bruits avec une loi de distribution connue particulièrement gaussienne ou non, avec un système dynamique continu ou non et bien d'autres. En tout cas, même si le système évolue en temps continu, le filtrage est traité en temps discret.

Problématiques et motivations

Le thème principal de cette thèse est le filtrage non linéaire de signaux. Des développements du point de vue théorique et algorithmique de filtrage sont exposés. Les points essentiels étudiés sont :

- le problème d'existence et de construction de solution théorique du filtre optimal. Il s'agit de construire, tout en restant dans le cas gaussien, le filtre optimal à partir du théorème de Girsanov.
- l'étude d'une propriété fondamentale du filtre optimal : l'oubli de la condition initiale du filtre. Cet oubli de la condition initiale permet d'apprécier les méthodes de filtrage établies, en particulier pour le filtrage particulaire.
- le problème de l'approximation particulaire appliquée au bruit d'observation quelconque, en particulier bruit uniforme. Le bruit uniforme est un bruit non informatif. L'application dans les filtres particuliers avec les bruits uniformes a une propriété très particulière.
- le problème de mise en œuvre des filtres prédits et corrigés dans le filtrage bayésien. L'oubli de la condition initiale est justifié par application de l'approximation de Monte Carlo dans la simulation.
- le problème de régularisation des filtres pour déterminer l'estimation des paramètres ou des commandes. La détermination d'une commande intervenant dans le système demande des hypothèses plus précises.
- le problème d'application du filtrage de Julier et Uhlmann au modèle à bruits d'observation quelconque. Le bruit d'observation est approximé par un mélange de bruits gaussiens.

Cette thèse traite des problèmes particuliers dans le filtrage et complète les théories et applications sur les filtres non linéaires.

1. minimum mean square error

Organisation du memoire

Le mémoire est organisé en quatre parties :

Un état de l'art sur les méthodes du filtrage non linéaire est exposé à la première partie.

Le premier chapitre est consacré aux méthodes analytiques du problème de filtrage. En particulier, le filtrage de système paramétré et le système de filtrage de système commandé sont introduits.

Le second chapitre concerne les différentes méthodes numériques classiques de filtrage non linéaire. Nous rappelons en particulier les filtrages de Kalman et ses variantes, les filtrages particuliers et ses variantes en particulier le filtrage par noyau de convolution.

La deuxième partie développe le thème principal de la thèse. Des cas de filtrages non linéaires particuliers sont étudiés. Des hypothèses particulières sont avancées et des solutions sont proposées.

Le troisième chapitre développe une théorie sur l'existence de solution optimale dans le cas gaussien et étudie le cas de modèle à bruit d'observation uniforme.

Dans la pratique, les filtres de Kalman et les filtres particuliers sont adaptés pour le filtrage non linéaires. La perturbation par des bruits blancs uniformes des observations est un phénomène particulier et son adaptation est assez simple. Le filtrage par noyau de convolution améliore les approximations particulières.

Le quatrième chapitre consiste à étudier l'oubli de la condition initiale du filtre. Il s'agit de mesurer la différence entre deux filtres déduits de deux différents états initiaux, à un horizon assez lointain. Dans ce mémoire, cet oubli peut se traduire par la mesure dans L^1 de cette différence ou la mesure de la variation totale. Nous allons montrer l'oubli de la condition initiale dans les différentes méthodes de filtrage.

La troisième partie est consacrée aux applications et simulations. Nous étudions des applications comme la poursuite de la trajectoire, la commande optimale d'un engin volant ou encore l'estimation paramétrique. L'oubli de la condition initiale est appliquée dans les simulations.

Les théories sont vérifiées par des applications sur des phénomènes classiques de filtrage. La discretisation dans l'implémentation entraîne des situations compliquées. Des procédures dans Scilab 5.4.1, dans R et dans Maple sont utilisés pour les simulations des trajectoires des filtres optimaux et leurs stabilités.

La quatrième partie concerne les annexes complétant des définitions et théories non développées dans les précédentes parties. Il s'agit

- Des méthodes de Monte-Carlo de chaînes de Markov,
- Des chaînes de Markov cachés,
- Des filtrages statistiques en traitement de signal,
- Des articles publiés dans MADA-Enelsa.

Première partie

État de l'art des méthodes de filtrages non linéaires

Introduction de la première partie

Le filtrage consiste à estimer l'état d'un système dynamique à partir d'observations partielles bruitées. En l'absence de bruit sur l'équation d'observation, on peut reconstruire directement toute la trajectoire de X_t avec des hypothèses adéquates.

En présence de bruit, on traite le problème de filtrage par une reconstitution de la trajectoire de X_t sachant les observations qu'on dispose ou par une détermination de la loi de trajectoire de X_t sachant la trajectoire des processus d'observations depuis l'instant initiale.

En général, les signaux sont analogiques. Un traitement numérique au niveau des capteurs est nécessaire dans l'étude. La discretisation du signal observé se fait dans un espace de dimension au plus à celle du signal non observé et la non linéarité du modèle d'observation ne permet pas d'avoir l'estimation par inversion. Des informations à priori sur le signal caché assurent l'existence et l'unicité du filtre. Un critère de minimisation est introduite [Le GLAND, 2011] et la propriété markovienne du signal caché est indispensable.

Des extensions du filtrage peuvent être étudiées selon le cas :

- le lissage consistant à extraire de l'information à l'instant n , à partir des données observées au-delà de l'instant n
- la prediction consistant à prédire la valeur d'une quantité d'intérêt à l'instant $n + \tau$, $\tau \geq 0$, connaissant cette quantité jusqu'à l'instant n inclus.

On dispose donc,

- d'une part, une équation d'état d'un processus X_n non observé,
- d'autre part, une équation d'observation où à chaque instant n , on recueille une observation Y_n représentée par une fonction $h(X_n)$ de l'état X_n et d'un bruit additif :

$$Y_n = h(X_n) + V_n \quad (1)$$

On suppose également que les caractéristiques statistiques du bruit de mesure V_n sont connues.

Le but est d'obtenir le plus d'informations possibles sur l'état du système X_n soit :

- en utilisant le filtrage de Kalman pour déterminer un estimateur \hat{X}_n de X_n à chaque instant n ,
- en utilisant le filtrage particulaire pour déterminer la loi conditionnelle de X_n sachant les observations Y_1, \dots, Y_n .

La mesure de l'état d'un système est en général impossible, pour différentes raisons entre autres, l'absence de matériels adaptés aux mesures. On fait alors des mesures observables à partir des capteurs qui sont évidemment liées à l'état du système non observé au cours du temps.

Le signal obtenu est en général perturbé par d'autres signaux d'origines inconnues supposés aléatoires.

Le filtrage est une opération qui consiste à estimer l'état d'un système dynamique à partir d'observations partielles et bruitées : on cherche à chaque instant n à estimer la vraie valeur de X_n , à partir des observations Y_1, \dots, Y_n .

L'estimation optimale de X_n est naturellement $\mathbb{E}(X_n|\mathcal{Y}_n)$. Le filtre optimal est défini par la loi de probabilité de cette estimation.

On établit une relation de récurrence pour le filtre avec un algorithme qui permet le traitement dans l'immédiat à chaque mesure observée. On décompose le traitement en deux étapes :

- une étape de prédiction où la loi prédite est :

$$\mu_n^-(dx) = \mathbb{P}(dx \in X_n | \mathcal{Y}_{n-1}) = \mathbb{P}_{X_n}(dx | \mathcal{Y}_{n-1}),$$

- une étape de correction où la loi corrigée, qui est le filtre optimal, est :

$$\mu_n(dx) = \mathbb{P}(dx \in X_n | \mathcal{Y}_n) = \mathbb{P}_{X_n}(dx | \mathcal{Y}_n)..$$

Ces étapes permettent de construire un algorithme récursif tel qu'à l'instant $n - 1$ avec l'observation Y_n , on peut obtenir $\mu_n(dx)$ sans reconsidérer \mathcal{Y}_{n-1} , c'est à dire que l'algorithme tourne au "temps réel".

Cette décomposition nécessite une structure markovienne du signal X_n non observé.

En général, la théorie sur l'existence du filtre ne permet pas d'avoir une expression analytique du filtre. Dans cette partie, nous allons exposer les méthodes classiques de filtrages non linéaires. Il s'agit des filtrages de Kalman et des filtrages particuliers. Leurs algorithmes sont présentés avec des hypothèses dans leur mise en application.

Un chapitre sur la résolution du filtrage en général montre les constructions des différents modèles en particulier, les modèles non linéaires paramétrés et les modèles non linéaires commandés. Des adaptations des méthodes de filtrage sur les modèles sont développées.

Chapitre 1

Méthodes analytiques de filtrages non linéaires

Sommaire

1.1 Introduction	9
1.2 Modélisation du problème de filtrage	10
1.2.1 Modélisation de la mesure d'état	10
1.2.2 Modèle de mesure d'observation	11
1.2.3 Modèle à temps discret	12
1.3 Filtrage dans des modèles particuliers	13
1.3.1 Filtrage dans un modèle de diffusion	13
1.3.2 Filtrage dans un modèle général markovien	15
1.3.3 Filtrage avec une loi initiale à densité	17
1.4 Filtrage non linéaire de système paramétré	19
1.4.1 Modélisation en temps discret	19
1.4.2 Modèle d'état linéaire paramétré	20
1.5 Filtrage non linéaire de système contrôlé	22
1.5.1 Modélisation en temps discret	23
1.5.2 Modèle d'état linéaire contrôlé	24
1.6 Conclusion	26

1.1 Introduction

Un processus aléatoire X_n en général n'est pas observé d'autant plus qu'il est entaché de bruit aléatoire. Un signal Y_n observable contient X_n sous une forme explicitée par une fonction. Ce signal est aussi perturbé par un bruit. L'idée du filtrage est de déterminer la loi de X_n conditionnée par $\mathcal{Y}_n = \sigma(Y_1; \dots, Y_n)$, la tribu des observations.

On dispose des observations Y_1, Y_2, \dots à des instants t_1, t_2, \dots .
On veut suivre les évolutions d'un état caché $X_k = X(t_k)$.

Dans la construction du modèle on a :

- un processus d'état X_k , avec un état initial X_0 défini par une loi initiale.
- une dynamique sur l'état X_k définie par une loi de transition.

- une relation état-observation définie par une loi de proposition

Le problème de filtrage est modélisé par un système d'équation d'état et d'équation d'observation.

La résolution du problème de filtrage se déroule en deux étapes :

Une étape de prédiction où on estime X_n à partir des observations Y_1, \dots, Y_{n-1} et on a la loi prédite par

$$\mu_n^-(dx) = \mathbb{P}(X_n \in dx | \mathcal{Y}_{n-1}) = \mathbb{P}_{X_n}(dx | \mathcal{Y}_{n-1}).$$

Une étape de correction où on corrige la loi prédite au vu de l'observation y_n et la loi corrigée est :

$$\mu_n(dx) = \mathbb{P}(X_n \in dx | \mathcal{Y}_n) = \mathbb{P}_{X_n}(dx | \mathcal{Y}_n).$$

Pour que le filtre optimal soit récursif, il faut et il suffit que le système soit markovien conditionnellement à l'observation [BERNIER et al., 1994].

1.2 Modélisation du problème de filtrage

1.2.1 Modélisation de la mesure d'état

En considérant $X_0 \in \mathbb{R}^d$ l'état initial, $h : \mathbb{R}^d \rightarrow \mathbb{R}^d$ un champ vectoriel. La dynamique de la variable d'état est définie par une équation différentielle :

$$\frac{dX}{dt}(t) = f(X(t)), \quad X(0) = x_0. \quad (1.1)$$

La solution, si elle est unique, est représentée par une trajectoire : $X : \mathbb{R}^+ \rightarrow \mathbb{R}^d$. Les trajectoires mesurées ne correspondent pas vraiment aux solutions analytiques de l'équation. Des effets aléatoires viennent se superposer à la trajectoire. On introduit un processus aléatoire perturbant le système (1.1).

$$\frac{dX}{dt}(t) = f(X(t)) + \sigma(X(t))\xi(t). \quad (1.2)$$

où $\xi(t)$ est un bruit blanc¹, dérivée temporelle du processus de Wiener.

$$\frac{dW}{dt}(t) = \xi(t)$$

On obtient une équation différentielle stochastique appelée **équation de Fokker-Planck**.

$$dX(t) = f(X(t))dt + \sigma(X(t))dW(t) \quad (1.3)$$

Sous certaines hypothèses, cette équation admet une solution unique au sens faible [PARDOUX, 1991] exprimée sous la forme d'intégrale stochastique :

$$X_t = x_0 + \int_0^t f(X_s)ds + \int_0^t \sigma(X_s)dW_s \quad (1.4)$$

En général, l'équation d'état est de la forme :

$$\frac{dX}{dt}(t) = f(t, X(t), W(t)) \text{ où } W_t \text{ est un bruit blanc.}$$

1. voir ANNEXE C, variable aléatoire d'espérance nulle

La fonction f n'est pas en général connue mais le processus $X(t)$ a toujours une structure markovienne. Cette propriété est essentielle dans le traitement récursif du filtrage.

Dans la suite, on admet que le bruit blanc est additif. Il est issu de perturbations aléatoires dans l'évolution de $X(t)$ et sa matrice de covariance est connue.

1.2.2 Modèle de mesure d'observations

Soit X_t un signal markovien, que l'on ne peut pas observer directement. On observe le processus Y_t perturbé par un bruit blanc d'observation η_t supposé être la dérivée d'un processus de Wiener V_t dans \mathbb{R}^k , éventuellement corrélé avec X_t .

On suppose que : H une fonction mesurable et bornée de $\mathbb{R}^+ \times \mathbb{R}^d$ dans \mathbb{R}^k et τ une fonction continue et bornée dans \mathbb{R}^k .

En général, le processus d'observation Y_t évolue dans un espace de dimension inférieure à celle de l'espace d'état : $k \leq d$.

$$Y_t = \int_0^t H(s, X_s) ds + \tau(X_t) \eta(t) \quad (1.5)$$

$$Y(0) = y_0, \text{ nulle en général}$$

Et le processus de Wiener est tel que :

$$\frac{dV}{dt}(t) = \eta(t)$$

Sous forme d'équation différentielle stochastique, on a :

$$dY_t = H(t, X_t) dt + \tau(X_t) dV_t \quad (1.6)$$

Plus précisément sous forme d'intégrales stochastique :

$$Y_t = y_0 + \int_0^t H(s, X_s) ds + \int_0^t \tau(X_s) dV_s \quad (1.7)$$

On peut poser $h(t) = \int_0^t H(s, X_s) ds$ et supposer que $V(t)$ un bruit blanc additif indépendant de X_t causé par les perturbations sur les mesures d'observation.

L'équation d'observation est de la forme :

$$Y_t = h(t, X_t) + V(t)$$

1.2.3 Modèle à temps discret

Des équations (1.3) et (1.7), on obtient le système d'équations différentielles stochastiques [PARDOUX,1981]

$$\begin{cases} dX_t &= f(X_t)dt + \sigma(X_t)dW_t \\ dY_t &= h(t, X_t)dt + \tau(X_t)dV_t \end{cases}$$

Le filtrage consiste à estimer de façon recursive l'état caché $\{X_t, t \geq 0\}$ au vu d'observations $\{Y_t, t \geq 0\}$.

Dans la suite, les mesures sont réalisées périodiquement et le système d'équations différentielles est approximé par un système d'équations recurrentes.

Plus précisément, on a le système :

$$\begin{cases} X_{n+1} &= f(X_n) + \sigma(X_n)W_n \\ Y_n &= h(X_n) + \tau(X_n)V_n \end{cases}$$

De plus, on fait l'hypothèse d'additivité des bruits blancs gaussiens i.i.d :

$$\begin{cases} X_{n+1} &= f(X_n)dt + \sigma_W \cdot W_n \text{ où } W_n \sim \mathcal{N}(0, I_d) \\ Y_n &= h(X_n)dt + \sigma_V \cdot V_n \text{ où } V_n \sim \mathcal{N}(0, I_k) \end{cases}$$

Plus simplement, l'équation d'état est

$$X_{n+1} = f(X_n) + W_n, \quad W_n \sim \mathcal{N}(0, \sigma_W I_d), \quad n \in \mathbb{N}. \quad (1.8)$$

W_n est indépendant de la chaîne $\{X_n, n \in \mathbb{N}\}$ et f est une fonction non nécessairement linéaire.

Comme f est mesurable, la chaîne est, par définition, markovienne.

Plus précisément, l'état du système non observé $X = \{X_n; n \geq 0\}$ est une chaîne de Markov à valeurs dans l'espace mesurable $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est de loi initiale μ_0 et de probabilité de transition π_n ,

$$\mathbb{P}(X_{n+1} \in dx | X_n = x') = \pi_n(x'; dx)$$

où pour toute fonction test mesurable bornée $g : \mathbb{R}^d \rightarrow \mathbb{R}^+$

$$\mathbb{E}[g(X_{n+1}) | X_n = y] = \int_{\mathbb{R}^d} g(x) \pi_n(y; dx)$$

$$\mathbb{E}[g(X_0)] = \int_{\mathbb{R}^d} g(x) \mu_0(dx)$$

Dans la pratique, la chaîne de Markov est considérée homogène, c'est-à-dire que $\pi_n = \pi$, la loi de transition ne dépend pas de n .

À chaque instant n , on veut calculer la loi conditionnelle de X_n , connaissant Y_k , $0 \leq k \leq n$.

$$\mu_n(dx) = \mathbb{P}(X_n \in dx | \mathcal{Y}_n)$$

\mathcal{Y}_n est la filtration des observations Y_k , $0 \leq k \leq n$.

On peut supposer que cette loi conditionnelle admet une densité par rapport à la mesure de Lebesgue dans $\mathcal{B}(\mathbb{R}^d)$.

De même, on admet comme équation d'observation

$$Y_n = h(X_n) + V_n, \quad n \geq 1 \quad (1.9)$$

1.3 Filtrage dans des modèles particuliers

1.3.1 Filtrage dans un modèle de diffusion

La variable d'état X_t prend ses valeurs dans \mathbb{R}^d et générée par l'équation différentielle stochastique :

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \quad (1.10)$$

W_t est un bruit blanc gaussien et X_t n'est observée qu'à travers un processus d'observation tel que :

$$Y_t = h_t(X_t) + V_t$$

V_t est un bruit blanc supposé i.i.d de W_t et h_t est une fonction mesurable.

L'équation différentielle stochastique (1.10) est mise sous la forme intégrale stochastique :

$$X_t = Z + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s. \quad (1.11)$$

ou chaque composante est

$$X_t^i = Z^i + \int_0^t b^i(s, X_s)ds + \sum_{j=1}^d \int_0^t \sigma_{i,j}(s, X_s)dW_s^j.$$

L'existence et l'unicité de la solution² pour l'équation (1.11) sont conditionnées par : b et σ sont des fonctions continues et il existe $K < +\infty$ tel que :

1. $|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq K|x - y|$
2. $|b(t, x)| + |\sigma(t, x)| \leq K(1 + |x|)$
3. $\mathbb{E}(Z^2) < +\infty$

De plus cette solution vérifie, pour tout $T \leq 0$, dans l'intervalle $[0, T]$

$$\mathbb{E} \left[\sup_{0 \leq s \leq T} |X_s|^2 \right] < +\infty$$

Comme cette solution n'est pas observable, la meilleure estimée de X_t sachant $\mathcal{Y}_t = \{Y_s; s \leq t\}$ est la moyenne conditionnelle $\mathbb{E}[X_t|\mathcal{Y}_t]$ et le filtre optimal est défini par sa mesure de probabilité telle que :

pour une fonction test mesurable bornée ϕ , on a :

$$\int_{\mathbb{R}^d} \phi(x) \mu_t(dx) = \mathbb{E}[\phi(X_t)|\mathcal{Y}_t]$$

2. La solution est appelée diffusion

de densité telle que

$$p_t(x)dx = \mu_t(dx)$$

Dans la discrétisation, on donne une approximation du filtre dans l'espace de mesure de probabilité.

Le lemme suivant montre l'association de l'équation d'état (1.10) avec un opérateur aux dérivées partielles,

Lemme 1 (formule d'Îto) *f est une fonction deux fois continûment différentiable et*

$$X(t) = X(0) + \int_0^t b(s, X(s))ds + \int_0^t \sigma(s, X(s))dW(s)$$

où W_t est un mouvement brownien standard. Alors

$$\begin{aligned} df(t, X(t)) &= \left[\frac{\partial f}{\partial t}(t, X(t)) + b(t, X(t)) \cdot \frac{\partial f}{\partial x}(t, X(t)) + \frac{1}{2} \sigma(t, X(t))^2 \frac{\partial^2 f}{\partial x^2}(t, X(t)) \right] dt \\ &\quad + \sigma(t, X(t)) \frac{\partial f}{\partial x}(t, X(t)) \cdot dW(t) \end{aligned}$$

Dans le cas multidimensionnel

$$\begin{aligned} f(t, X(t)) &= f(0, X(0)) + \int_0^t \left[\frac{\partial f}{\partial s}(s, X(s)) + \sum_{i=1}^d b_s^i(s, X(s)) \cdot \frac{\partial f}{\partial x_i}(s, X(s)) \right. \\ &\quad \left. + \frac{1}{2} \sum_{i,j=1}^d [\sigma(s, X(s))\sigma(s, X(s))^t]_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(s, X(s)) \right] ds \\ &\quad + \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x_i}(s, X(s)) \cdot \sigma(s, X(s)) dW_s. \end{aligned}$$

On peut donc associer à l'équation (1.10), un opérateur aux dérivées partielles du second ordre, appelé générateur infinitésimal [BERNIER et al.,1994] du processus de diffusion X_t :

$$L_t = \frac{1}{2} \sum_{i,j=1}^d a_t^{i,j} \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^d b_t^i \frac{\partial}{\partial x_i} \text{ avec } a_t = \sigma(t, \cdot) \cdot \sigma(t, \cdot)^t. \quad (1.12)$$

La solution du problème de filtrage non linéaire est alors donnée par la solution de l'équation aux dérivées partielles stochastique :

$$dq_t(x) = L_t^t q_t(x)dt + q_t(x)h_t^t dY_t \quad (1.13)$$

En général, la solution de cette équation différentielle est obtenue numériquement. Si q_t est solution de (1.13), la densité du filtre non normalisée est

$$p_t(x) = \frac{q_t(x)}{\int_{\mathbb{R}^d} q_t(x')dx'}$$

Plus précisément, pour la prédiction, de l'équation de Fokker-Planck

$$\frac{\partial p_t(x)}{\partial t} = L_t^* p_t(x)$$

, on obtient la densité prédite p_n^-

Pour la correction, on obtient la densité par normalisation

$$p_n(x) = c_n \psi_n(x) p_n^-$$

où c_n est une constante de normalisation et

$$\psi_n(x) = \exp \left[-\frac{1}{2\Delta} |y_n - h_n(x)|^2 \right]$$

la fonction de vraisemblance de l'estimation de X_n sachant \mathcal{Y}_n .

1.3.2 Filtrage dans un modèle général markovien

Considérons le système tel que l'état non observé $X = \{X_n; n \geq 0\}$ est une chaîne de Markov homogène à valeurs dans l'espace mesurable $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est de loi initiale μ_0 et de probabilité de transition π où pour toute fonction mesurable $f : \mathbb{R}^d \rightarrow \mathbb{R}^+$

$$\mathbb{E}[f(X_{n+1}) | X_n = y] = \int_{\mathbb{R}^d} f(x) \pi(y; dx) \quad (1.14)$$

La loi initiale est définie par :

$$\mathbb{E}[f(X_0)] = \int_{\mathbb{R}^d} f(x) \mu_0(dx) \quad (1.15)$$

L'observation est un processus tel que :

$$Y = \{Y_n; n \geq 0\}, Y_n = h(X_n) + V_n, n \geq 0.$$

h est une fonction mesurable de \mathbb{R}^d dans \mathbb{R}^k et $V = (V_n; n \geq 0)$ est une suite de variables i.i.d de loi commune $\nu(dx)$, le bruit blanc de l'observation non nécessairement gaussien.

Supposons que la loi commune est à densité $f_V : \nu(dx) = f_V(x)dx$.

La loi prédite et la loi corrigée sont déterminées par l'application de la formule de Bayes. Avant cette application, introduisons les lemmes suivants :

Lemme 2 *La loi de Y_n sachant \mathcal{Y}_{n-1} est donnée par :*

$$\mathbb{P}_{Y_n}(dy | \mathcal{Y}_{n-1}) = \left[\int_{\mathbb{R}^d} f_V(y - h(x)) \mu_n^-(dx) \right] dy$$

Preuve

En effet, si ϕ est une fonction test mesurable bornée, et comme V_n est indépendante de (X_n, \mathcal{Y}_{n-1})

$$\begin{aligned}
\mathbb{E}(\phi(Y_n)|\mathcal{Y}_{n-1}) &= \mathbb{E}(\phi(h(X_n) + V_n)|\mathcal{Y}_{n-1}) \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \phi(h(x) + v) \mathbb{P}_{(X_n, V_n)}(dx, dv|\mathcal{Y}_{n-1}) \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \phi(h(x) + v) \mathbb{P}_{X_n}(dx|\mathcal{Y}_{n-1}) \mathbb{P}_{V_n}(dv) \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \phi(h(x) + v) f_V(v) dv \mu_n^-(dx) \\
&= \int_{\mathbb{R}^d} \phi(y) \left[\int_{\mathbb{R}^d} f_V(y - h(x)) \mu_n^-(dx) \right] dy
\end{aligned}$$

Lemme 3 *La loi conditionnelle de (X_n, Y_n) sachant \mathcal{Y}_{n-1} est donnée par :*

$$\mathbb{P}_{(X_n, Y_n)}(dx, dy|\mathcal{Y}_{n-1}) = f_V(y - h(x)) \mu_n^-(dx) dy$$

Preuve

En effet, si ϕ est une fonction test mesurable bornée, et comme V_n est indépendante de (X_n, \mathcal{Y}_{n-1})

$$\begin{aligned}
\mathbb{E}(\phi(X_n, Y_n)|\mathcal{Y}_{n-1}) &= \mathbb{E}(\phi(X_n, h(X_n) + V_n)|\mathcal{Y}_{n-1}) \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \phi(x, h(x) + v) \mathbb{P}_{(X_n, V_n)}(dx, dv|\mathcal{Y}_{n-1}) \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \phi(x, h(x) + v) \mathbb{P}_{X_n}(dx|\mathcal{Y}_{n-1}) \mathbb{P}_{V_n}(dv) \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \phi(x, h(x) + v) f_V(v) dv \mu_n^-(dx) \\
&= \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} \phi(x, y) f_V(y - h(x)) \mu_n^-(dx) \right] dy
\end{aligned}$$

On a la proposition suivante sur la prédiction et la correction d'après la formule de Bayes :

Proposition 4 *La loi prédite $\mathbb{P}_{X_n}(dx|\mathcal{Y}_{n-1})$ de X_n sachant les observations jusqu'à l'instant $n - 1$ est :*

$$\mu_n^-(dx) = \int_{\mathbb{R}^d} \pi(x'; dx) \mu_{n-1}(dx'). \quad (1.16)$$

La loi corrigée $\mathbb{P}_{X_n}(dx|\mathcal{Y}_n)$ de X_n sachant les observations jusqu'à l'instant n est :

$$\mu_n(dx) = C_n f_V(Y_n - h(x)) \mu_n^-(dx). \quad (1.17)$$

C_n est une constante de normalisation

$$\text{telle que : } C_n = \int_{\mathbb{R}^d} f_V(Y_n - h(x')) \mu_n^-(dx')$$

Preuve

si ϕ est une fonction mesurable bornée,

$$\begin{aligned}\mathbb{E}(\phi(X_n|\mathcal{Y}_{n-1})) &= \mathbb{E} \left[\int_{\mathbb{R}^d} \phi(x') \pi(X_{n-1}; dx') | \mathcal{Y}_{n-1} \right] \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \phi(x') \pi(x; dx') \mu_n^-(dx)\end{aligned}$$

On a effectivement la loi prédite (1.16) :

$$\mu_n^-(dx) = \int_{\mathbb{R}^d} \pi(x'; dx) \mu_{n-1}(dx')$$

Avec la formule de Bayes, on a

$$\begin{aligned}\mathbb{P}_{(X_n, Y_n)}(dx, dy | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}) &= \\ &= \mathbb{P}_{Y_n}(dy | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}) \times \mathbb{P}_{X_n}(dx | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}, Y_n = y_n)\end{aligned}$$

D'après le lemme 2 et le lemme 3,

$$\begin{aligned}f_V(y - h(x)) \mathbb{P}_{X_n}(dx | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}) dy &= \\ \left[\int_{\mathbb{R}^d} f_V(y - h(x')) \mathbb{P}_{X_n}(dx' | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}) \right] dy &\times \mathbb{P}_{X_n}(dx | Y_1 = y_1, \dots, Y_n = y)\end{aligned}$$

Donc, $\mathbb{P}_{X_n}(dx, dy | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}, Y_n = y_n) =$

$$= \frac{f_V(y - h(x)) \mathbb{P}_{X_n}(dx | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1})}{\left[\int_{\mathbb{R}^d} f_V(y - h(x')) \mathbb{P}_{X_n}(dx' | Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}) \right]}$$

Ainsi, la loi corrigée (1.17) est telle que

$$\mu_n(dx) = \frac{f_V(Y_n - x) \mu_n^-(dx)}{\int_{\mathbb{R}^d} f_V(Y_n - x') \mu_n^-(dx')}$$

1.3.3 Filtrage avec une loi initiale à densité

Supposons que la loi de la condition initiale X_0 admet une densité $p_0(x)$. La loi prédite $\mu_1^-(dx)$ est telle que : pour toute fonction test ϕ mesurable bornée,

$$\mathbb{E}[\phi(X_1)] = \int_{\mathbb{R}^d} \phi(x) \mu_1^-(dx) = \int_{\mathbb{R}^d} \phi(x) p_0(x) dx$$

et $\mu_1^-(dx) = p_0(x) dx = p_1^-(x) dx$.

Avec le système,

$$\begin{cases} X_{n+1} &= f(X_n) + W_n \\ Y_n &= h(X_n) + V_n \end{cases}$$

$$\begin{aligned}\mathbb{E}[\phi(X_1)|\mathcal{Y}_1] &= \mathbb{E}[\phi(f(X_0) + W_0)|\mathcal{Y}_1] \\ &= \end{aligned}$$

$$\int_{\mathbb{R}^d} f(x)\mu_1(dx)$$

De même pour la loi corrigée $\mu_1(dx)$

$$\mathbb{E}[f(X_1)|\mathcal{Y}_1] = \int_{\mathbb{R}^d} f(x)\mu_1(dx) = \int_{\mathbb{R}^d} f(x) \frac{f_V(Y_1 - x)p_0(x)}{\int_{\mathbb{R}^d} f_V(Y_1 - x')p_0(x')dx'} dx.$$

Et $\mu_1(dx)$ est à densité telle que $p_1(x) = \frac{f_V(Y_1 - x)p_0(x)}{\int_{\mathbb{R}^d} f_V(Y_1 - x')p_0(x')dx'}$.

Par récurrence, supposons que la loi corrigée $\mu_{n-1}(dx) = p_{n-1}(x)dx$ donc à densité,

La loi conditionnelle prédite admet donc une densité telle que :

$$\begin{aligned}\mu_n^-(dx) &= p_n^-(x)dx \\ p_n^-(x) &= \int_{\mathbb{R}^d} f_W(x - f(x'))p_{n-1}(x')dx'.\end{aligned}\tag{1.18}$$

De même, la loi conditionnelle corrigée admet une densité telle que :

$$\begin{aligned}\mu_n(dx) &= p_n(x)dx \\ p_n(x) &= \frac{f_V(Y_n - h(x))p_n^-(x)}{\int_{\mathbb{R}^d} f_V(Y_n - h(x'))p_n^-(x')dx'}\end{aligned}\tag{1.19}$$

À partir de la condition initiale de densité p_0 ,

$$p_1(x) = \frac{f_V(Y_1 - h(x))p_0(x)}{\int_{\mathbb{R}^d} f_V(Y_1 - h(x'))p_0(x')dx'}$$

En général, les densités des filtres n'ont pas d'expressions analytiques simples d'autant plus que h est non linéaire et la détermination du filtre se fait numériquement même si p_0 et f_V sont gaussiennes

..

Dans la suite, par application des constructions des filtres optimaux, considérons les modèles contenant des paramètres inconnus. Il s'agit de déterminer à la fois les paramètres et les filtres optimaux.

1.4 Filtrage non linéaire de système paramétré

Les équations d'état et d'observation peuvent contenir des paramètres inconnus θ comme dans le modèle ARMA [ANNEXE C] sur l'équation d'état. Le filtrage consiste à estimer à la fois les variables d'état non observés X_n et les paramètres inconnus constants θ . On pourrait directement estimer θ par une statistique $T(\theta)$ et appliquer les méthodes classiques de maximum de vraisemblance ou de minimum de variance.

Les étapes de prédiction et de correction dans le filtrage exige l'inséparabilité des deux estimations.

On introduit alors une loi à priori π_θ et notons la nouvelle variable d'état par $X_n^\varepsilon = (X_n, \theta_n)$.

Le paramètre θ est donc estimé à chaque observation n et cette estimation est utilisée dans la prochaine observation $n + 1$.

1.4.1 Modélisation en temps discret

Considérons le modèle général avec un système d'équation d'état et d'équation d'observation tel que

$$\begin{cases} X_{n+1} &= f_n(X_n, \theta_n, \varepsilon_n) \\ Y_n &= h_n(X_n, \theta_n, \eta_n) \end{cases}$$

θ_n est un paramètre à estimer et h_n une fonction non linéaire caractéristique dans l'équation d'observation.

ε_n et η_n sont des bruits blancs indépendants de l'état et de l'observation.

Dans le cas habituel de bruits additifs,

$$\begin{cases} X_{n+1} &= f_n(X_n, \theta_n) + W_n \\ Y_n &= h_n(X_n, \theta_n) + V_n \end{cases}$$

V_n et W_n sont des bruits blancs non nécessairement gaussiens.

Avec la nouvelle variable d'état augmenté $X_n^\theta = (X_n, \theta_n)$, on modifie le système en :

$$\begin{cases} X_{n+1}^\theta &= f_n(X_n^\theta) + W_n \\ Y_n^\theta &= h_n(X_n^\theta) + V_n \end{cases}$$

Avec la loi à priori π_θ , on obtient une estimation θ_0 de θ à l'instant initial.

Le filtre à l'instant initial est μ_0 et la condition initiale est (μ_0, θ_0) .

Dans le cas des bruits à densité, le filtre prédit paramétré par θ est :

$$\mu_n^{-\theta}(dx) = \mathbb{P}(X_n^\theta \in dx | \mathcal{Y}_{n-1}) = \int_{\mathbb{R}^d} f_W(x - f_n(x')) \mu_{n-1}^\theta(dx'). dx \quad (1.20)$$

Le filtre corrigé paramétré par θ est :

$$\mu_n^\theta(dx) = \mathbb{P}(X_n^\theta \in dx | \mathcal{Y}_n) = \frac{f_V(Y_n - h_n(x)) \mu_n^{-\theta}(dx)}{\int_{\mathbb{R}^d} f_V(Y_n - h_n(x')) \mu_n^{-\theta}(dx')} \quad (1.21)$$

Le filtre optimal est alors défini par

$$\mu_n(dx) = \mathbb{P}(X_n \in dx | \mathcal{Y}_n) = \int_{\mathbb{R}^d} \mathbb{P}(X_n^\theta \in dx | \mathcal{Y}_n) \cdot \pi_\theta(d\theta) = \int_{\mathbb{R}^d} \mu_n^\theta(dx) \cdot \pi_\theta(d\theta).$$

L'estimation $\hat{\theta}_n$ de θ à l'instant n est obtenue par l'équation :

$$\hat{X}_{n+1} = f_n(\hat{X}_n, \hat{\theta}_n)$$

où $\hat{X}_{n+1} = \mathbb{E}(X_{n+1} | \mathcal{Y}_{n+1})$ et $\hat{X}_n = \mathbb{E}(X_n | \mathcal{Y}_n)$.

1.4.2 Modèle d'état linéaire paramétré

Pour le modèle d'état supposé linéaire et stationnaire :

$$\begin{cases} Y_t & = MX_t + V_t \\ \frac{dX_t}{dt} & = \theta X_t + \sigma \cdot \varepsilon_t \end{cases}$$

$X_t \in \mathbb{R}^n$ est le vecteur d'état du système,

$\theta \in \mathbb{R}$ est le paramètre déterministe à estimer

$\varepsilon_t \in \mathbb{R}^q$ est le vecteur des signaux aléatoires inconnus qui viennent perturber directement l'équation d'état du système à travers une matrice d'entrée $M_{n \times q}$,

$Y_t \in \mathbb{R}^p$ est le vecteur des mesures d'observation,

$V_t \in \mathbb{R}^p$ est le vecteur des signaux aléatoires, bruit des mesures, qui polluent les mesures Y_t .

M est une matrice $p \times n$ et σ un réel.

Sans bruit de mesure d'état, on a l'intégrale de l'équation différentielle linéaire du premier ordre

$$x_t = \exp(\theta(t - t_0))x_{t_0}$$

x_t est la trajectoire de référence ou la trajectoire nominale.

Comme les mesures d'observation se font en temps discret, considérons le système d'équations

$$\begin{cases} Y_n & = MX_n + V_n \\ X_{n+1} & = \theta X_n + W_n \end{cases}$$

Par application de la section précédente, le filtre prédit paramétré est tel que :

$$\mu_n^{-\theta}(dx) = \int_{\mathbb{R}^d} f_W(x - \theta_n \cdot x') \mu_{n-1}^{\theta}(dx') \cdot dx \quad (1.22)$$

et le filtre corrigé paramétré est :

$$\mu_n^{\theta}(dx) = \frac{f_V(Y_n - Mx) \mu_n^{-\theta}(dx)}{\int_{\mathbb{R}^d} f_V(Y_n - Mx') \mu_n^{-\theta}(dx')} \quad (1.23)$$

On obtient le filtre optimal par :

$$\mu_n(dx) = \int_{\mathbb{R}^d} \mu_n^{\theta}(dx) \cdot \pi_{\theta}(d\theta). \quad (1.24)$$

L'estimation $\hat{\theta}_n$ de θ à l'instant n est telle que :

$$\hat{X}_{n+1} = \hat{\theta}_n \cdot \hat{X}_n.$$

\hat{X}_k est l'estimation de l'état à l'instant k définie par $\mathbb{E}(X_k | \mathcal{Y}_k)$.

1.5 Filtrage non linéaire de système contrôlé

Considérons un système dynamique tel qu'à chaque instant t , on applique un contrôle³ u_t .
 u_t est une variable déterministe.

On a le système de l'équation d'état et de l'équation d'observation :

$$\begin{cases} \frac{dx_t}{dt} = f_t(x_{t-1}, u_t, \varepsilon_t), & t > 0, \quad x_0 \text{ est connue.} \\ y_t = h_t(x_t, \eta_t). \end{cases}$$

- $x_t \in \mathbb{R}^d$ variable d'état non observée.
- $y_t \in \mathbb{R}^k$ variable d'observation observée.
- ε_t bruit blanc du modèle d'état.
- η_t bruit blanc du modèle d'observation.
- $u_t \in C_u \subset \mathbb{R}^l$ variable de contrôle dans un compact C_u .

Comme la variable de contrôle est déterministe et connue à chaque instant, on peut considérer le système équivalent,

$$\begin{cases} \frac{dx_t}{dt} = \tilde{f}_t(x_{t-1}, \varepsilon_t), & t > 0, \quad x_0 \text{ est connue.} \\ y_t = h_t(x_t, \eta_t) \end{cases}$$

où $\tilde{f}_t(.,.,.) = f_t(., u_t, .)$ est une fonction non linéaire et on peut appliquer les méthodes de filtrage non linéaire.

En général, le contrôle est déterminé suivant un critère d'optimisation. On considère alors un problème en boucle fermée ou les contrôles sont pris au cours du temps et dépend des observations faites par le système au cours de son évolution.

On introduit une fonction coût définie entre deux instant t_0 et t telle que

$$J(t, u(t)) = \mathbb{E} \left[\int_{t_0}^t F(s, x_u(s), u(s)) ds + G(s, x_u(s)) \right] \quad (1.25)$$

Pour tout instant t , le contrôle u_t est une variable aléatoire.

L'objectif est de déterminer les trajectoires $x_{u^*}(t)$ où le contrôle optimal est défini par

$$u^*(t) = \underset{u \in \mathcal{U}}{\operatorname{argmin}} J(t, u(t))$$

\mathcal{U} est l'ensemble des contrôles admissibles.

3. ou commande

1.5.1 Modélisation en temps discret

Le système dynamique décrit l'évolution du processus périodiquement. Il en est de même pour le processus d'observation.

Considérons le système d'équations :

$$\begin{cases} X_n &= f(X_{n-1}, u_n, W_n) \\ Y_n &= h(X_n, u_n, V_n) \end{cases}$$

et la fonction coût associée à chaque période est

$$J_n(u) = \mathbb{E} \left[\sum_{k=j}^{n-1} g_k(X_k, u_k, W_k) + G_n(X_n) \right], \quad 0 \leq j \leq n.$$

Contrairement aux problèmes en boucle ouverte où les décisions peuvent être prises à priori avant que le système n'évolue, c'est à dire que la commande est la même quelque soit la réalisation, entre deux observations, on considère une stratégie⁴ pour déterminer l'estimation de X_n .

Notons alors $\tilde{X}_n = \mathbb{E}[X_n | Y_1, \dots, Y_n, u_1, \dots, u_n]$ avec $(n-1) \leq k \leq n$. La fonction coût à minimiser est :

$$J_n(u) = \sum_{k=j}^n L_k(\tilde{X}_k, u_k), \quad 0 \leq j \leq n.$$

L_k est une fonction associée à g_k .

La commande optimale est telle que : $u^* = (u_j^*, \dots, u_n^*)$,

$$J_n(u^*) = \min_{u \in U^{n-j}} J(u)$$

Une situation fréquente est la suivante :

Pour maintenir X_n proche de la trajectoire nominale x_n , on peut définir la commande optimale :

$$(u_j^*, \dots, u_n^*) = \operatorname{argmin}_{u_j, \dots, u_n} \sum_{k=j}^n \|x_k^* - \tilde{X}_n\|^2.$$

La commande optimale est introduite dans le système d'équation dynamique d'état et d'équation d'observation.

Remarques

On peut ne considérer que les décisions antérieures assez proches du temps de l'observation. L'intérêt de la commande optimale dans le problème de trajectographie est le pouvoir de mieux contrôler l'évolution du système dynamique et de mener l'engin à s'approcher le plus près possible de la trajectoire prédéfinie.

4. une suite de contrôles

Le noyau de transition est tel que :

$$\begin{aligned}\pi_n(x; dx') &= \mathbb{P}(X_{n+1} \in dx' | X_n = x) \\ &= \mathbb{P}[f(X_n) + W_n \in dx' | X_n = x] \\ &= \mathbb{P}[f(x) + W_n \in dx'] \\ &= f_{W'}(x') dx'\end{aligned}$$

À partir du système

$$\begin{cases} X_n &= f(X_{n-1}, u_n) + W_n \\ Y_n &= h(X_n, u_n) + V_n \end{cases}$$

et d"après (1.16), la loi prédite est telle que :

$$\mu_n^-(dx) = \int_{\mathbb{R}^d} f_W(x - \tilde{f}(x')) \mu_{n-1}(dx') = \int_{\mathbb{R}^d} f_W(x - f(x', u_n)) \mu_{n-1}(dx').$$

de même d"après (1.17), la loi corrigée est telle que :

$$\mu_n(dx) = \frac{f_V(Y_n - \tilde{h}(x)) \mu_n^-(dx)}{\int_{\mathbb{R}^d} f_V(Y_n - \tilde{h}(x')) \mu_n^-(dx')} = \frac{f_V(Y_n - h(x, u_n)) \mu_n^-(dx)}{\int_{\mathbb{R}^d} f_V(Y_n - h(x', u_n)) \mu_n^-(dx')}.$$

La commande u_n est estimée à partir de la relation :

$$Y_n = h(\hat{X}_n, u_n)$$

où \hat{X}_n est l'estimation de l'état X_n à l'instant n obtenue du filtre optimal $\mu_n(dx)$.

Dans la pratique, le filtre μ_n est une mesure à densité.

1.5.2 Modèle d'état linéaire contrôlé

Avec un modèle d'état supposé linéaire et stationnaire⁵, on a le système :

$$\begin{cases} y_t &= Cx_t + Du_t + v_t \\ \frac{dx_t}{dt} &= Ax_t + Bu_t + Mw_t \end{cases}$$

$x_t \in \mathbb{R}^n$ est le vecteur d'état du système,

$u_t \in \mathbb{R}^m$ est le vecteur des entrées déterministes et connues (commandes,...),

C est une matrice $p \times n$ et D est une matrice $p \times m$.

A est une matrice $n \times n$ et B est une matrice $n \times p$.

$w_t \in \mathbb{R}^q$ est le vecteur des signaux aléatoires inconnus qui viennent perturber directement l'équation d'état du système à travers une matrice d'entrée M de dimension $n \times q$,

5. définition à l'annexe A

$y_t \in \mathbb{R}^p$ est le vecteur des mesures,

$v_t \in \mathbb{R}^p$ est le vecteur des signaux aléatoires, bruit des mesures, qui polluent les mesures y_t .

Sans perturbation, et dans le cas des entrées déterministes u_t sur un horizon fini $t \in [t_0; t_1]$, la solution de l'équation dynamique est telle que, avec des conditions initiales x_{t_0} :

$$x_t = \exp(A(t - t_0))x_{t_0} + \int_{t_0}^t \exp(A(t - \tau))Bu_\tau d\tau$$

x_t est la trajectoire de référence ou la trajectoire nominale.

$$y_t = Cx_t + Du_t + v_t$$

Comme dans le cas non linéaire, on détermine à chaque instant, la commande optimale et on l'introduit dans le modèle discret de filtrage.

$$\begin{cases} X_n &= A.X_{n-1} + B.u_{n-1} + W_n \\ Y_n &= C.X_n + D.u_n + V_n \end{cases}$$

La densité du filtre prédit est telle que :

$$p_n^-(x) = \int_{\mathbb{R}^d} f_W(x - A.x' - B.u_{n-1})p_{n-1}(x')dx'$$

La densité du filtre corrigé est telle que :

$$p_n(x) = \frac{f_V(Y_n - C.x - D.u_{n-1})p_n^-(x)}{\int_{\mathbb{R}^d} f_V(Y_n - C.x' - D.u_{n-1})p_n^-(x')dx'}$$

La commande à l'instant n est déterminée par :

$$D.u_n = Y_n - C.\hat{X}_n$$

où \hat{X}_n est l'estimation de X_n .

1.6 Conclusion

Dans ce chapitre, dans la modélisation du problème de filtrage, la structure markovienne du processus d'état non observable se présente sous deux formes différentes : soit on connaît exactement l'équation d'état dans laquelle le bruit est aussi connu, soit on connaît son noyau de transition. Les bruits de mesure d'état et de mesure d'observation sont les principaux facteurs dans la détermination du filtre.

Nous avons considéré deux méthodes analytiques :

-d'une part le problème de filtrage non linéaire est associé à un problème d'équations aux dérivées partielles [PARDOUX,1981]. Dans le modèle de diffusion, avec la formule d'Îto, l'existence et l'unicité de la solution sont justifiées avec des hypothèses bien adaptées. Le filtre est une solution de l'équation aux dérivées partielles stochastiques.

-d'autre part, le filtre est calculé directement à partir des lois de distribution des bruits. Cette solution paraît idéale mais le filtre n'a pas de formes très explicites même dans le cas où la condition initiale a une loi à densité.

Les deux approches nécessitent des calculs numériques pour pouvoir être exploitées.

Ces calculs analytiques peuvent être appliquées en filtrage non linéaires de système paramétré ou de système contrôlé. Les différentes méthodes numériques sont choisies en fonction de leurs facilités d'adaptation pour chaque situation.

Dans le chapitre suivant, nous énonçons le filtrage de Kalman pour un système linéaire gaussien. Les résolutions de ses extensions dans les systèmes non linéaires sont basées sur des approximations. Dans le cas plus général de filtrage non linéaire, les méthodes particulières sont introduites.

Chapitre 2

Méthodes numériques de filtrages non linéaires

Sommaire

2.1 Introduction	27
2.2 Filtre de Kalman	28
2.2.1 Filtre linéaire gaussien	29
2.2.2 Filtre Kalman étendu	33
2.3 Filtrage particulaire	36
2.3.1 Formules de Bayes	37
2.3.2 Echantillonnage d'importance	37
2.3.3 Filtre particulaire ou bootstrap	38
2.3.4 Résolution par filtrage particulaire	40
2.4 Filtrage par noyau de convolution	41
2.4.1 Noyau de convolution	41
2.4.2 Estimation par noyau de convolution	44
2.4.3 Filtrage par noyau de convolution	47
2.4.4 Résolution par filtrage par noyau	49
2.5 Conclusion	51

2.1 Introduction

En théorie, l'existence du filtre optimal peut être vérifiée par différentes méthodes mais en général, il n'existe pas de solution explicite. Des méthodes numériques sont utiles dans les applications. Dans ce memoire en particulier, il s'agit du filtrage de Kalman , du filtrage particulaire et du filtrage par noyau de convolution :

- Le filtrage de Kalman, l'une des premières méthodes de résolution de problème de filtre, est utilisé dans beaucoup de domaines et son application ne présente pas trop de difficultés. Ce filtrage présente quand même de défaut dans son adaptation. En effet, il exige l'hypothèse de gaussianité de la loi initiale et des bruits. À l'origine, il était limité à des systèmes linéaires. Une extension sur des systèmes non linéaires est adaptée avec des hypothèses supplémentaires.

- Le filtrage particulaire est une méthode bien adaptée pour des systèmes non linéaires et sans condition de gaussianité sur les bruits de mesures. Il a la particularité de gérer des particules pour déterminer le filtre optimal défini par sa loi de probabilité. L'un des défauts du filtrage particulaire se trouve dans le problème de convergence. En effet, les particules divergent rapidement pour un système à bruits faibles.

- Le filtrage par noyau de convolution est une méthode particulière qui est adaptée à des systèmes à bruits faibles. Cette méthode est basée sur la théorie des distributions et des estimations paramétriques et non paramétriques.

2.2 Filtre de Kalman

Le filtre de Kalman [ALAZARD,2006], [ENZO et al.,2010] permet de donner un estimé de l'état du système à partir d'informations à priori sur l'évolution de cet état et de mesures réelles. Il est utilisé pour estimer des conditions initiales inconnues, dans le problème balistique, dans la prédiction des trajectoires des mobiles [GUECHI, 2010] [MASCARILLA,2004], dans le problème de trajectographie, dans la localisation d'un engin par un radar ou un sonar [GIREMUS,2005], [COLLIN et al.,1999] et également dans l'implantation des lois de commandes fondées sur un estimateur de l'état et un retour d'état [RACICOT et THEORET,2005] comme dans le problème commande quadratique gaussienne.

Les bases de traitement de signal sur lesquelles repose le filtre de Kalman seront également utiles dans beaucoup de domaines entre autres la détermination expérimentale de certains paramètres du modèle [BREHARD, 2006].

L'objectif est d'estimer le vecteur aléatoire X_n à partir des observations Y_1, \dots, Y_n de façon optimale et récursive. Avec le critère du minimum de variance, il s'agit de calculer la distribution de probabilité conditionnelle du vecteur aléatoire X_n sachant Y_1, \dots, Y_n .

Le filtrage de Kalman se déroule en deux phases : la prédiction et la mise à jour :

- la prédiction utilise l'état estimé de l'instant précédent pour produire une estimation de l'état courant.
- la mise à jour avec les observations à l'instant courant est appliquée pour corriger l'état prédit dans le but d'avoir une estimation plus précise.

Le filtrage de Kalman traite uniquement le modèle linéaire où tout est gaussien. Les filtres prédits et corrigés sont gaussiens et il suffit de calculer les paramètres pour la loi corrigée :

-la moyenne

$$\hat{X}_n = \mathbb{E}[X_n | \mathcal{Y}_n] \quad (2.1)$$

-la matrice de covariance

$$P_n = \mathbb{E} \left[(X_n - \hat{X}_n)(X_n - \hat{X}_n)^t | \mathcal{Y}_n \right] \quad (2.2)$$

On définit également les quantités suivantes du filtre prédit :

-la moyenne

$$\hat{X}_n^- = \mathbb{E}[X_n | \mathcal{Y}_{n-1}] \quad (2.3)$$

-la matrice de covariance

$$P_n^- = \mathbb{E}[(X_n - \hat{X}_n)(X_n - \hat{X}_n)^t | \mathcal{Y}_{n-1}] \quad (2.4)$$

En fait, les covariances conditionnelles sont indépendantes des observations

$$P_n = \mathbb{E}[(X_n - \hat{X}_n)(X_n - \hat{X}_n)^t] \text{ et } P_n^- = \mathbb{E}[(X_n - \hat{X}_n)(X_n - \hat{X}_n)^t]. \quad (2.5)$$

2.2.1 Filtre linéaire gaussien

On considère une suite d'états non observable $\{X_n, n \geq 0\}$ à valeurs dans \mathbb{R}^d , vérifiant la relation de récurrence

$$X_{n+1} = F_{n+1}X_n + f_{n+1} + G_{n+1}W_{n+1} \quad (2.6)$$

La suite de bruits $\{W_n, n \geq 0\}$ est à valeurs dans \mathbb{R}^d .

$\{Y_n, n \geq 1\}$ est une suite d'observations à valeurs dans \mathbb{R}^p vérifiant

$$Y_n = H_n X_n + h_n + V_n \quad (2.7)$$

De plus, on fait les hypothèses suivantes :

- la condition initiale X_0 est gaussienne, de moyenne \bar{X}_0 et de matrice de covariance Q_0
- W_n est un bruit blanc gaussien, de matrice de covariance Q_n^W
- V_n est un bruit blanc gaussien, de matrice de covariance Q_n^V
- les suites $\{W_n\}$ et $\{V_n\}$ et X_0 sont mutuellement indépendantes.

On a le lemme suivant pour le vecteur (X_n, Y_n) :

Lemme 5 *La suite $\{U_n = (X_n, Y_n)\}$ est un processus aléatoire gaussien à valeurs dans \mathbb{R}^{d+p} et tel que $\mathbb{E}(U_n) = \begin{bmatrix} \bar{X}_n \\ \bar{Y}_n \end{bmatrix}$ et de matrice de covariance $Q_Z = \begin{bmatrix} Q_n^X & Q_n^{XY} \\ Q_n^{YX} & Q_n^Y \end{bmatrix}$.*

En supposant que $Q_n^Y > 0$, la loi de probabilité conditionnelle de X_n sachant Y_n est gaussienne $\mathcal{N}(\hat{X}_n, \hat{S}_n)$ telle que :

$$\hat{X}_n = \bar{X}_n + Q_n^{XY} \cdot (Q_n^Y)^{-1} \cdot (Y_n - \bar{Y}_n) \quad (2.8)$$

$$\hat{S}_n = Q_n^X - Q_n^{XY} \cdot (Q_n^Y)^{-1} \cdot Q_n^{YX} \quad (2.9)$$

Preuve

Les moyennes de X_n et Y_n sont déterminées par :

$$\bar{X}_n = \mathbb{E}[X_n] = F_n \mathbb{E}[X_{n-1}] + f_n + G_n \mathbb{E}[W_n] = F_n \bar{X}_{n-1} + f_n$$

$$\bar{Y}_n = \mathbb{E}[Y_n] = H_n \mathbb{E}[X_n] + h_n + \mathbb{E}[V_n] = H_n \bar{X}_n + h_n$$

De mêmes les variances de X_n et Y_n sont déterminées par :

$$\begin{aligned}
Q_n^X &= \mathbb{E} [(X_n - \bar{X}_n)(X_n - \bar{X}_n)^t] \\
&= \mathbb{E} [(F_n(X_{n-1} - \bar{X}_{n-1}) + G_n W_n)(F_n(X_{n-1} - \bar{X}_{n-1}) + G_n W_n)^t] \\
&= F_n \mathbb{E} [(X_{n-1} - \bar{X}_{n-1})(X_{n-1} - \bar{X}_{n-1})^t] F_n^* + G_n \mathbb{E} [W_n W_n^t] G_n^t + \\
&+ F_n \mathbb{E} [(X_{n-1} - \bar{X}_{n-1}) W_n^t] G_n^t + G_n \mathbb{E} [W_n (X_{n-1} - \bar{X}_{n-1})^t] F_n^t \\
&= F_n Q_{n-1}^X F_n^t + G_n Q_n^W G_n^t
\end{aligned}$$

$$\begin{aligned}
Q_n^Y &= \mathbb{E} [(Y_n - \bar{Y}_n)(Y_n - \bar{Y}_n)^t] \\
&= \mathbb{E} [(H_n(X_{n-1} - \bar{X}_{n-1}) + V_n)(H_n(X_{n-1} - \bar{X}_{n-1}) + V_n)^t] \\
&= H_n \mathbb{E} [(X_{n-1} - \bar{X}_{n-1})(X_{n-1} - \bar{X}_{n-1})^t] H_n^t + \mathbb{E} [V_n V_n^t] + \\
&+ H_n \mathbb{E} [(X_{n-1} - \bar{X}_{n-1}) V_n^t] + \mathbb{E} [V_n (X_{n-1} - \bar{X}_{n-1})^t] H_n^t \\
&= H_n Q_n^X H_n^t + Q_n^V
\end{aligned}$$

La matrice de covariance de X_n et Y_n est telle que :

$$\begin{aligned}
Q_n^{XY} &= \mathbb{E} [(X_n - \bar{X}_n)(Y_n - \bar{Y}_n)^t] \\
&= \mathbb{E} [(X_n - \bar{X}_n)(H_n(X_n - \bar{X}_n) + V_n)^t] \\
&= \mathbb{E} [(X_n - \bar{X}_n)(X_n - \bar{X}_n)^t] H_n^t + \mathbb{E} [(X_n - \bar{X}_n) V_n^t] \\
&= Q_n^X H_n^t
\end{aligned}$$

En posant $\tilde{X}_n = X_n - \hat{X}_n$, $\begin{pmatrix} \tilde{X}_n \\ Y_n \end{pmatrix}$ est un vecteur aléatoire gaussien tel que $\text{cov}(\tilde{X}_n, Y_n) = 0$. \tilde{X}_n et Y_n sont indépendantes et \hat{X}_n est une fonction de Y_n .

Soit $\varphi \in \mathcal{C}_b(\mathbb{R}^d)$,

$$\begin{aligned}
\mathbb{E}[\varphi(X_n)|Y_n] &= \mathbb{E}[\varphi(\hat{X}_n + \tilde{X}_n)|Y_n] \\
&= \int_{\mathbb{R}^d} \varphi(\hat{X}_n + x) \mathbb{P}(\tilde{X}_n \in dx | Y_n) \\
&= \int_{\mathbb{R}^d} \varphi(\hat{X}_n + x) \mathbb{P}_{\tilde{X}_n}(dx) \\
&= \int_{\mathbb{R}^d} \varphi(x) \mathbb{P}(X_n \in dx | Y_n).
\end{aligned}$$

Et

$$\begin{aligned}
\text{cov}(\tilde{X}_n) &= \text{cov}(X_n - \hat{X}_n) \\
&= \text{cov}(X_n - \bar{X} - Q_n^{XY} (Q_n^Y)^{-1} (Y_n - \bar{Y}_n)) \\
&= Q_n^X - 2Q_n^{XY} (Q_n^Y)^{-1} Q_n^{YX} + Q_n^{XY} (Q_n^Y)^{-1} Q_n^{YX} \\
&= Q_n^X - Q_n^{XY} (Q_n^Y)^{-1} Q_n^{YX}
\end{aligned}$$

On a donc, d'après le lemme 5 : Dans l'étape de prédiction, avec l'équation (2.6), compte tenu de l'indépendance de W_n et la tribu \mathcal{Y}_{n-1} , la moyenne estimée du filtre prédit est

$$\begin{aligned}\hat{X}_n^- &= \mathbb{E}[X_n | \mathcal{Y}_{n-1}] \\ &= \mathbb{E}[F_n X_{n-1} + f_n + G_n W_n | \mathcal{Y}_{n-1}] \\ &= F_n \mathbb{E}[X_{n-1} | \mathcal{Y}_{n-1}] + f_n \\ &= F_n \hat{X}_{n-1} + f_n\end{aligned}$$

et l'estimation de la variance du filtre prédit est

$$\begin{aligned}P_n^- &= \mathbb{E}[(X_n - \bar{X}_n)(X_n - \bar{X}_n)^t | \mathcal{Y}_{n-1}] \\ &= \mathbb{E}[(X_n - \bar{X}_n)(X_n - \bar{X}_n)^t] \\ &= F_n Q_{n-1}^X F_n^t + G_n Q_n^W G_n^t \\ &= F_n P_{n-1} F_n^t + G_n Q_n^W G_n^t\end{aligned}$$

Dans l'étape de correction avec l'équation (2.7), on introduit le processus d'innovation avec la nouvelle observation Y_n ,

$$Z_n = Y_n - \mathbb{E}[Y_n | \mathcal{Y}_{n-1}]$$

, compte tenu de l'indépendance de V_n et la tribu \mathcal{Y}_{n-1}

$$\begin{aligned}Z_n &= Y_n - (H_n \mathbb{E}[Y_n | \mathcal{Y}_{n-1}] + h_n + \mathbb{E}[V_n | \mathcal{Y}_{n-1}]) \\ &= Y_n - (H_n \hat{X}_n^- + h_n) \\ &= H_n (X_n - \hat{X}_n^-) + V_n.\end{aligned}$$

Z_n apporte les nouvelles informations de Y_n par rapport aux observations passées \mathcal{Y}_n .

Lemme 6 *Le processus $\{Z_n\}$ est un processus gaussien à valeurs dans \mathbb{R}^d de moyenne nulle et de matrice de covariance, indépendant de la tribu \mathcal{Y}_{n-1} .*

$$Q_n^Z = H_n P_n^- H_n^t + Q_n^V \quad (2.10)$$

Preuve

En effet, compte tenu de l'indépendance de V_n avec $(X_n - \hat{X}_n^-)$

$$\begin{aligned}Q_n^Z &= \mathbb{E}[Z_n Z_n^t] \\ &= \mathbb{E}[(H_n (X_n - \hat{X}_n^-) + V_n)(H_n (X_n - \hat{X}_n^-) + V_n)^t] \\ &= H_n \mathbb{E}[(X_n - \hat{X}_n^-)(X_n - \hat{X}_n^-)^t] H_n^t + \mathbb{E}[V_n (X_n - \hat{X}_n^-)^t] H_n^t + \\ &\quad + H_n \mathbb{E}[(X_n - \hat{X}_n^-) V_n^t] + \mathbb{E}[V_n V_n^t] \\ &= H_n P_n^- H_n^t + Q_n^V.\end{aligned}$$

Et on a aussi $\mathbb{E}[(X_n - \hat{X}_n^-).Z_n^t] = P_n^- .H_n^t$.

Les informations contenues dans $\{\mathcal{Y}_{n-1}, Y_n\}$ et dans $\{\mathcal{Y}_{n-1}, Z_n\}$ sont identiques.

Notons le gain de Kalman

$$K_n = P_n^- H_n^t [H_n P_n^- H_n^t + Q_n^V]^{-1} = P_n^- H_n^t [Q_n^Z]^{-1} \quad (2.11)$$

On a la moyenne du filtre corrigé :

$$\hat{X}_n = \mathbb{E}[X_n | \mathcal{Y}_n] = \hat{X}_n^- + K_n \cdot [Y_n - (H_n \hat{X}_n^- + h_n)] \quad (2.12)$$

et la matrice de covariance du filtre corrigé

$$P_n = \mathbb{E}[(X_n - \bar{X}_n)(X_n - \bar{X}_n)^t | \mathcal{Y}_n] \quad P_n = [I - K_n H_n] P_n^- \quad (2.13)$$

En résumé, on a l'algorithme du filtre de Kalman :

Initialisation

$$\begin{aligned} \hat{X}_0 &= \mathbb{E}[X_0] \\ Q_0 &= \text{cov}(X_0) \end{aligned}$$

Prédiction

$$\begin{aligned} \hat{X}_n^- &= F_n \hat{X}_{n-1} + f_n \\ P_n^- &= F_n P_{n-1} F_n^t + G_n Q_n^W G_n^t \end{aligned}$$

Innovation (équation 2.10)

$$\begin{aligned} Z_n &= H_n (X_n - \hat{X}_n^-) + V_n \\ Q_n^Z &= H_n P_n^- H_n^t + Q_n^V \end{aligned}$$

Gain de Kalman (équation 2.11)

$$K_n = P_n^- H_n^t [Q_n^Z]^{-1}$$

Correction (équations 2.12 et 2.13)

$$\begin{aligned} \hat{X}_n &= \hat{X}_n^- + K_n [Y_n - (H_n \hat{X}_n^- + h_n)] \\ P_n &= [I - K_n H_n] P_n^- \end{aligned}$$

On a aussi les relations de récurrence suivantes :

$$\hat{X}_n = F_n \hat{X}_{n-1} + f_n + K_n [Y_n - H_n \cdot F_n \cdot \hat{X}_{n-1} - H_n \cdot f_n - h_n] \quad (2.14)$$

$$P_n = [I - K_n H_n] [F_n P_{n-1} F_n^t + G_n Q_n^W G_n^t] \quad (2.15)$$

Le filtrage de Kalman est adapté uniquement au système linéaire gaussien. Une extension est établie sur un modèle non linéaire gaussien. Toutefois, la non linéarité ne doit pas être forte et les fonctions dans le modèle doivent être différentiables localement pour pouvoir faire une linéarisation avec les développements classiques.

2.2.2 Filtre de Kalman étendu

Le filtre de Kalman est adapté pour les filtres linéaires gaussiens. Dans le cas non linéaire, dans l'extension du filtre de Kalman, on linéarise par une approximation fonctionnelle le système [ROSSI, 2005].

Considérons le système non linéaire gaussien [LAFARGE,2009] tel que :

$$\begin{cases} Y_n & = h_n(X_n) + V_n & (1) \\ X_{n+1} & = f_n(X_n) + g_n(X_n)W_n & (2) \end{cases}$$

V_n représente le bruit de mesure, bruit blanc gaussien de matrice de covariance Q_n^V . L'état X_n du système dynamique (2) n'est pas observé et son équation d'évolution est perturbée par un bruit blanc gaussien de matrice de covariance Q_n^W et, de condition initiale X_0 gaussienne $\mathcal{N}(\bar{X}_0, Q_0)$.

Les bruits $\{W_n\}$, $\{V_n\}$ et la condition initiale X_0 sont mutuellement indépendants.

Le critère d'optimalité est de minimiser la variance de l'erreur d'estimation donc de déterminer la loi conditionnelle de X_n sachant \mathcal{Y}_n .

Soit une suite déterministe \bar{x}_n dans \mathbb{R}^d , solution approchée du système, appelée **trajectoire nominale**. Les fonctions f_n et g_n sont supposées dérivables. On linéarise f_n et g_n autour de \bar{x}_n autour des développements habituels et on approxime

$$\begin{aligned} f_n(x) &\simeq f_n(\bar{x}_n) + \vec{\nabla} f_n(\bar{x}_n)(x - \bar{x}_n), \\ g_n(x) &\simeq g_n(\bar{x}_n) \end{aligned}$$

Il en est de même pour h_n autour de \bar{x}_n ,

$$h_n(x) \simeq h_n(\bar{x}_n) + \vec{\nabla} h_n(\bar{x}_n)(x - \bar{x}_n)$$

On obtient le système linéarisé suivant :

$$\begin{cases} Y_n & = H_n(X_n - \bar{x}_n) + \tilde{h}_n + V_n \\ X_{n+1} & = F_n(X_n - \bar{x}_n) + \tilde{f}_n + G_n W_n \end{cases}$$

où $f_n(\bar{x}_n) = \tilde{f}_n$, $F_n = \vec{\nabla} f_n(\bar{x}_n)$, $G_n = g_n(\bar{x}_n)$, $H_n = \vec{\nabla} h_n(\bar{x}_n)$ et $h_n(\bar{x}_n) = \tilde{h}_n$.

En notant, $\tilde{X}_n = X_n - \bar{x}_n$, $\tilde{f}_n = f_n - \bar{x}_{n+1}$ et $\tilde{W}_n = G_n W_n$, on a un nouveau système linéaire :

$$\begin{cases} Y_n & = H_n \tilde{X}_n + \tilde{h}_n + V_n & (3) \\ \tilde{X}_{n+1} & = F_n \tilde{X}_n + \tilde{f}_n + \tilde{W}_n & (4) \end{cases}$$

Les bruits $\{\tilde{W}_n\}$, $\{V_n\}$ et la condition initiale $\tilde{X}_0 = X_0 - \bar{x}_0$ sont mutuellement indépendants. La condition initiale est gaussienne $\mathcal{N}(\bar{X}_0 - \bar{x}_0, Q_0)$

\tilde{W}_n , la perturbation aléatoire est un bruit blanc gaussien de matrice de covariance $Q_n^{\tilde{W}}$. Dans le cas gaussien, seules la moyenne $\hat{X}_n = \mathbb{E}[\tilde{X}_n | \mathcal{Y}_n]$ et la matrice de covariance

$R_n = \mathbb{E}[(\tilde{X}_n - \hat{X}_n)(\tilde{X}_n - \hat{X}_n)^t]$ sont nécessaires à la définition de cette loi.

Pour la prévision,

$$\begin{aligned}\hat{X}_{n^-} &= \mathbb{E}[\tilde{X}_n | \mathcal{Y}_{n-1}] \\ R_{n^-} &= \mathbb{E}[(\tilde{X}_n - \hat{X}_{n^-})(\tilde{X}_n - \hat{X}_{n^-})^t]\end{aligned}$$

On construit le processus d'innovation défini par l'information apportée par Y_n par rapport aux observations passées \mathcal{Y}_{n-1} : $I_n = Y_n - \hat{Y}_{n^-}$ où $\hat{Y}_{n^-} = \mathbb{E}[Y_n | \mathcal{Y}_{n-1}]$.

L'innovation est donc : $I_n = Y_n - (H_n \hat{Y}_{n^-} + \tilde{h}_n)$.

Le processus I_n est un processus gaussien à valeurs dans \mathbb{R}^k , en particulier I_n est un vecteur aléatoire gaussien de moyenne nulle et de matrice de covariance $Q_n^I = H_n R_{n^-} H_n^t + Q_n^V$ indépendant de \mathcal{Y}_{n-1} . Et on a le théorème :

Théorème de Kalman-Bucy[PARDOUX,2006]

Si la matrice de covariance Q_n^V est inversible pour tout $n \in \mathbb{N}$, alors les processus \hat{X} et R_n sont définies par les équations suivantes :

Prédiction

$$\begin{aligned}\hat{X}_n^- &= F_n \hat{X}_{n-1} + \tilde{f}_n \\ R_n^- &= F_n R_{n-1} F_n^t + Q_n^{\tilde{W}}\end{aligned}$$

Corréction

$$\begin{aligned}\hat{X}_n &= \hat{X}_n^- + K_n [Y_n - (H_n \hat{X}_n^- + \tilde{h}_n)] \\ R_n &= [I - K_n H_n] R_n^-\end{aligned}$$

où le **gain de Kalman** est la matrice $K_n = R_n^- H_n^t [H_n R_n^- H_n^t + Q_n^V]^{-1} = R_n^- H_n^t [Q_n^I]^{-1}$ avec les initialisations :

$$\hat{X}_0^- = \bar{X}_0 - \bar{x}_0 = \mathbb{E}[X_0] - \bar{x}_0, \quad R_0^- = Q_0^X$$

Le gain matriciel de Kalman K_n est à déterminer de façon à fournir une estimation \hat{X}_n optimale au sens de l'erreur quadratique moyenne.

Plus précisément, pour la phase de prédiction on linéarise l'équation d'état autour de \hat{X}_{k-1}

$$\begin{aligned}X_n &= f(X_{n-1}) + g(X_{n-1})W_{n-1} \\ &\simeq f(\hat{X}_{n-1}) + F_{n-1}(X_{n-1} - \hat{X}_{n-1}) + G_{n-1}W_{n-1}\end{aligned}$$

$$F_{n-1} = \vec{\nabla} f(\hat{X}_{n-1}) \text{ et } G_{n-1} = g(\hat{X}_{n-1})$$

et pour la phase de correction, on linéarise l'équation d'observation autour de \hat{X}_n^-

$$\begin{aligned}Y_n &= h(X_n) + V_n \\ &\simeq h(\hat{X}_n^-) + H_n(X_n - \hat{X}_n^-) + V_n\end{aligned}$$

où $H_n = \vec{\nabla} h(\hat{X}_n^-)$

et on a l'algorithme du filtre de Kalman étendu pour le filtrage non linéaire gaussien :

Initialisation

$$X_0 \sim \mathcal{N}(\bar{X}_0, Q_0)$$

Prédiction

$$F_{n-1} = \vec{\nabla} f(\hat{X}_{n-1})$$

$$G_{n-1} = g(\hat{X}_{n-1})$$

$$\hat{X}_n^- = f(\hat{X}_{n-1})$$

$$P_n^- = F_{n-1} P_{n-1} F_{n-1}^t + G_{n-1} Q_{n-1}^W G_{n-1}^t$$

Innovation

$$H_n = \vec{\nabla} h(\hat{X}_n^-)$$

$$Q_n^Z = H_n P_n^- H_n^t + Q_n^V$$

Gain de Kalman

$$K_n = P_n^- H_n^t [Q_n^Z]^{-1}$$

Correction

$$\hat{X}_n = \hat{X}_n^- + K_n [Y_n - h(\hat{X}_n^-)]$$

$$P_n = [I - K_n H_n] P_n^-$$

On définit ainsi l'erreur à posteriori : $e_n = X_n - \hat{X}_n$
avec sa matrice de covariance $p_n = \mathbb{E}[e_n e_n^t]$.

De même on définit l'erreur à priori $e_n^- = X_n - \hat{X}_n^-$,
ainsi que la matrice de covariance de l'erreur à priori : $p_n^- = \mathbb{E}[e_n^- (e_n^-)^t]$.

L'application du filtrage de Kalman en système non linéaire gaussien nécessite la dérivabilité locale des fonctions d'état et d'observation.

L'approche par filtrage particulaire peut résoudre ce problème mais avec des hypothèses plus précises.

Le filtrage particulaire est une méthode basée sur la théorie des mesures de probabilité. Sa convergence dépend des lois classiques des grands nombres et du théorème de la limite centrale.

Comme dans le filtrage de Kalman, les phases de prédiction et correction apparaissent encore dans le filtrage particulaire mais avec les phases de mutation et pondération et en général avec une phase de sélection.

2.3 Filtrage particulière

Le filtrage particulière est une méthode numérique de filtrage adaptée aux cas de modèles non linéaires munis des bruits non-gaussiens.

Il est parfois difficile, voire impossible d'établir explicitement les filtres optimaux.

L'idée du filtrage particulière [BAEHR,2009], [BEROZZI,2003] est d'approximer à chaque instant n , le filtre $p_k(x)$ par une suite de loi discrète $p_k^N(x) = \sum_{i=1}^N \omega_k^i \delta_{\xi_k^i}(x)$, où δ est la mesure de Dirac.

Dans cette méthode, dite méthode de Monte Carlo séquentielle, les étapes de prédiction et correction sont représentées par des systèmes de particules ξ_k^i où $\xi_k^i \sim p_k(x)$. À chaque simulation, $p_k^N(\xi_k^i) = \omega_k^i$ est le poids de la particule ξ_k^i .

Il se peut que dans la simulation, des particules ont des poids ω_k^i très faibles ou presque nuls et ne peuvent plus contribuer à l'approximation. Cette dégénérescence de particules implique une redistribution. Les particules de poids négligeables devront être remplacées par duplication des particules de poids plus importants non négligeables. On procède alors à un regroupement de particules dans une région d'état pour ralentir la dégénérescence. Une redistribution des particules de poids plus importants est nécessaire pour compléter le nombre total des particules. Cette redistribution ne doit pas se faire sur un petit nombre de particules sinon on perd l'idée de convergence¹ de la méthode.

On peut considérer un rééchantillonnage multinomial : à partir des particules ξ_k^1, \dots, ξ_k^N avec les probabilités $\omega_k^1, \dots, \omega_k^N$, on fait un échantillonnage de N particules de poids $\frac{1}{N}$. Cela revient à multiplier les particules les plus importantes.

2.3.1 Formules de Bayes

Les formules de Bayes permettent de déterminer les densités conditionnelles en fonction de la densité conjointe et des lois marginales.

Considérons une variable aléatoire X à valeurs dans \mathbb{R}^d et de densité p_X , i.e

$$\mathbb{E}\phi(X) = \int_{\mathbb{R}^d} \phi(x)p_X(x)dx \quad \forall \phi \in \mathcal{C}_b(\mathbb{R}^d)$$

Par définition, la densité conjointe $p_{X,Y}$ de X et Y est telle que $\forall \phi \in \mathcal{C}_b(\mathbb{R}^d \times \mathbb{R}^d)$

$$\mathbb{E}\phi(X, Y) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \phi(x, y)p_{X,Y}(x, y)dxdy$$

1. loi des grands nombres

On définit la densité conditionnelle de X sachant $Y = y$, notée $p_{X|Y=y}$ telle que :
 $\forall \phi \in \mathcal{C}_b(\mathbb{R}^d)$

$$\begin{aligned}\mathbb{E} [\phi(X)|Y = y] &= \int_{\mathbb{R}^d} \phi(x) p_{X|Y=y}(x) dx \\ &= \frac{\int_{\mathbb{R}^d} \phi(x) p_{X,Y}(x, y) dx}{\int_{\mathbb{R}^d} p_{X,Y}(x, y) dy}\end{aligned}$$

Et

$$\begin{aligned}p_{X|Y=y}(x) &:= \frac{p_{X,Y}(x, y)}{p_Y(y)} = \frac{p_{X,Y}(x, y)}{\int_{\mathbb{R}^d} p_{X,Y}(x, y) dx} \\ p_{X|Y=y}(x) &= \frac{p_{Y|X=x}(y) \cdot p_X(x)}{p_Y(y)} = \frac{p_{Y|X=x}(y)}{p_Y(y)} \cdot p_X(x)\end{aligned}\quad (2.16)$$

$$p_{X|Y=y}(x) = \frac{p_{Y|X=x}(y)}{\int_{\mathbb{R}^d} p_{X,Y}(x, y) dy} \cdot \int_{\mathbb{R}^d} p_{X,Y}(x, y) dy \quad (2.17)$$

2.3.2 Echantillonnage d'importance

L'échantillonnage d'importance est une étape nécessaire dans le cas où la loi est difficilement simulable voire impossible. Une autre loi connue et qui est plus facile à manier, intervient dans la simulation. Le choix de la loi de remplacement, en général, est justifié par le fait d'avoir une variance plus petite.

Considérons une loi $p(x)$ absolument continue par rapport à une autre loi $\tilde{p}(x)$
 $p(x) = 0 \Rightarrow \tilde{p}(x) = 0$.

$$\mathbb{E}\varphi(X) = \int_{\mathbb{R}^d} \varphi(x) p(x) dx = \int_{\mathbb{R}^d} \varphi(x) \frac{p(x)}{\tilde{p}(x)} \tilde{p}(x) dx$$

Si on simule un N -échantillon $\tilde{\xi}^1, \dots, \tilde{\xi}^N$ à partir de la loi $\tilde{p}(x)$, alors on peut avoir l'approximation par la méthode de Monte Carlo

$$\mathbb{E}\varphi(X) \simeq \sum_{i=1}^N \omega^i \varphi(\tilde{\xi}^i)$$

où $\tilde{\xi}^i \sim \tilde{p}(x)$ et $\omega^i = p(\tilde{\xi}^i)/\tilde{p}(\tilde{\xi}^i)$.

Dans ce contexte, si on dispose d'un N -échantillon $(\xi^i, 1 \leq i \leq N)$ de la loi à priori p_X , avant toute observation, on a l'approximation :

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

où $\xi^i \sim p_X$, ($\forall 1 \leq i \leq N$)

Après observation de Y , d'après la formule de Bayes (2.17), on obtient une approximation de la loi à posteriori telle que :

$$p_{X|Y=y}(x) \simeq p_{X|Y}^N(x) = \sum_{i=1}^N \omega^i \delta_{\xi^i}(x) \quad (2.18)$$

$$\text{où } \omega^i = \frac{p_{Y|X=\xi^i}(y)}{\sum_{j=1}^N p_{Y|X=\xi^j}(y)}, \quad (\forall 1 \leq i \leq N)$$

2.3.3 Filtre particulière ou bootstrap

La base du filtrage particulière [DEL MORAL et al.,2006] est de déterminer des approximations du filtre prédit et du filtre corrigé, sous la forme :

$$\mu_{n-}^-(dx) = \mathbb{P}_{X_n}(dx|\mathcal{Y}_{n-1}) \simeq \sum_{i=1}^N \omega_{n-}^i \times \delta_{\xi_{n-}^i}(x) dx$$

$$\mu_n(dx) = \mathbb{P}_{X_n}(dx|\mathcal{Y}_n) \simeq \sum_{i=1}^N \omega_n^i \times \delta_{\xi_n^i}(x) dx$$

À chaque instant k , on détermine des suites $(\omega_{n-}^i, \xi_{n-}^i)$ et (ω_n^i, ξ_n^i) , $i = 1, \dots, N$, par échantillonnage d'importance.

$$\text{Avec le modèle : } \begin{cases} Y_n &= h(X_n) + V_n \\ X_n &= f(X_{n-1}, W_n) \end{cases}$$

$X_0 \sim \mu_0(dx)$, $W_n \sim p_W(w)dw$ et $V_n \sim p_V(v)dv$ et notons les densités prédite et corrigée par :

$$p_{n-}(x) = \sum_{i=1}^N \omega_{n-}^i \times \delta_{\xi_{n-}^i}(x)$$

$$p_n(x) = \sum_{i=1}^N \omega_n^i \times \delta_{\xi_n^i}(x)$$

En utilisant la formule récurrente du modèle d'état, on obtient le filtre SIS² :

Filtre SIS

Dans l'étape de **prédiction**, les particules sont obtenues directement sans difficultés.

$$\xi_k^i = f(\xi_{k-1}^i, w_k^i) \text{ avec } w_k^i \sim p_W(w)dw$$

L'étape de **correction** correspond à une normalisation avec une introduction de l'observation $Y_n = y_n$:

$$\omega_n^i = \frac{\omega_{n-1}^i p_V(y_n - h(\xi_n^i))}{\sum_{j=1}^N \omega_{n-1}^j p_V(y_n - h(\xi_n^j))}$$

En général, les poids sont assez faibles et les particules ne peuvent plus contribuer à l'approximation.

Une redistribution des particules est nécessaire avant la prédiction : c'est le principe du filtre SIR³.

Filtre SIR

Avec les particules obtenues, on passe à une étape de sélection, c'est à dire avec un **rééchantillonnage**, on choisit $\hat{\xi}_{n-1}^i$ parmi $(\xi_{n-1}^1, \dots, \xi_{n-1}^N)$ en fonction des poids $(\omega_{n-1}^1, \dots, \omega_{n-1}^N)$.

Cette étape de sélection est conditionnée par le paramètre

$$N_n^{eff} = \frac{1}{\sum_{i=1}^N (\omega_n^i)^2} \in [1, N]$$

qui représente le nombre efficace des particules.

Quand N_n^{eff} est proche de N , alors les particules sont d'égale importance. En général, ce n'est pas le cas et on doit redistribuer les particules pour avoir des poids très proches de $\frac{1}{N}$.

La redistribution se fait selon la loi multinomiale qui élimine les particules de poids négligeables en faveur de celles qui apportent plus de contributions dans l'approximation.

La **prédiction** est une étape de mutation,

$$\xi_n^i = f(\hat{\xi}_{n-1}^i, w_n^i) \text{ avec } w_n^i \sim p_W(w)dw$$

$\hat{\xi}_{n-1}^i$ sont les particules sélectionnées après l'observation de $Y_{n-1} = y_{n-1}$.

La **correction** est une étape de pondération en introduisant l'observation $Y_n = y_n$, telle que

$$\omega_n^i = \frac{p_V(y_n - h(\xi_n^i))}{\sum_{j=1}^N p_V(y_n - h(\xi_n^j))}$$

La dégénérescence des poids peut être résolue par une redistribution de Kitagawa, consistant à construire une suite arithmétique de raison $1/N$ de N nombres telle que le premier terme est simulé selon la loi uniforme $\mathcal{U}[0, \frac{1}{N}]$.

Cet algorithme a une complexité de l'ordre de N mais n'est pas toujours efficace la plupart du temps ; les particules peuvent s'éloigner et sortent dans la zone d'intérêt.

2.3.4 Résolution par filtrage particulière

Considérons le système d'équations stochastiques :

$$\begin{cases} X_t &= f_t(X_{t-1}, u_t, W_t) \\ Y_t &= h_t(X_t, u_t, V_t) \end{cases} \quad (2.19)$$

- X_t variables d'état
- u_t variables de contrôle déterministes dans un compact C_u
- Y_t variables observées
- W_t bruit blanc de mesure d'état de densité p_W
- V_t bruit blanc de mesure d'observation de densité p_V

Considérons le système équivalent :

$$\begin{cases} X_t &= \tilde{f}_t(X_{t-1}, W_t) \\ Y_t &= \tilde{h}_t(X_t, V_t) \end{cases} \quad (2.20)$$

où $f_t(\cdot, u_t, \cdot) = \tilde{f}_t(\cdot, \cdot)$ la fonction dans le modèle d'état et $h_t(\cdot, u_t, \cdot) = \tilde{h}_t(\cdot, \cdot)$ dans le modèle d'observation.

On suppose que le bruit d'observation est additif c'est-à-dire le le modèle d'observation est :

$$Y_t = \tilde{h}_t(X_t) + V_t$$

Dans l'étape de mutation : $\xi_n^i = \tilde{f}(\hat{\xi}_{n-1}^i) + w_n^i = f(\tilde{\xi}_{n-1}, u_{n-1}) + w_n^i$

et le filtre prédit est approximé par : $p_n^-(x) = \sum_{i=1}^N \omega_{n-}^i \times \delta_{\xi_{n-}^i}(x)$

L'étape de correction est : $\omega_n^i = \frac{p_V(y_n - \tilde{h}(\xi_n^i))}{\sum_{j=1}^n p_V(y_n - \tilde{h}(\xi_n^j))} = \frac{p_V(y_n - h(\xi_n^i, u_n))}{\sum_{j=1}^n p_V(y_n - h(\xi_n^j, u_n))}$

En calculant le nombre de particules efficaces : $N_n^{eff} = \frac{1}{\sum_{i=1}^N (\omega_n^i)}$

on passe à l'étape de redistribution des particules pesantes : $\hat{\xi}_n^i$

Le filtre optimal est alors approximé par : $p_n(x) = \sum_{i=1}^N \omega_n^i \times \delta_{\hat{\xi}_n^i}(x)$

Une étape de régularisation de la densité du filtre donne :

$$p_n^N(x|\mathcal{Y}_n) \simeq \frac{1}{N} \sum_{i=1}^N K_h(x - \hat{\xi}_n^i)$$

La commande u_n utilisée à l'instant n est déterminée dans l'équation : $y_n = h(\hat{x}_n, u_n)$.

$$\hat{x}_n = \mathbb{E}(X_n|\mathcal{Y}_n) = \int_{R^d} x.p_n^N(x).dx$$

est l'état estimé par le filtrage par noyau de convolution.

Si les bruits sont trop faibles, le filtrage particulaire peut diverger. [ROSSI et VILA,2003]. Une régularisation des filtres par la méthode des noyaux de convolution est une étape nécessaire pour mieux contrôler les particules. Cette méthode est aussi appliquée pour réduire les erreurs d'estimation des paramètres.

2.4 Filtrage par noyau de convolution

Lorsque les bruits d'observation et d'état sont faibles, il y a dégénérescence des particules et le filtrage particulaire diverge. Le filtrage par noyau de convolution consiste à régulariser le filtre estimé obtenu des particules par une densité continue et permet d'éviter la divergence du filtrage.

Si X_n est la variable d'état non observable à l'instant n et $(Y_1 = y_1, \dots, Y_n = y_n)$ la suite de variables observées suivant la loi μ_n^Y , on considère la variable $Z_n = (X_n, Y_1, \dots, Y_n)$ suivant une loi μ_n^Z .

Si les lois sont à densité telles que $\mu_n^Z(dz) = p_n^Z(z)dz$ et $\mu_n^Y(dy) = p_n^Y(y)dy$, par définition la densité du filtre optimal, d'après la formule de Bayes, est telle que :

$$p_n(x) = \frac{p_n^Z(x, y_1, \dots, y_n)}{p_n^Y(y_1, \dots, y_n)}$$

On estime la densité du filtre par la méthode de l'histogramme. On construit une estimation de la densité à partir de l'approximation discrète de la mesure. Cette régularisation est le principe du filtrage par noyau de convolution [ROSSI et VILLA, 2004].

La méthode de Parzen est une généralisation de la méthode d'estimation par histogramme des densités.

Dans l'application du filtrage particulaire, une étape de régularisation est introduite soit après l'échantillonnage du filtre optimal, le filtrage par noyau de convolution post-régularisé, soit après l'échantillonnage du filtre prédit, le filtrage par noyau de convolution pré-régularisé.

Le filtrage par noyau de convolution [CHEN et al.,2012] est aussi le mieux adapté pour l'estimation de paramètres inconnus dans le modèle d'état.

2.4.1 Noyau de convolution

Soit Ω un sous ensemble ouvert de \mathbb{R} et soit f une application intégrable de Ω dans \mathbb{R} associée à une distribution et κ une fonction à support compact de Ω dans \mathbb{R} .

Le **produit de convolution** de f et κ est une fonction $\hat{f}_\kappa(x)$ de Ω dans \mathbb{R} telle que :

$$\hat{f}_\kappa(x) = \int_{\Omega} f(\omega)\kappa(x - \omega)d\omega \quad (2.21)$$

En notant κ_x , la **fonction translatée** de κ en x ,

$$(\forall \omega \in \Omega), \kappa_x(\omega) = \kappa(x - \omega), \quad (2.22)$$

Et on a,

$$\hat{f}_\kappa(x) = \langle f, \kappa_x \rangle = \int_{\Omega} f(\omega) \kappa_x(\omega) d\omega \quad (2.23)$$

Par définition, **une approximation de l'unité** est une suite (φ_k) dans $L^1(\mathbb{R}^d)$ telle que

$$\begin{aligned} \sup_{k \geq 1} \int_{\mathbb{R}^d} |\varphi_k(x)| dx &< +\infty \\ \int_{\mathbb{R}^d} \varphi_k(x) dx &= 1 \text{ pour tout } k \geq 1, \\ (\forall \alpha > 0), \lim_{k \geq 1} \int_{\{y: |y| > \alpha\}} |\varphi_k(x)| dx &= 0 \end{aligned}$$

Proposition 7 *La suite (φ_k) est une approximation de l'unité telle que :*

- *Pour toute fonction mesurable bornée g et pour tout point de continuité x_0 de g , la suite $(\varphi_k * g)(x_0)$ converge vers $g(x_0)$.*
- *pour toute fonction uniformément continue bornée g , la suite $(\varphi_k * g)$ converge uniformément vers g .*
- *Pour toute fonction $g \in L^p, 1 \leq p < \infty$, la suite $(\varphi_k * g)$ converge vers g dans L^p .*

Preuve

Soit g une fonction mesurable bornée,

$$d_k(x_0) = (\varphi_k * g)(x_0) - g(x_0) = \int_{\mathbb{R}^d} [g(x_0 - y) - g(x_0)] \varphi_k(y) dy$$

Pour construire des approximations de l'unité, on se donne une fonction φ intégrable

sur \mathbb{R}^d telle que $\int_{\mathbb{R}^d} \varphi(x) dx = 1$ et on considère la suite (φ_k) définie par :

$$\varphi_k(x) = k^d \varphi(kx) \quad k \geq 1.$$

Avec le changement de variable $y = kx$, on obtient les propriétés de l'approximation de l'unité.

On peut donc construire une approximation de l'unité par d'après proposition suivante :

Proposition 8 *Si φ est une fonction continue positive à support compact sur \mathbb{R}^d telle que :*

$$x \neq 0 \Rightarrow \varphi(x) < \varphi(0),$$

la suite φ_n telle que

$$\varphi_n(x) = \left(\int_{\mathbb{R}^d} \varphi(y)^n dy \right)^{-1} \cdot \varphi(x)^n$$

est une approximation de l'unité.

Par définition, un **noyau** K est une application de $\mathbb{R}^d \rightarrow \mathbb{R}$, bornée, positive, symétrique, intégrable par rapport à la mesure de Lebesgue et d'intégrale 1.

Un noyau de **Parzen-Rosenblatt** est un noyau vérifiant :

$$\lim_{\|x\| \rightarrow \infty} \|x\|^d K(x) = 0$$

Lemme 9 *L'estimateur f_n associé à K de la densité f des variables aléatoires X_1, \dots, X_n indépendantes identiquement distribuées, est*

$$f_n(x) = \frac{1}{nh_n^d} \sum_{t=1}^n K \left[\frac{x - X_t}{h_n} \right] = (K_{h_n} * \mu_n)(x), \quad x \in \mathbb{R}^d$$

h_n est le paramètre de lissage lié à n .

$\mu_n = \frac{1}{n} \sum_{t=1}^n \delta_{X_t}$ est la densité empirique obtenue en régularisant la mesure empirique des X_1, \dots, X_n par convolution avec $\frac{1}{h_n^d} K \left(\frac{\cdot}{h_n} \right)$.

$$K_{h_n}(y) = \frac{1}{h_n^d} K \left(\frac{y}{h_n} \right)$$

Preuve

De l'échantillon (X_1, \dots, X_n) , on régularise la mesure empirique représentée par un histogramme par la densité :

$$\mu_n = \sum_{i=1}^n \delta_{X_i}$$

Si K est un noyau de Parzen-Rosenblatt associé à la densité à estimer f , notons, $K_{h_n}(y) = \frac{1}{h_n^d} K \left(\frac{y}{h_n} \right)$. Le produit de convolution de K_{h_n} par μ_n est

$$\begin{aligned} (K_{h_n} * \mu_n)(x) &= \int_{\mathbb{R}^d} \frac{1}{h_n^d} K \left(\frac{x - y}{h_n} \right) \sum_{i=1}^n \delta_{X_i}(y) dy \\ &= \frac{1}{nh_n^d} \sum_{t=1}^n K \left[\frac{x - X_t}{h_n} \right] \end{aligned}$$

Noyaux usuels

Les noyaux généralement utilisés sont obtenus dans la théorie d'approximation numérique de fonction.

Noyaux polynomiaux

- $K(y) = P(y)$ pour $|y| < C$, où P est un polynôme, $K(y) = 0$ pour $|y| \geq C$
- noyau unité $K(y) = 1$, si $|y| < 1$
- noyau d'Epanechnikov [DEHEUVELS, 1977] $K(y) = 1 - 1.5y^2$, si $|y| < 1$
- noyau de Legendre d'ordre 1 : $K(y) = \frac{3}{8}(3 - 5y^2)$, $|y| < 1$
- noyau de Legendre d'ordre 2 : $K(y) = \frac{15}{128}(15 - 70y^2 + 63y^4)$, si $|y| < 1$

Noyaux de Gram-Charlier

- En général, $K(y) = P(y)e^{-\frac{1}{2}y^2}$, P étant un polynôme.
- noyau normal : $K(y) = (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}y^2}$
 - noyau de Gram-Charlier d'ordre 1 : $K(y) = \frac{1}{2}(3 - y^2)(2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}y^2}$
 - noyau de Gram-Charlier d'ordre 2 : $K(y) = \frac{1}{8}(y^4 - 10y^2 + 15)(2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}y^2}$

Noyaux de Laguerre

- En général, $K(y) = P(|y|)\exp(-|y|)$, P étant un polynôme,
- noyau de Picard : $K(y) = \frac{1}{2}\exp(-|y|)$
 - noyau de Laguerre d'ordre 1 : $K(y) = \frac{1}{2}\left[\frac{1}{2}y^2 - 3|y| + 3\right]\exp(-|y|)$

Noyaux divers

- noyau de Cauchy d'ordre r : $K(y) = \frac{2^{2r-2}\Gamma(r)^2}{\pi(2r-1)(1+y^2)^r}$
- noyau triangulaire : $K(y) = 1 - |y|$
- noyau de Fourier : $K(y) = \pi^{-1}\left(\frac{1}{y}\sin y\right)$
- noyau cosinus : $K(y) = \cos y$, $|y| < \frac{1}{2}\pi$

2.4.2 Estimation par noyau de convolution

L'histogramme est considéré comme un estimateur non-paramétrique de densité [BILAL,2010]. On partitionne l'intervalle de référence en N -intervalles I_k , $k \in \{1, \dots, N\}$ et on compte le nombre d'observations dans I_k . L'histogramme est régulier si les intervalles ont la même

amplitude. Cette amplitude est appelé la granularité h ou le pas de l'histogramme.

Quand le nombre d'observations tend vers l'infini et quand la granularité tend vers 0, on peut construire l'estimateur d'une densité de probabilité f telle que :

$$\hat{f}_h^n(x) = \frac{1}{h} \sum_{k=1}^N \mathbb{P}(I_k) \mathbb{I}_{I_k}(x).$$

où \mathbb{I}_{I_k} est la fonction caractéristique de l'intervalle I_k .

Dans un histogramme, la densité en un point x est donc estimée par la proportion d'observations x_1, \dots, x_N dans le voisinage de x . Un rectangle ou une fenêtre dans l'histogramme en x a une largeur déterminée selon un paramètre de lissage h ou largeur de la fenêtre.

Par construction, l'estimation est discontinue. La méthode du noyau consiste à remplacer le rectangle centré en x et de largeur h par une courbe gaussienne centrée en x et est adaptée pour une estimation continue. Les valeurs numériques correspondantes aux observations par la courbe sont fonctions de leurs proximités par rapport à x . Finalement, l'estimateur est obtenu par la moyenne de courbes gaussiennes.

L'estimateur à noyau converge plus vite que d'autres estimateurs. En effet, la vitesse de convergence $n^{-1/5}$ est plus faible que la vitesse typique des méthodes paramétriques, généralement n^{-1} .

La méthode de Parzen Rozenblatt [STRAUSS,2008] consiste à construire une estimation de la densité de probabilité d'une variable aléatoire X à partir d'un échantillon de n observations indépendantes et identiquement distribuées (x_1, \dots, x_n) . Un voisinage pondéré sous la forme d'une fonctionnelle intégrable à 1 est nécessaire dans la construction de l'estimateur en tout point de Ω .

Soit κ est un noyau à valeurs dans \mathbb{R}^+ définie sur Ω vérifiant :

$$\int_{\Omega} \kappa(\omega) d\omega = 1$$

À partir de κ , on peut définir un noyau κ_h^x translaté en $x \in \Omega$ et dilaté d'une largeur de bande (ou paramètre de lissage) $h > 0$, par :

$$(\forall u \in \Omega), \kappa_h^x(u) = \frac{1}{h} \kappa\left(\frac{u-x}{h}\right)$$

Par convention, $\kappa(u) = \kappa_1^0(u)$.

L'estimation \hat{f}_κ est donnée en tout point de $x \in \Omega$ par :

$$\hat{f}_\kappa(x) = \frac{1}{n} \sum_{i=1}^n \kappa(x - x_i) \tag{2.24}$$

La plupart des noyaux couramment utilisés en analyse fonctionnelle sont monomodaux, symétriques et centrés.

On a aussi,

$$\hat{f}_\kappa(x) = \frac{1}{n} \sum_{i=1}^n \kappa_x(x_i)$$

L'estimation \hat{f}_κ en chaque point $x \in \Omega$ peut être réécrite comme le produit scalaire du noyau κ_x avec la mesure empirique e_n :

$$\hat{f}_\kappa(x) = \langle e_n, \kappa_x \rangle$$

avec

$$e_n(x) = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}(x)$$

où δ_{x_i} est l'impulsion de Dirac translatée en x_i .

En effet,

$$\begin{aligned} \langle e_n, \kappa_x \rangle &= \int_{\Omega} \frac{1}{n} \sum_{i=1}^n \delta_{x_i}(\omega) \kappa_x(\omega) d\omega \\ &= \frac{1}{n} \sum_{i=1}^n \int_{\Omega} \delta_{x_i}(\omega) \kappa_x(\omega) d\omega \\ &= \frac{1}{n} \sum_{i=1}^n \kappa_x(x_i) = \hat{f}_\kappa(x). \end{aligned}$$

L'utilisation pratique de cette méthode requiert deux choses :

- le noyau K (généralement la densité d'une loi statistique),
- le paramètre de lissage h .

On peut supposer que l'échantillon est distribué selon la loi normale $\mathcal{N}(\mu; \sigma^2)$ et prenons

$$h = 1.06 \hat{\sigma}_n^{-1/5}.$$

Une autre façon d'opérer est de chercher à fixer h de manière optimale. Prenons

$$R(f, \hat{f}(x)) \approx \frac{1}{4} \sigma_K^4 h^4 \int_{\Omega} (f''(x))^2 dx + \frac{\int_{\Omega} K^2(x) dx}{nh}$$

où

$$\sigma_K^2 = \int_{\Omega} x^2 K(x) dx$$

La fenêtre optimale est obtenue en minimisant la fonction de risque $R(f, \hat{f}(x))$ et vaut :

$$h^* = \frac{c_1^{-2/5} c_2^{1/5} c_3^{-1/5}}{n^{1/5}}$$

où

$$c_1 = \int_{\Omega} x^2 K(x) dx, \quad c_2 = \int_{\Omega} K(x)^2 dx, \quad c_3 = \int_{\Omega} (f''(x))^2 dx.$$

Le paramètre h est toujours proportionnel à $n^{-1/5}$.

2.4.3 Filtrage par noyau de convolution

En considérant le système d'équation d'état et d'équation d'observation,

$$\begin{cases} X_{n+1} = f_n(X_n, W_n) \\ Y_n = h_n(X_n, V_n) \end{cases} \quad (2.25)$$

f_n est la fonction d'évolution du modèle d'état.

h_n est la fonction d'observation,

W_n est un bruit blanc dans le modèle d'état,

V_n est un bruit blanc dans le modèle d'observation,

$\mu_0(dx)$ est la loi de distribution de l'état initial X_0 .

Dans l'application de la méthode de filtrage par noyau de convolution, on a besoin de pouvoir :

- générer X_0 par la loi initiale μ_0 ,
- générer X_{n+1} sachant X_n ,
- générer Y_n sachant X_n .

On génère d'abord M - particules de la loi initiale :

$$\tilde{\xi}_0^1, \dots, \tilde{\xi}_0^M \sim \mu_0.$$

Après, les particules évoluent suivant le système (2.25) avec la loi μ_1 de l'état X_1 en :

$$\tilde{\xi}_1^1, \dots, \tilde{\xi}_1^M \sim \mu_1.$$

Plus généralement, du filtre optimal $\mu_{n-1}(dx)$ à l'instant $n-1$, on obtient par échantillonnage la mesure $\frac{1}{M} \sum_{i=1}^M \delta_{\tilde{\xi}_{n-1}^i}(dx)$ approximative particulaire. On construit le filtre

prédit par évolution des particules à partir du système (2.25) tel que $\mu_n^-(dx) = \frac{1}{M} \sum_{i=1}^M \delta_{\tilde{\xi}_{n-1}^i}(dx)$.

Une régularisation de la mesure prédite par un noyau de convolution entraîne la mesure $\nu_n(dx) = K_h * \mu_n^-(dx)$.

Le filtre optimal à l'instant n est obtenue en corrigeant $\nu_n(dx)$ par normalisation telle que : $\mu_n(dx) = \Psi_n \cdot \nu_n(dx)$.

Remarque

L'étape de régularisation peut être considérée après l'échantillonnage du filtre optimal dans le cas de post-régularisation.

Une autre approche de filtrage par noyau de convolution est d'estimer la densité conjointe $p_{X_n, Y_1, \dots, Y_n}(x, y_1, \dots, y_n)$ et la densité marginale des observations $p_{Y_1, \dots, Y_n}(y_1, \dots, y_n)$ par des noyaux de Panzen-Rosenblatt.

Par définition, la densité conditionnelle de X_n sachant $Y_1 = y_1, \dots, Y_n = y_n$ est telle que

$$p_{X_n}(x|Y_1 = y_1, \dots, Y_n = y_n) = \frac{p_{X_n, Y_1, \dots, Y_n}(x, y_1, \dots, y_n)}{p_{Y_1, \dots, Y_n}(y_1, \dots, y_n)}$$

Les principes fondamentaux dans ce filtrage par noyau de convolution sont :

- l'estimation de p_{X_n, Y_1, \dots, Y_n} et de p_{Y_1, \dots, Y_n} par la méthode des noyaux de convolution.
- la construction des estimateurs par la générations de $(\tilde{x}_n, \tilde{y}_n)$ à partir du système (2.25).

On commence par générer les M -états initiaux \tilde{x}_0^i suivant la loi μ_0 .
Les M -états initiaux évoluent suivant l'équation d'état dans (2.25) en \tilde{x}_1^i .

On génère ensuite les M -observations \tilde{y}_1^i suivant l'équation d'observation dans (2.25).
On obtient alors M -états dans la loi conjointe :

$$\tilde{z}_n^1, \dots, \tilde{z}_n^M, \text{ avec } \tilde{z}_n^i = (\tilde{x}_n^i, \tilde{y}_n^i).$$

On construit alors une approximation particulière de la mesure de la loi conjointe :

$$\mu_{XY}(dx) = \frac{1}{M} \sum_{i=1}^M \delta_{\tilde{z}_n^i}(dx)$$

Cette mesure empirique est convoluée avec un noyau L et on a l'estimation de la densité conjointe :

$$p_{XY}^n(z) = \frac{1}{M} \sum_{i=1}^M L_{h_M}(z - \tilde{z}_n^i)$$

De même pour l'estimation de la densité marginale par convolution à un noyau K est

$$p_Y^n(y) = \frac{1}{M} \sum_{i=1}^M K_{h_M}(y - \tilde{y}_n^i)$$

Le filtre optimal $p_{X_n}(x|\mathcal{Y}_n)$ est donc estimé par :

$$p_n(x|y_1, \dots, y_n) = \frac{p_{XY}^n(z)}{p_Y^n(y_1, \dots, y_n)} = \frac{\sum_{i=1}^M L_{h_M}(z - \tilde{z}_n^i)}{\sum_{i=1}^M K_{h_M}(y_1 - \tilde{y}_1^i, \dots, y_n - \tilde{y}_n^i)}$$

avec $z - \tilde{z}_n^i = (x - \tilde{x}_n^i, y_i - \tilde{y}_n^i)$

2.4.4 Resolution par filtrage par noyau

Le filtrage à noyaux de convolution [CHEN et al.,2012], [VILA et al,2009] permet d'estimer simultanément les paramètres et les états non observés du système dynamique en temps réel en fonction des observations.

Considérons le système dynamique :

$$\begin{cases} X_{n+1} &= f_n(X_n, \theta_x, W_n) \\ Y_n &= h_n(X_n, \theta_y, V_n) \end{cases} \quad (2.26)$$

f_n est la fonction d'évolution du modèle d'état.

h_n est la fonction d'observation,

W_n est un bruit blanc dans le modèle d'état de densité f_W ,

V_n est un bruit blanc dans le modèle d'observation de densité f_V ,

On veut estimer à l'instant n , le paramètre $\theta = (\theta_x, \theta_y)$ et l'état x_n de X_n au vu des observations $Y_1 = y_1, \dots, Y_n = y_n$.

À l'instant n , on obtient les bruits et les particules avec :

pour le bruit d'état, $w_n \sim f_W(w)dw$

pour le bruit d'observation, $v_n \sim f_V(v)dv$

et les particules de la loi dynamique de l'état obtenues par l'équation d'évolution

$$\bar{x}_n^i = f_n(\bar{x}_n^i, \bar{\theta}_{n-1}^i, w_{n-1})$$

de même, les particules obtenues par l'équation d'observation

$$\bar{y}_n^i = h_n(\bar{x}_n^i, \bar{\theta}_{n-1}^i, v_n)$$

On traite les paramètres comme dans les variables d'état $x_n^\theta = (x_n, \theta_n)$.

$$\hat{p}_n^M(x^\theta | \mathcal{Y}_n) = \hat{p}_n^M(x, \theta | \mathcal{Y}_n) = \frac{\sum_{i=1}^M K_{h_M}^y(\bar{y}_n^i - y_n) \times K_{h_M}^\theta(\bar{\theta}_n^i - \theta) \times K_{h_M}^x(\bar{x}_n^i - x)}{\sum_{i=1}^M K_{h_M}(\bar{y}_n^i - y_n)} \quad (2.27)$$

On dispose alors,

- d'une loi à priori sur les paramètres $p_\theta(t)$,
- d'une loi d'évolution des paramètres, $\theta_{n+1} = \theta_n$,
- d'estimation de la densité $p_{X_n}(x | y_1, \dots, y_n)$ et

$$p_{X_n}^M(x | y_1, \dots, y_n) = \frac{\sum_{i=1}^M K_{h_M}^y(y_n - \tilde{y}_n^i) \times K_{h_M}^x(x - \tilde{x}_n^i)}{\sum_{i=1}^M K_{h_M}^y(y_n - \tilde{y}_n^i)} \quad (2.28)$$

– d'estimation de la densité $p_\theta(t|y_1, \dots, y_n)$

$$p_{\theta_n}^M(t|y_1, \dots, y_n) = \frac{\sum_{i=1}^M K_{h_M}^y(y_n - \tilde{y}_n^i) \times K_{h_M}^\theta(t - \tilde{\theta}_n^i)}{\sum_{i=1}^M K_{h_M}^y(y_n - \tilde{y}_n^i)} \quad (2.29)$$

$K_{h_M}^y$ est un noyau de Parzen-Rosenblatt relatif à l'estimation de Y , $K_{h_M}^x$ celui relatif à l'estimation de X et $K_{h_M}^\theta$ celui relatif à l'estimation de θ .

$\hat{p}(x_{n+1}^\theta, y_{n+1}|y_{0:n})$ est un estimateur à noyau de $p(x_{n+1}^\theta, y_{n+1}|y_{0:n})$.

Les particules sont initialisées à partir d'une loi à priori de densité $p^\theta(x)$.
 À l'étape $n, n \geq 1$, on extrait M -particules selon la loi de densité $\hat{p}_n^\theta(x|\mathcal{Y}_n)$.
 À travers l'équation d'évolution du système, on propage les particules.

L'estimé du paramètre θ est tel que :

$$\hat{\theta}_n = \mathbb{E}(\theta|\mathcal{Y}_n) = \int_{\mathbb{R}} t \cdot p_{\theta_0}^M(t|y_1, \dots, y_n) dt$$

Finalement, le filtrage par noyau de convolution n'est tout simplement qu'une régularisation des filtres particuliers.

2.5 Conclusion

Les différentes méthodes exposées ont chacune ses particularités. Leurs adaptations nécessitent des hypothèses bien définies. Il est bien nécessaire de modéliser avant tout le problème de filtrage.

Comme le modèle est un système d'équation d'état et d'équation d'observation, il est naturel de les établir selon les circonstances. Le filtrage de Kalman est une méthode d'estimation paramétriques d'un système dynamique à partir des mesures perturbées. Il a non seulement la capacité de prédire les paramètres mais la capacité de rectifier aussi les erreurs dans modelisation.

Le filtrage de Kalman est bien adaptée au système gaussien. Il est assez facile à appliquer dans la pratique et donne en effet une solution analytique du problème que ce soit linéaire ou non linéaire. Mais la non-linéarité du modèle peut entraîner la multi-modalité de la loi conditionnelle de l'état sachant l'observation et rend le filtre de Kalman inadapté.

De même, si le système est fortement non linéaire, le filtre peut diverger.

Dans le cas non linéaire non nécessairement gaussien, le filtrage particulaire est le mieux adapté. Il consiste à approximer le filtre optimal par une "peigne de Dirac" de la forme $\frac{1}{N} \sum_{i=1}^N \delta_{\xi_i^N}$ où ξ_i^N sont des particules simulées par le filtre même. C'est un filtrage utilisant la simulation de Monte Carlo.

C'est une methode qui repose sur une exploration de l'espace d'état par des particules. Des particules sont conservées et multipliées et d'autres tout simplement éliminées selon leurs contributions dans la construction du filtre.

Une grande qualité de cette methode est sa facilité de mis en œuvre et à son adaptation à presque n'importe bruit d'observation, en particulier avec le bruit d'observation uniforme. Toutefois, cette méthode necessite une connaissance analytique des fonctions de vraisemblance. De plus, elle demande une nombre assez considerable de particules pour qu'on puisse espérer une convergence avec les lois des grands nombres et le théorème central limite.

Avec le filtrage particulaire, on peut introduire des régularisations soit au niveau du filtre prédit et appliquer le théorème de Bayes, soit au niveau du filtre optimal pour résoudre les estimations de paramètres inconnus intervenant dans le modèle. Le filtrage par noyau de convolution est une extension du filtrage particulaire nécessaire dans les cas où les bruits sont trop faibles.

Conclusion de la première partie

cette première partie concerne les méthodes analytiques et numériques de filtrage non linéaires. Les méthodes numériques sont plus exposées. Il s'agit des filtrages de Kalman et des filtrages particulaires. Chacune de ces méthodes demande des hypothèses bien précises mais la mise en œuvre de ces différentes méthodes présente parfois des difficultés.

Le filtrage de Kalman est adapté pour les bruits blancs gaussiens. À l'origine, le filtre de Kalman est une solution pour les systèmes linéaires gaussiens. Une extension avec le filtrage de Kalman étendu demande une linéarisation dans le système mais si le système a une forte non linéarité, le filtrage de Kalman n'est plus adapté.

D'autres extensions de filtrage de Kalman sont proposées. Le filtre de Kalman inodore [FROGERAIS,2008] qui est une extension permettant de se passer des calculs du gradient par des formules de quadrature ou le filtre de Kalman d'ensemble introduisant de l'approximation particulière sur les moments dans l'élaboration du filtre avec les moments empiriques. Cette dernière technique permet de résoudre le problème de la grande dimension des matrices de covariance.

Une autre approche du filtrage de Kalman-Bucy est la résolution d'équations différentielles stochastiques de Riccati de la matrice de covariance du filtre et de l'estimation de sa moyenne [VALADIER,1972]. Des hypothèses de régularité des coefficients sont nécessaires.

Le filtrage particulaire est la méthode la plus utilisée en filtrage non linéaire non nécessairement gaussien. Son application est mis en défaut dans les cas où les bruits sont trop faibles. La régularisation à partir des noyaux de convolution peut résoudre ce problème. Le filtrage particulaire peut être appliqué pour un système non linéaire à bruit d'observation uniforme.

À propos des modèles classiques de problème de filtrage, nous avons développé des méthodes. En particulier, avec un modèle de diffusion, la solution du filtrage est liée à la solution d'une équation aux dérivées partielles. Cette équation aux dérivées partielles est établie en introduisant un générateur infinitésimal.

Le filtrage d'un système non linéaire paramétré est résolu par le filtrage particulaire par noyau de convolution et le système non linéaire contrôlé est résolu par le filtrage particulaire. Des régularisations sont utiles pour avoir plus de précision. Des applications sont exposées à la troisième partie.

Dans la partie suivante, nous allons proposer une théorie sur l'existence du filtre optimal et étudier des cas particuliers de modèles de système non linéaire. Une étude de l'oubli de la condition initiale est détaillée.

Deuxième partie

Études de cas de Filtrage non
linéaire et Études de l'oubli de la
condition initiale

Introduction de la deuxième partie

Cette partie développe le thème principal de la thèse. Elle contient les différents points essentiels dans nos recherches tout en précisant des originalités et innovations :

- Dans ce mémoire, nous avons développé une théorie dans laquelle le théorème de Girsanov permet d'établir une relation de récurrence dans la construction des filtres. Ce développement facilite la mise en œuvre des filtres dans les applications avec un modèle non linéaire où tous les bruits sont gaussiens. Ce développement est une autre approche de construction du filtre optimal qui peut être étendu dans les cas de bruits non gaussiens par approximation.

- L'application de bruit uniforme dans le modèle d'observation présente un cas particulier. En général, l'existence de bruit d'observation uniforme est une conséquence de l'imprécision des mesures. Cette imprécision est a priori connue selon les capteurs utilisés. Le bruit d'observation uniforme est étudié dans le filtrage de Kalman et le filtrage particulière.

- Plus généralement, dans le filtrage non linéaire à bruits quelconques, nous avons distingué en particulier, les bruits d'observation uniformes et les mélanges de bruits d'observation gaussiens. Avec les résultats obtenus, on a pu généraliser ces applications dans des modèles contenant des paramètres inconnus et des commandes.

- Pour les différentes méthodes de filtrage, une caractéristique à étudier est la stabilité. Il s'agit d'oublier la condition initiale définie par la loi de X_0 . En considérant deux lois initiales différentes, on obtient pratiquement les mêmes propriétés des filtres optimaux déduits.

L'oubli de la condition initiale pour chaque type de filtrage demande des hypothèses bien précises. Dans cet oubli, si on utilise une loi initiale erronée, l'impact sur la propriété du filtre n'est pas très significatif. Le quatrième chapitre est principalement consacré à l'étude complète de cet oubli.

Chapitre 3

Filtrage non linéaire en temps discret

Sommaire

3.1 Introduction	59
3.2 Filtrage non linéaire d'un modèle à bruits gaussiens	60
3.3 Filtrage particulière d'un modèle à bruit d'observation uniforme ..	69
3.4 Filtrage par noyau de convolution d'un modèle commandé	71
3.5 Filtrage de Kalman inodore d'un modèle à bruit d'observation quelconque	76
3.7 Conclusion	82

3.1 Introduction

Avec un modèle général, le filtre optimal est explicité mais sa mise en œuvre dans la pratique n'est pas évidente. Nous proposons une solution théorique dans le cas gaussien où le modèle d'état n'est représenté que par un noyau de transition et un bruit additif gaussien.

Avec un modèle à bruits gaussiens, nous avons développé une théorie dans laquelle le théorème de Girsanov [CLIMESCU, 2004], [OUVRARD,1973] permet de construire une suite de mesures de probabilité. Cette suite de mesures permet d'établir de façon récursive le filtre optimal en passant par le filtre prédit.

La théorie utilisée avec le théorème de Girsanov n'est adaptée que dans le cas où tous les bruits sont gaussiens. Dans le cas où le bruit d'observation n'est pas gaussien, en particulier, avec un bruit d'observation uniforme, nous avons appliqué le filtrage particulière [Le GLAND, 2003], [DEL MORAL, 2009]. Dans l'application de ce filtrage, la redistribution des particules n'est plus nécessaire.

En ce qui concerne le filtrage de système paramétré ou de système contrôlé, le filtrage particulière par noyau de convolution [CHEN et al.,2012], [ROSSI et VILLA, 2004], [ROSSI, 2004] est le mieux adapté. Nous avons considéré une grille fixe comme système de particules et les poids dans l'échantillonnage sont tout simplement relatifs aux nombres

d'apparition des particules.

Pour un modèle plus général, le bruit peut être approximé par un mélange de lois gaussiennes [PROSPERI, 1997]. Nous avons appliqué le filtrage de Kalman inodore dans [JULIER et UHLMANN, 2004], [MERWE et WAN, 2001], [COLLIN et al., 2001] pour des modèles à bruits non nécessairement gaussiens.

3.2 Filtrage non linéaire d'un modèle à bruits gaussiens

On considère une chaîne de Markov $X = (X_n)_{n \in \mathbb{N}}$ [BAKRI, 2003] supposée homogène et non observée dans \mathbb{R}^d de loi initiale μ_0 et de matrice de transition π perturbée par un bruit blanc gaussien W_n à chaque instant n . On observe une suite $Y = (Y_n)_{n \in \mathbb{N}}$ à valeurs dans \mathbb{R}^k perturbée par une suite de bruits blancs gaussiens $V = (V_n)_{n \in \mathbb{N}}$.

On dispose d'une :

- loi initiale $\mu_0(dx) = \mathbb{P}(X_0 \in dx) = \mathbb{P}_{X_0}(dx)$, d'une
- loi de transition ou noyau $\pi(x, dx') = \mathbb{P}(X_n \in dx' | X_{n-1} = x)$, $\forall n \geq 1$ et d'une
- loi d'émission $\Gamma_n(x, dy) = \mathbb{P}(Y_n \in dy | X_n = x)$, $\forall n \geq 1$

En effet, à partir de l'équation d'état,

$$X_n = f(X_{n-1}) + W_n \quad (3.1)$$

où $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ une fonction mesurable et $W = \{W_n; n \geq 1\}$ une suite de v.a i.i.d gaussienne additive de loi commune $\mathcal{N}(0, Q^W)$, $Q^W > 0$ de densité :

$$f_W(w) = \frac{1}{(2\pi)^{d/2} |\det Q^W|^{1/2}} \exp \left[-\frac{1}{2} \langle w, (Q^W)^{-1} \cdot w \rangle \right]$$

$$f_W(w) = \frac{1}{(2\pi)^{d/2} |\det Q^W|^{1/2}} \exp \left[-\frac{1}{2} |(Q^W)^{-1/2} \cdot w|^2 \right]$$

et de l'équation d'observation,

$$Y_n = h(X_n) + V_n \quad (3.2)$$

où $h : \mathbb{R}^d \rightarrow \mathbb{R}^k$ une fonction mesurable et $V = \{V_n; n \geq 1\}$ une suite de v.a i.i.d gaussienne additive de loi commune $\mathcal{N}(0, Q^V)$, $Q^V > 0$ de densité :

$$f_V(v) = \frac{1}{(2\pi)^{k/2} |\det Q^V|^{1/2}} \exp \left[-\frac{1}{2} \langle v, (Q^V)^{-1} \cdot v \rangle \right]$$

$$f_V(v) = \frac{1}{(2\pi)^{k/2} |\det Q^V|^{1/2}} \exp \left[-\frac{1}{2} |(Q^V)^{-1/2} \cdot v|^2 \right]$$

on obtient la loi de transition et la loi d'émission par le lemme suivant,

Lemme 10 *Le noyau de transition est tel que :*

$$\pi(x, dx') = f_W(x' - f(x))dx'$$

Et la loi d'émission est telle que :

$$\Gamma(x, dy) = f_V(y - h(x))dx$$

Preuve

Avec l'équation (3.1), pour toute fonction test ϕ mesurable et bornée,

$$\begin{aligned} \mathbb{E}[\phi(X_n)|X_{n-1} = x] &= \mathbb{E}[\phi(f(X_{n-1}) + W_n)|X_{n-1} = x] \\ &= \mathbb{E}[\phi(f(x) + W_n)] \\ &= \int_{\mathbb{R}^d} \phi(f(x) + w)\mathbb{P}(W_n \in dw) \\ &= \int_{\mathbb{R}^d} \phi(f(x) + w)P_{W_n}(dw) \\ &= \int_{\mathbb{R}^d} \phi(f(x) + w)f_W(w)dw \\ &= \int_{\mathbb{R}^d} \phi(x')f_W(x' - f(x))dx' \quad \square \end{aligned}$$

le noyau de transition est : $\pi(x, dx') = f_W(x' - f(x))dx'$

De forme analytique :

$$\begin{aligned} \pi(x, dx') &= \frac{1}{(2\pi)^{d/2}|\det Q^W|^{1/2}} \exp \left[-\frac{1}{2} \langle x' - f(x), (Q^W)^{-1} \cdot (x' - f(x)) \rangle \right] dx' \\ &= \frac{\exp \left(-\frac{1}{2} |x'|^2 \right)}{(2\pi)^{d/2}|\det Q^W|^{1/2}} \exp \left[\langle x', (Q^W)^{-1} \cdot f(x) \rangle - \frac{1}{2} |(Q^W)^{-1/2} \cdot f(x)|^2 \right] dx' \end{aligned}$$

Avec l'équation (3.2), pour toute fonction test ϕ mesurable et bornée,

$$\begin{aligned} \mathbb{E}[\phi(Y_n)|X_n = x] &= \mathbb{E}[\phi(h(X_n) + V_n)|X_n = x] \\ &= \mathbb{E}[\phi(h(x) + V_n)] \\ &= \int_{\mathbb{R}^d} \phi(h(x) + v)\mathbb{P}(V_n \in dv) \\ &= \int_{\mathbb{R}^k} \phi(h(x) + v)P_{V_n}(dw) \\ &= \int_{\mathbb{R}^k} \phi(h(x) + v)f_V(w)dv \\ &= \int_{\mathbb{R}^k} \phi(y)f_V(y - h(x))dy \quad \square \end{aligned}$$

la loi d'émission est : $\Gamma(x, dy) = f_V(y - h(x))dy$

De forme analytique :

$$\begin{aligned}\Gamma(x, dy) &= \frac{1}{(2\pi)^{k/2} |\det Q^V|^{1/2}} \exp \left[-\frac{1}{2} \langle y - h(x), (Q^V)^{-1} \cdot (y - h(x)) \rangle \right] \cdot dy \\ &= \frac{\exp \left(-\frac{1}{2} |y|^2 \right)}{(2\pi)^{d/2} |\det Q^V|^{1/2}} \exp \left[\langle y, (Q^V)^{-1} \cdot h(x) \rangle - \frac{1}{2} |(Q^V)^{-1/2} \cdot h(x)|^2 \right] dy\end{aligned}$$

La loi jointe de (X_0, \dots, X_n) est telle que :

$$\begin{aligned}\mathbb{P}(X_0 \in dx_0, \dots, X_n \in dx_n) &= \mathbb{P}(X_0 \in dx_0) \cdot \mathbb{P}(X_1 \in dx_1 | X_0 = x_0) \dots \\ &\quad \dots \mathbb{P}(X_n \in dx_n | X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \\ &= \int_{\mathbb{R}^d} \dots \int_{\mathbb{R}^d} \mu_0(dx_0) Q(x_0, dx_1) \dots Q(x_{n-1}, dx_n) \\ &= \int_{\mathbb{R}^d} \dots \int_{\mathbb{R}^d} \mu_0(dx_0) f_W(x_1 - f(x_0)) dx_1 \dots \\ &\quad \dots f_W(x_n - f(x_{n-1})) dx_n \\ &= \int_{\mathbb{R}^d} \dots \int_{\mathbb{R}^d} \mu_0(dx_0) f_W(x_1 - f(x_0)) \dots \\ &\quad \dots f_W(x_n - f(x_{n-1})) dx_1 \dots dx_n\end{aligned}$$

Lemme 11 (Cameron Martin) *Si X_1, \dots, X_n variables gaussiennes reduites indépendantes, et comme $X = (X_1, \dots, X_n)$ est un vecteur gaussien où pour tout vecteur $\mu = (\mu_1, \dots, \mu_n)$, $\langle \mu, X \rangle$ est une variable gaussienne et donc*

$$\mathbb{E} \left[\exp \left(\langle \mu, X \rangle - \frac{|\mu|^2}{2} \right) \right] = 1$$

Preuve

Si $X = (X_1, \dots, X_d)$ une v.a à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, sa fonction caractéristique est définie pour $\xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d$ par

$$\Phi_X(\xi) := \mathbb{E}(\exp(i \langle \xi, X \rangle)) = \mathbb{E} \left[\exp i \left(\sum_{j=1}^d \xi_j \cdot X_j \right) \right]$$

□

En particulier, la fonction caractéristique d'un vecteur gaussien X dans \mathbb{R}^d de moyenne μ et de matrice de covariance Q est telle que

$$\Phi_X(\xi) = \exp \left[i \langle \xi, \mu \rangle - \frac{\langle \xi, Q\xi \rangle}{2} \right], \quad \forall \xi \in \mathbb{R}^d$$

Inversement si X est une v.a à valeurs dans \mathbb{R}^d et λ , la mesure de Lebesgue sur \mathbb{R}^d , alors sa densité est donnée par

$$f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \exp(-i \langle \xi, x \rangle) \Phi_X(\xi) d\lambda(\xi)$$

Pour $d = 1$, $\mathbb{E}(X) = i\phi'_X(0)$ et $\mathbb{E}(X^2) = -\phi''_X(0)$.

On a aussi le resultat suivant :

(X^1, \dots, X^n) sont indépendantes si et seulement si $(\forall \xi^i \in \mathbb{R}^d)$

$$\Phi_{(X^1, \dots, X^n)}(\xi^1, \dots, \xi^n) = \prod_{i=1}^n \Phi_{X^i}(\xi^i)$$

La fonction caracteristique d'une variable gaussienne centrée reduite est telle que :

$$\Phi_X(\xi) = \exp \left[-\frac{|\xi|^2}{2} \right]$$

$$\mathbb{E} [\exp \langle \mu, X \rangle] = \mathbb{E} \left[\exp \sum_{i=1}^n \mu_i X_i \right] = \prod_{i=1}^n \mathbb{E} \exp [\mu_i X_i] = \prod_{i=1}^n \exp \left[\frac{\mu_i^2}{2} \right] = \exp \left[\frac{|\mu|^2}{2} \right]$$

et

$$\mathbb{E} \left(\exp \left[-\frac{|\mu|^2}{2} \right] \times \exp (\langle \mu, X \rangle) \right) = \mathbb{E} \left[\exp \left(\langle \mu, X \rangle - \frac{|\mu|^2}{2} \right) \right] = 1$$

$$\text{ou encore } \mathbb{E} \left[\exp \sum_{i=1}^n \left(\mu_i X_i - \frac{\mu_i^2}{2} \right) \right] = \prod_{i=1}^n \mathbb{E} \left[\exp \left(\mu_i X_i - \frac{\mu_i^2}{2} \right) \right] = 1$$

□

Lemme 12 Si Z est une variable aléatoire positive telle que $\mathbb{E}(Z) = 1$, alors on définit une probabilité \mathbb{Q} telle que $\mathbb{Q}(A) = \mathbb{E}(Z\mathbb{I}_A)$ et si X est \mathbb{Q} -intégrable, alors $\mathbb{E}_{\mathbb{Q}}(X) = \mathbb{E}_{\mathbb{P}}(Z.X)$.

Preuve

D'après le lemme de Cameron-Martin et le théorème de Radon-Nikodym comme dans [JEANBLANC et SIMON,2006], on construit une nouvelle probabilité \mathbb{Q} , en posant :

$$\begin{aligned} Z(\omega) &= \exp \left(\langle \mu, X(\omega) \rangle - \frac{|\mu|^2}{2} \right) \\ &= \exp \sum_{i=1}^n \left(\mu_i X_i(\omega) - \frac{\mu_i^2}{2} \right) \\ &= \prod_{i=1}^n \exp \left(\mu_i X_i(\omega) - \frac{\mu_i^2}{2} \right) \end{aligned}$$

$$\frac{\mathbb{Q}(d\omega)}{\mathbb{P}(d\omega)} = Z(\omega) \Leftrightarrow \mathbb{Q}(d\omega) = Z(\omega)\mathbb{P}(d\omega)$$

Ou encore, pour tout ensemble A mesurable,

$$\mathbb{Q}(A) = \mathbb{E}_{\mathbb{P}}(Z\mathbb{I}_A) = \int_A Z(\omega)\mathbb{P}(d\omega)$$

Application au modèle

Supposons que, $V_n \sim \mathcal{N}(0, Q)$, avec $Q > 0$.

Dans le modèle, d'après le lemme10, prenons la vraisemblance

$$Z(x, y) := \exp \left[\langle Q^{-1}h(x), y \rangle - \frac{1}{2}|Q^{-1/2}h(x)|^2 \right]$$

Pour un échantillon de taille n : $(X_1, Y_1), \dots, (X_n, Y_n)$ considérons

$$\begin{aligned} L_n &= Z(X_1, Y_1) \times \dots \times Z(X_n, Y_n) \\ &= \prod_{j=1}^n \exp \left[\langle Q^{-1}h(X_j), Y_j \rangle - \frac{1}{2}|Q^{-1/2}h(X_j)|^2 \right] \\ &= \prod_{j=1}^n \exp \left[\langle Q^{-1}h(X_j), h(X_j) + V_j \rangle - \frac{1}{2}|Q^{-1/2}h(X_j)|^2 \right] \end{aligned}$$

$$\begin{aligned} L_n^{-1} &= \prod_{j=1}^n \exp \left[- \langle Q^{-1}h(X_j), Y_j \rangle + \frac{1}{2}|Q^{-1/2}h(X_j)|^2 \right] \\ &= \prod_{j=1}^n \exp \left[- \langle Q^{-1}h(X_j), h(X_j) + V_j \rangle + \frac{1}{2}|Q^{-1/2}h(X_j)|^2 \right] \\ &= \prod_{j=1}^n \exp \left[- \langle Q^{-1}h(X_j), V_j \rangle - \frac{1}{2}|Q^{-1/2}h(X_j)|^2 \right] \\ &= \exp \sum_{j=1}^n \left[- \langle Q^{-1}h(X_j), V_j \rangle - \frac{1}{2}|Q^{-1/2}h(X_j)|^2 \right] \end{aligned}$$

Pour tout $n \geq 1$, comme V_i sont indépendantes entre-eux et par rapport aux X_j , d'après le théorème

$$\begin{aligned} \mathbb{E} \left[\exp \sum_{j=1}^n [- \langle Q^{-1}h(x), V_j \rangle] \right] &= \mathbb{E} \left[\prod_{j=1}^n \exp [- \langle Q^{-1}h(x), V_j \rangle] \right] \\ &= \prod_{j=1}^n \mathbb{E} [\exp [- \langle Q^{-1}h(x), V_j \rangle]] \\ &= \exp \left[-\frac{1}{2}|Q^{-1/2}h(x)|^2 \right] \end{aligned}$$

$$\mathbb{E} \left(\exp \left[\frac{1}{2}|Q^{-1/2}h(x)|^2 \right] \times \left[\exp \sum_{j=1}^n [- \langle Q^{-1}h(x), V_j \rangle] \right] \right) = 1$$

on a $\mathbb{E}(L_n^{-1}) = 1$ et on définit alors la probabilité \bar{P}_n sur (Ω, \mathcal{F}) par

$$\frac{d\bar{P}_n}{dP} = L_n^{-1}$$

Théorème de Girsanov

Sous $\bar{\mathbb{P}}_n$, (X_0, \dots, X_n) est une chaîne de Markov [PARDOUX,1991], [PARDOUX,2006] de loi initiale μ_0 et de probabilité de transition π indépendante de la suite (Y_1, \dots, Y_n) qui est i.i.d de loi $\mathcal{N}(0, Q)$

Preuve

En effet,

$$\mathbb{E} \exp [- \langle Q^{-1}h(x), V_j \rangle] = \exp \left[-\frac{1}{2}|Q^{-1/2}h(x)|^2 \right]$$

pour toute f mesurable, : comme

$$\begin{aligned} \exp \left[i \sum_{j=1}^n \langle u_j, Y_j \rangle \right] \cdot L_n^{-1} &= \exp \left[i \sum_{j=1}^n \langle u_j, h(X_j) + V_j \rangle \right] \cdot L_n^{-1} = \\ &= \exp \left[\sum_{j=1}^n i \langle u_j, h(X_j) + V_j \rangle - \langle Q^{-1}h(X_j), V_j \rangle - \frac{1}{2}|Q^{-1/2}h(X_j)|^2 \right] \\ &= \exp \left[\sum_{j=1}^n i \langle u_j, h(X_j) \rangle - \frac{1}{2}|Q^{-1/2}h(X_j)|^2 \right] \cdot \exp \left[\sum_{j=1}^n i \langle u_j, V_j \rangle - \langle Q^{-1}h(X_j), V_j \rangle \right] \end{aligned}$$

Notons $\bar{\mathbb{E}}_n$ l'espérance mathématique sous la probabilité $\bar{\mathbb{P}}_n$.

$$\begin{aligned} \bar{\mathbb{E}}_n \left[f(X_0, \dots, X_n) \exp \left(i \sum_{j=1}^n \langle u_j, Y_j \rangle \right) \right] &= \bar{\mathbb{E}}_n [f(X_0, \dots, X_n) e^{i \sum_{j=1}^n \langle u_j, h(X_j) + V_j \rangle}] \\ &= \mathbb{E} \left[f(X_0, \dots, X_n) e^{i \sum_{j=1}^n \langle u_j, h(X_j) + V_j \rangle} L_n^{-1} \right] = \\ &= \mathbb{E} \left[f(X_0, \dots, X_n) \mathbb{E} \left[e^{i \sum_{j=1}^n \langle u_j, h(X_j) + V_j \rangle} L_n^{-1} | X_0, \dots, X_n \right] \right] \end{aligned}$$

et

$$\begin{aligned} &= \mathbb{E} \left[e^{i \sum_{j=1}^n \langle u_j, h(X_j) + V_j \rangle} L_n^{-1} | X_0, \dots, X_n \right] = \\ &= \mathbb{E} \left[\exp \left[\sum_{j=1}^n i \langle u_j, h(X_j) + V_j \rangle - \langle Q^{-1}h(X_j), V_j \rangle - \frac{1}{2}|Q^{-1/2}h(X_j)|^2 \right] \right] = \\ &= \exp \left[\frac{1}{2} \sum_{j=1}^n \langle Qu_j, u_j \rangle \right] \end{aligned}$$

Donc

$$\bar{\mathbb{E}}_n \left[f(X_0, \dots, X_n) \exp \left(i \sum_{j=1}^n \langle u_j, Y_j \rangle \right) \right] = \mathbb{E} [f(X_0, \dots, X_n)] \exp \left[\frac{1}{2} \sum_{j=1}^n \langle Qu_j, u_j \rangle \right]$$

Avec la loi conditionnelle $\mu_n(dx) = \mathbb{P}_{X_n}(dx|\mathcal{Y}_n)$ de X_n sachant \mathcal{Y}_n est telle que

$$(\forall f \in \mathcal{C}_b(E)), \int_{\mathbb{R}^d} f(x)\mu_n(dx) = \mathbb{E}[f(X_n)|\mathcal{Y}_n]$$

et on a :

$$\mathbb{E}[f(X_n)|\mathcal{Y}_n] \times \bar{\mathbb{E}}_n[L_n|\mathcal{Y}_n] = \bar{\mathbb{E}}_n[f(X_n)L_n|\mathcal{Y}_n]$$

plus précisément, [SZPIRGLAS, 1979],

Formule 13 (Kallianpur-Striebel ¹)

$$\mathbb{E}[f(X_n)|\mathcal{Y}_n] = \frac{\bar{\mathbb{E}}[f(X_n)L_n|\mathcal{Y}_n]}{\bar{\mathbb{E}}[L_n|\mathcal{Y}_n]} \quad (\forall f \in \mathcal{C}_b(E))$$

En d'autres termes, il existe une mesure de probabilité σ_n , telle que

$$(\forall f \in \mathcal{C}_b(E)), \int_{\mathbb{R}^d} f(x)\sigma_n(dx) = \bar{\mathbb{E}}[f(X_n)L_n|\mathcal{Y}_n]$$

$$\mu_n(dx) = \frac{\sigma_n(dx)}{\int_{\mathbb{R}^d} \sigma_n(dy)}$$

Preuve

Comme sous la loi $\bar{\mathbb{P}}_n$, X_1, \dots, X_n et Y_1, \dots, Y_n sont indépendantes, on peut donc définir une loi \bar{P} , d'espérance $\bar{\mathbb{E}}$ de la façon suivante

$$\bar{\mathbb{E}}_n[\phi(X_1, \dots, X_n, Y_1, \dots, Y_n)|Y_1, \dots, Y_n] = \bar{\mathbb{E}}[\phi(X_1, \dots, X_n, Y_1, \dots, Y_n)]$$

Remarques

Dans le cas plus général, on définit :

$$Z_t = \exp \left[\int_0^t h(s, X_s) dY_s - \frac{1}{2} \int_0^t |h(s, X_s)|^2 ds \right]$$

et les filtrations

$$\mathcal{G}_t = \sigma\{(X_\theta, Y_\theta), 0 \leq \theta \leq t\}$$

$$\mathcal{F}_t = \sigma\{Y_\theta, 0 \leq \theta \leq t\}$$

On a bien,

$$\frac{d\bar{P}}{dP|_{\mathcal{G}_t}} = (Z_t)^{-1}$$

et pour toute fonction mesurable bornée f ,

$$\mathbb{E}[f(X_t)|\mathcal{F}_t] = \frac{\bar{\mathbb{E}}[f(X_t)Z_t|\mathcal{F}_t]}{\bar{\mathbb{E}}[Z_t|\mathcal{F}_t]} \quad (3.3)$$

Si on pose

$$\int_{\mathbb{R}^d} f(x)p(t, x)dx = \bar{\mathbb{E}}[f(X_t)Z_t|\mathcal{F}_t]$$

, alors

$$\rho(t, x) = p(t, x) \cdot \left[\int_{\mathbb{R}^d} p(t, x) dx \right]^{-1}$$

est la densité de la loi de X_t , conditionnée par \mathcal{F}_t .

On obtient les formes analytiques des filtres optimaux par la proposition suivante :

Proposition 14 *Les mesures σ_n et μ_n sont solutions des équations de récurrences suivantes :*

$$\sigma_n(dx) = Z(x, Y_n) \int_{\mathbb{R}^d} \pi(x'; dx) \sigma_{n-1}(dx'), \quad X_0 \sim \sigma_0(dx)$$

$$\mu_n(dx) = C_n Z(x, Y_n) \int_{\mathbb{R}^d} \pi(x'; dx) \mu_{n-1}(dx'), \quad X_0 \sim \mu_0(dx)$$

où C_n est une constante de normalisation telle que :

$$C_n^{-1} := \int_{\mathbb{R}^d} Z(x, Y_n) \int_{\mathbb{R}^d} \pi(x'; dx) \mu_{n-1}(dx').$$

Preuve : Pour toute fonction test $f \in \mathcal{C}_b(\mathbb{R}^d)$, sachant que $\{X_n, n \geq 0\}$ est homogène,

$$\begin{aligned} \int_{\mathbb{R}^d} f(x) \sigma_n(dx) &= \int_{\mathbb{R}^d} f(x) \bar{\mathbb{P}}(X_n \in dx) \\ &= \int_{\mathbb{R}^d} f(x) L_n \mathbb{P}(X_n \in dx) \\ &= \int_{\mathbb{R}^d} f(x) L_{n-1} Z(x, Y_n) \mathbb{P}(X_n \in dx) \\ &= \int_{\mathbb{R}^d} f(x) L_{n-1} Z(x, Y_n) \int_{\mathbb{R}^d} \mathbb{P}(X_{n-1} \in dx') \cdot \mathbb{P}(X_n \in dx | X_{n-1} = x') \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x) \cdot Z(x, Y_n) \cdot \pi(x', dx) L_{n-1} \mathbb{P}(X_{n-1} \in dx') \\ &= \int_{\mathbb{R}^d} f(x) \cdot \left[Z(x, Y_n) \int_{\mathbb{R}^d} \pi(x', dx) \bar{\mathbb{P}}(X_{n-1} \in dx') \right] \\ &= \int_{\mathbb{R}^d} f(x) \cdot \left[Z(x, Y_n) \int_{\mathbb{R}^d} \pi(x', dx) \sigma_{n-1}(dx') \right] \\ &\text{et } \sigma_n(dx) = Z(x, Y_n) \cdot \int_{\mathbb{R}^d} \pi(x'; dx) \sigma_{n-1}(dx') \end{aligned}$$

De même,

$$\begin{aligned} \mu_n(dx) &= \left(\int_{\mathbb{R}^d} \sigma_n(dx) \right)^{-1} \cdot \sigma_n(dx) \\ &= K_n \cdot \sigma_n(dx) \\ &= K_n \cdot Z(x, Y_n) \cdot \int_{\mathbb{R}^d} \pi(x'; dx) \sigma_{n-1}(dx') \\ &= K_n \cdot Z(x, Y_n) \cdot \int_{\mathbb{R}^d} \pi(x'; dx) K_{n-1} \mu_{n-1}(dx') \\ &= C_n \cdot Z(x, Y_n) \cdot \int_{\mathbb{R}^d} \pi(x'; dx) \mu_{n-1}(dx'). \end{aligned}$$

C_n est une constante de normalisation.

On établit alors une relation de récurrence pour $\mu_n(dx)$ avec un algorithme récursif qui permet le traitement dans l'immédiat, la méthode classique de décomposition en deux étapes, en loi prédite $\mu_n^-(dx) = \mathbb{P}_{X_n}(dx|\mathcal{Y}_{n-1})$ puis en loi corrigée $\mu_n(dx) = \mathbb{P}_{X_n}(dx|\mathcal{Y}_n)$.

Prédiction

$$\begin{aligned}\mu_n^-(dx) &= \mathbb{P}(X_n \in dx|\mathcal{Y}_{n-1}) = \int_{\mathbb{R}^d} \mathbb{P}(X_{n-1} \in dx'|\mathcal{Y}_{n-1}) \cdot \mathbb{P}(X_n \in dx|X_{n-1} = x') \\ &= \int_{\mathbb{R}^d} \pi(x'; dx) \mu_{n-1}(dx')\end{aligned}$$

Corréction

$$\mu_n(dx) = C_n \cdot Z(x, Y_n) \cdot \int_{\mathbb{R}^d} \pi(x'; dx) \mu_{n-1}(dx') = C_n Z(x, Y_n) \mu_n^-(dx)$$

avec

$$Z(x, Y_n) = \exp \left[\langle Q^{-1}h(x), Y_n \rangle - \frac{1}{2} |Q^{-1/2}h(x)|^2 \right]$$

En supposant que la condition initiale X_0 admet une densité $p_0(x)$, le filtre optimal admet une densité $p_n(x)$, la densité conditionnelle de X_n sachant \mathcal{Y}_n telle que :

$$P_i = \begin{cases} Y_n &= h(X_n) + V_n \\ X_n &= f(X_{n-1}) + W_n \end{cases} \quad (3.4)$$

La prédiction est

$$p_n^-(x) dx = \int_{\mathbb{R}^d} \pi(x'; dx) p_{n-1}(x') dx'$$

La corréction est

$$p_n(x) = C_n Z(x, Y_n) p_n^-(x)$$

$$C_n^{-1} := \int_{\mathbb{R}^d} Z(x, Y_n) \int_{\mathbb{R}^d} \pi(x'; dx) p_{n-1}(x') dx'$$

$$\text{avec } Z(x, Y_n) = \exp \left[\langle Q^{-1}h(x), Y_n \rangle - \frac{1}{2} |Q^{-1/2}h(x)|^2 \right]$$

Ces résultats sont destinés à des systèmes à bruits gaussiens.

3.3 Filtrage particulière avec de bruit d'observation uniforme

Considérons le système tel que l'état non observé $X = \{X_n; n \geq 0\}$ est régi par une équation :

$$X_{n+1} = f(X_n) + W_n$$

, où f est une fonction mesurable non nécessairement linéaire et W_n est un bruit blanc de densité f_W avec μ_0 est la loi initiale de X_0 :

$$\mathbb{E}[f(X_0)] = \int_{\mathbb{R}^d} f(x)\mu_0(dx) \quad (3.5)$$

L'observation est un processus tel que :

$$Y = \{Y_n; n \geq 0\}, Y_n = h(X_n) + V_n, n \geq 0.$$

où h est une fonction mesurable non nécessairement linéaire et V_n est un bruit blanc de densité f_V

D'après la proposition 4 de la section 1.3.3, on a :

La loi prédite $\mu_n^-(dx) = \mathbb{P}_{X_n}(dx|\mathcal{Y}_{n-1})$ de X_n sachant les observations jusqu'à l'instant $n - 1$ est :

$$\mu_n^-(dx) = \int_{\mathbb{R}^d} f_W(x - f(x'))\mu_{n-1}^-(x')dx'.$$

La loi corrigée $\mu_n(dx) = \mathbb{P}_{X_n}(dx|\mathcal{Y}_n)$ de X_n sachant les observations jusqu'à l'instant n est :

$\mu_n(dx) = C_n f_V(Y_n - h(x))\mu_n^-(dx)$ où C_n est une constante de normalisation telle que :

$$C_n = \int_{\mathbb{R}^d} f_V(Y_n - h(x'))\mu_n^-(dx')$$

Si μ_0 est une loi à densité, alors les filtres sont à densité tels que :

$$p_n^-(x) = \int_{\mathbb{R}^d} f_W(x - f(x'))p_{n-1}^-(x')dx'.$$

$$p_n(x) = \frac{f_V(Y_n - h(x))p_n^-(x)}{\int_{\mathbb{R}^d} f_V(Y_n - h(x'))p_n^-(x')dx'}$$

Supposons que le bruit d'observation est uniforme sur $\mathcal{D} = [-a; a]^d$, $a > 0$ et que la condition initiale est à densité.

La densité du bruit est telle que,

$$f_V(x) = \frac{1}{m(\mathcal{D})} \mathbb{I}_{\mathcal{D}}(x)$$

où m est la mesure de Lebesgue sur \mathbb{R}^d .

La densité du filtre prédit (1.18) est telle que :

$$p_n^-(x) = \int_{\mathbb{R}^d} f_W(x - f(x')) p_{n-1}(x') dx'$$

La densité du filtre corrigé est, d'après (1.19),

$$p_n(x) = C_n \cdot \mathbb{I}_{\mathcal{D}}(Y_n - h(x)) p_n^-(x)$$

avec la constante de normalisation :

$$C_n^{-1} = \int_{\mathbb{R}^d} \mathbb{I}_{\mathcal{D}}(Y_n - h(x)) p_n^-(x) dx$$

On obtient, en tenant compte que la mesure de Lebesgue est invariante par translation, :

$$C_n^{-1} = \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} f_W(x - f(x')) p_{n-1}(x') dx' \right] \mathbb{I}_{\mathcal{D}}(Y_n - h(x)) dx$$

$$p_n(x) = C_n \cdot \left[\int_{\mathbb{R}^d} f_W(x - f(x')) p_{n-1}(x') dx' \right] \mathbb{I}_{\mathcal{D}}(Y_n - h(x))$$

Notons $\mathcal{D}_n = \{x \in \mathbb{R}^k : Y_n - h(x) \in \mathcal{D}\} = \{x \in \mathbb{R}^k : h(x) \in \mathcal{D} + Y_n\}$

$$\text{Pour } x \in \mathcal{D}_n, p_n(x) = C_n \int_{\mathbb{R}^d} f_W(x - f(x')) p_{n-1}(x') dx'$$

$$\text{avec } C_n^{-1} = \int_{\mathcal{D}_n} \left[\int_{\mathbb{R}^d} f_W(x - f(x')) p_{n-1}(x') dx' \right] dx$$

On obtient donc une relation de récurrence sur les filtres corrigés, $p_n(x)$ en fonction de $p_{n-1}(x)$.

En application du filtrage particulière, où le bruit d'observation V_n suit la loi uniforme $\mathcal{U}([-a; a]^k)$, $a > 0$.

et $\mathcal{D} = [-a; a]^k$ où $m(\mathcal{D})$ est la mesure de Lebesgue de \mathcal{D} dans \mathbb{R}^k .

Après avoir sélectionné les particules $(\hat{\xi}_{n-1}^i, i = 1, \dots, N)$ après l'observation $Y_{n-1} = y_{n-1}$, l'étape de mutation est

$$\xi_n^i = f(\hat{\xi}_{n-1}^i) + w_n^i \text{ avec } w_n^i \sim p_W(w) dw$$

et l'étape de pondération avec l'observation $Y_n = y_n$, est

$$\omega_n^i = \frac{p_V(y_n - h(\xi_n^i))}{N \sum_{j=1}^N p_V(y_n - h(\xi_n^j))}$$

où

$$p_V(y_n - h(\xi_n^i)) = \frac{1}{m(\mathcal{D})} \mathbb{I}_{[-a; a]}(y_n - h(\xi_n^i))$$

Cette quantité est soit nulle, soit $\frac{1}{m(\mathcal{D})}$ et les particules ξ_n^i retenues au vu de l'observation $Y_n = y_n$, sont celles correspondantes à $\omega_n^i \neq 0$, plus précisément quand $y_n - h(\xi_n^i) \in \mathcal{D}$.

Les particules sélectionnées sont redistribuées suivant la loi multinomiale pour compléter le nombre total de particules.

Il n'est pas vraiment nécessaire de calculer le nombre efficace des particules.

Le filtre prédit est approximé par

$$\mu_n^-(dx) = \mathbb{P}_{X_n}(dx|\mathcal{Y}_{n-1}) \simeq p_n^-(x)dx = \sum_{i=1}^N \omega_{n-}^i \cdot \delta_{\xi_{n-}^i}(x)dx$$

avec

$$\xi_n^i = f(\hat{\xi}_{n-1}^i) + w_n^i \text{ avec } w_n^i \sim p_W(w)dw$$

Les particules $\hat{\xi}_{n-1}^i$ sont sélectionnées au temps $n-1$ et $\omega_{n-}^i = \frac{1}{N}$.

Le filtre corrigé est approximé par

$$\mu_n(dx) = \mathbb{P}_{X_n}(dx|\mathcal{Y}_n) \simeq p_n(x)dx = \sum_{i=1}^N \omega_n^i \cdot \delta_{\xi_n^i}(x)dx$$

Les particules ξ_n^i sont sélectionnées quand $y_n - h(\xi_n^i) \in \mathcal{D}$ et dupliquées de poids $\omega_n^i = \frac{k_i}{N}$, k_i est le nombre de duplication de ξ_n^i .

3.4 Filtrage par noyau de convolution d'un système commandé

Un signal (X_n) , $(n \in \mathbb{N})$ est contrôlé par une commande (U_n) , $(n \in \mathbb{N})$ mais inobservable. L'équation d'état de (X_n) , $(n \in \mathbb{N})$ est perturbée par un bruit blanc (W_n) , $(n \in \mathbb{N})$. Le signal observé (Y_n) , $(n \in \mathbb{N}^*)$ est régi par une équation non linéaire et aussi perturbée par un bruit blanc (V_n) , $(n \in \mathbb{N})$.

$$\begin{cases} X_{n+1} &= f_n(X_n, U_n) + W_n \\ Y_n &= h_n(X_n) + V_n \end{cases}$$

Le filtrage par noyau de Parzen-Rosenblatt [CHEN et al.,2012] est une solution efficace surtout pour l'estimation de paramètres et en particulier l'estimation des commandes.

Le problème avec le filtrage par noyau est la lenteur de l'algorithme. Le choix d'une grille fixe support des particules pourrait améliorer cette vitesse.

3.4.1 Application du filtrage particulaire

En supposant que la loi conditionnelle prédite admet une densité telle que :

$$p_n^-(x) = \int_{\mathbb{R}^d} f_w(x - f_n(x', u_n)) p_{n-1}(x') dx' \quad (3.6)$$

la loi conditionnelle corrigée est déterminée par la densité :

$$p_n(x) = \frac{f_v(Y_n - h_n(x)) p_n^-(x)}{\int_{\mathbb{R}^d} f_v(Y_n - h_n(x')) p_n^-(x') dx'} \quad (3.7)$$

En application du filtrage particulaire, l'étape de mutation correspond à la prédiction telle que : $\xi_n^i = f_n(\hat{\xi}_{n-1}^i, u_n) + W_n^i$ et l'étape de pondération correspond à la correction telle que les poids

$$\omega_n^i = \frac{p_V(Y_n - h_n(\xi_n^i))}{N \sum_{j=1}^N p_V(Y_n - h_n(\xi_n^j))}$$

peuvent être trop faibles et entraîneraient une redistribution des particules en utilisant la loi multinomiale par exemple à l'instant $n = 1$. Les nouvelles particules redistribuées $\hat{\xi}_n^i$ déterminent le filtre optimal à l'instant n . Une régularisation par noyau de Parzen-Rosenblatt permet d'avoir la densité du filtre optimal.

Une régularisation du filtre prédit peut être considérée avant l'étape de correction et un échantillonnage à partir de la densité du filtre corrigé permet de déterminer l'approximation particulaire du filtre optimal.

Plus précisément, pour maintenir X_n proche de la trajectoire nominale x_n à chaque instant n , la commande optimale est obtenue par exemple :

$$(u_j^*, \dots, u_n^*) = \operatorname{argmin}_{u_j, \dots, u_n} \sum_{k=j}^n \|x_k - \mathbb{E}(X_k | Y_1, \dots, Y_n, u_1, \dots, u_n)\|^2$$

Avec le système d'équations,

$$\frac{\partial}{\partial u_i} \sum_{k=j}^n \|x_k - \mathbb{E}(X_k | Y_1, \dots, Y_n, u_1, \dots, u_n)\|^2 = 0$$

on peut déterminer les contrôles u_j^*, \dots, u_n^* .

3.4.2 Noyau de Parzen-Rosenblatt

La densité d'une variable aléatoire X peut être estimée par la méthode de noyau de Parzen-Rosenblatt telle que :

$$\tilde{f}_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{x - \xi_i}{h_n}\right) \text{ où } K \text{ est un noyau défini dans } \mathbb{R}^d, \text{ borné et intégrable par rapport à la mesure de Lebesgue et tel que } \lim_{\|n\| \rightarrow \infty} \|x\|^d K(x) = 0, h_n \text{ est le paramètre de}$$

lissage. (ξ_1, \dots, ξ_n) est un échantillon tiré de la variable aléatoire X . Le noyau d'Epanechnikov $K(y) = 1 - \frac{3}{2}y^2$ pour $y < 1$ et le noyau de Picard $K(y) = \exp(-\frac{|y|}{2})$ pour tout y de \mathbb{R} sont parmi d'autres des noyaux de Parzen-Rosenblatt.

3.4.3 Estimation de la densité conditionnelle

Considérons les estimations par noyau de la densité conjointe de 2 variables aléatoires X et Y où X_i et Y_i sont des particules générées des lois de X et de Y .

$$\hat{p}_{XY}(x, y) = \frac{1}{nh^2} \sum_{i=1}^n K_h^{XY} \left(\frac{x - X_i}{h}, \frac{y - Y_i}{h} \right), \text{ et de la densité marginale de } Y,$$

$$\hat{p}_Y(y) = \frac{1}{nh} \sum_{i=1}^n K_h^Y \left(\frac{y - Y_i}{h} \right). \text{ La loi conditionnelle de } X \text{ sachant } Y = y \text{ est définie, d'après le théorème de Bayes, par}$$

$$\hat{p}_{X|Y=y}^{(x)} = \frac{\frac{1}{nh^2} \sum_{i=1}^n K_h^{XY} \left(\frac{x - X_i}{h}, \frac{y - Y_i}{h} \right)}{\frac{1}{nh} \sum_{i=1}^n K_h^Y \left(\frac{y - Y_i}{h} \right)}.$$

K_h est un noyau de Parzen-Rosenblatt de paramètre de lissage h en une dimension ou deux.

$K_h(x, y) = (1 - 3/2x^2)(1 - 3/2y^2)$ dans le rectangle $] - 1; 1[\times] - 1; 1[$ pour le noyau d'Epanechnikov et $K_h(x, y) = 1/4 \exp(-(|x| + |y|))$ pour le noyau de Picard dans \mathbb{R}^2 .

3.4.4 Résolution par filtrage particulière d'un système paramétré

La méthode de filtrage non linéaire muni d'un paramètre θ [AMIRI, 2010] avec les phases prédiction-correction n'est adaptée que pour des états markoviens. Ces phases sont incontournables mais rencontrent des situations complexes dans leurs applications. Supposant que θ admet une loi à priori π_θ et on obtient une estimation θ_0 de θ à l'instant initial.

Considérons le système dynamique $\begin{cases} X_{n+1} &= f_n(X_n, \theta_n) + W_n \\ Y_n &= h_n(X_n, \theta_n) + V_n \end{cases}$ où les bruits sont supposés additifs.

L'objectif est de déterminer à la fois une estimation de X_n et une estimation θ_n du paramètre inconnu θ à chaque instant n par filtrage particulière à noyau de convolution [YOUNDJÉ, 2011], [CHEN et al., 2012].

Notons $X_n^\theta = (X_n, \theta_n)$ l'état caché augmenté, une nouvelle variable et on obtient un nouveau système : $\begin{cases} X_{n+1}^\theta &= f_n(X_n^\theta) + W_n \\ Y_n^\theta &= h_n(X_n^\theta) + V_n. \end{cases}$

Théoriquement, en supposant que la loi conditionnelle prédite admet une densité telle que :

$$p_n^{\theta-}(x) = \int_{\mathbb{R}^d} f_w(x - f_n(x')) \cdot p_{n-1}^\theta(x') dx'$$

La loi conditionnelle corrigée est déterminée par la densité :

$$p_n^\theta(x) = \frac{f_v(Y_n - h_n(x))p_n^{\theta-}(x)}{\int_{\mathbb{R}^d} f_v(Y_n - h_n(x'))p_n^{\theta-}(x')dx'}$$

La densité du filtre du filtre optimal est obtenu par marginalisation telle que :

$$p_n(x) = \int_{\mathbb{R}^d} p_n^\theta(x) d\theta.$$

En application du filtrage particulaire, on dispose de la densité initiale $p_0^\theta(x)dx = \mathbb{P}(X_0^\theta \in dx)$, des densités de probabilités $p_{n,n-1}^\theta(x|y)dx = \mathbb{P}(X_n^\theta dx | X_{n-1}^\theta = y)$ de transition correspondant à l'équation d'évolution et des densités d'émission : $p_{n,Y}^{X,\theta}(y|x)dy = P(Y_n \in dy | X_n^\theta = x)$ correspondant à l'équation d'observation ou la densité de Y sachant $X_n = x$ est : $p_{n,Y}^{X,\theta}(y|x) = f_V(y - h(x))$.

Les particules sont initialisées à partir de $p_0^\theta(x)$. À l'étape $n - 1$, $\hat{\xi}_{n-1}^{\theta(i)}$ sont les M -particules tirées de l'estimation de la loi de densité $\hat{p}_{n-1}^\theta(x|y_1, \dots, y_{n-1})$ de l'état augmenté sachant les observations y_1, \dots, y_{n-1} .

L'étape de mutation correspond à la prédiction en propageant les particules à partir de l'équation d'évolution : $\xi_n^{\theta(i)-} = f(\hat{\xi}_{n-1}^{\theta(i)}) + W_n^{(i)}$.

Avec la loi d'émission, on établit les $p_{n,Y}^{X,\theta}(y|\xi_n^{\theta(i)-})$ pour déterminer une estimation de la loi conjointe de (X_n^θ, Y_n) sachant les observations y_1, \dots, y_{n-1} par convolution à un noyau de Parzen-Rosenblatt :

$$\hat{p}_{XY,n}^\theta(x, y|y_1, \dots, y_{n-1}) = \frac{1}{M} \sum_{i=1}^M K_h(x - \xi_n^{\theta(i)-}) p_{n,Y}^{X,\theta}(y|\xi_n^{\theta(i)-})$$

L'étape de correction correspond à l'estimation de la densité du filtre optimal :

$$\hat{p}_n^\theta(x|y_1, \dots, y_n) = \frac{\sum_{i=1}^M K_h(x - \xi_n^{\theta(i)-}) p_{n,Y}^{X,\theta}(y_n|\xi_n^{\theta(i)-})}{\sum_{i=1}^M p_{n,Y}^{X,\theta}(y_n|\xi_n^{\theta(i)-})}$$

$$\frac{p_{n,Y}^{X,\theta}(y_n|\xi_n^{\theta(i)-})}{\sum_{i=1}^M p_{n,Y}^{X,\theta}(y_n|\xi_n^{\theta(i)-})} = \omega_n^{\theta(i)}$$

est le poids de la particule $\xi_n^{\theta(i)}$.

Les particules peuvent être trop faibles et entraîneraient une redistribution en utilisant la loi multinomiale par exemple à l'instant $n = 1$. Les nouvelles particules redistribuées $\hat{\xi}_n^{\theta(i)}$ déterminent le filtre optimal à l'instant n .

À chaque itération, les estimateurs du paramètre et de l'état caché sont déterminés à partir de $\hat{X}_n^\theta = (\hat{\xi}_n, \hat{\theta}_n)$ où $\hat{X}_n^\theta = \sum_{i=1}^M \omega_n^{\theta(i)} \hat{\xi}_n^{\theta(i)} = \sum_{i=1}^M \omega_n^{\theta(i)} (\hat{\xi}_n^{(i)}, \hat{\theta}_n^{(i)})$,

$$\hat{\theta}_n = \sum_{i=1}^M \omega_n^{\theta(i)} \hat{\theta}_n^{(i)} \text{ et } \hat{\xi}_n = \sum_{i=1}^M \omega_n^{\theta(i)} \hat{\xi}_n^{(i)}$$

Algorithme de filtrage par noyau

On a deux formes d'algorithme de filtrage par noyau : le filtrage par noyau pré-régularisé et le filtrage par noyau post-régularisé [ROSSI, 2005].

Dans le premier cas, le filtre prédit est régularisé par un noyau. L'étape de correction permet d'obtenir le filtre optimal. Une sélection de particules permet de passer à l'étape suivante.

Dans le second cas, après l'étape de mutation des particules, l'étape de pondération permet d'obtenir les particules correspondantes au filtre optimal. Une régularisation par un noyau détermine la densité du filtre optimal.

Dans le filtrage non linéaire d'un système contrôlé,

$$\begin{cases} X_n &= f(X_{n-1}, U_n) + W_n \\ Y_n &= h(X_n) + V_n \end{cases}$$

le paramètre θ est substitué par une variable de contrôle U_n déterministe en chaque instant.

Plus précisément, dans la poursuite de trajectoire, la variable de contrôle à chaque instant est déterministe définie par :

$$u_n^* = \arg_u \|f(\hat{\xi}_{n-1}, u) - x_n\|^2$$

où x_n est la trajectoire réelle à l'instant n .

Un échantillonnage de la loi à priori de θ permet d'obtenir des particules $(\theta_1, \dots, \theta_n)$. À l'étape suivante, on peut supposer que ces particules n'évoluent pas.

Avec l'équation d'observation, on obtient des particules représentant la loi de la variable d'observation à l'étape n . Les estimations du filtre et du paramètre sont déterminées par convolution à un noyau au système de particules.

Le choix du noyau [DEHEUVELS, 1977] n'est pas absolument nécessaire. Avec un test d'approximation de la densité exponentielle avec le noyau d'Epanechnikov et le noyau de Picard, il n'y a pas de remarques particulières.

Une régularisation par convolution à un noyau détermine une approximation du filtre prédit.

Le filtre prédit est corrigé à partir de l'observation en cet instant. Un échantillonnage de ce filtre optimal définit les particules dans l'itération suivante. Avec une grille fixe, le nombre d'apparition de chaque particules est supposé son poids.

L'estimation du paramètre est obtenue par l'espérance de la densité estimée du filtre.

Dans le cas de système commandé, le paramètre estimé est la commande u .

$$\begin{cases} X_n &= f_n(X_{n-1}, u_{n-1}) + W_n \\ Y_n &= h_n(X_n) + V_n \end{cases}$$

La commande est déterminée à chaque instant par

$$u_n^* = \operatorname{argmin}_{1 \leq i \leq N} \|f_n(\hat{\xi}_{n-1}, u_{n-1}^i) - x_n\|^2$$

x_n est la trajectoire nominale connue à l'instant n et u_{n-1}^i sont les particules obtenues par l'estimation de la densité de u à l'instant $n - 1$.

3.5 Filtrage de Kalman inodore d'un modèle à bruit d'observation quelconque

Considérons le système non linéaire avec les bruits blancs W_n et V_n :

$$\begin{cases} X_n &= f_n(X_{n-1}) + W_n \\ Y_n &= h_n(X_n) + V_n \end{cases}$$

Le filtrage de Kalman inodore [JULIER et UHLMANN, 2004], [KJETIL, 2007] consiste à remplacer les distributions conditionnelles par des distributions gaussiennes et approcher les moyennes et les matrices de covariance conditionnelles par des formules de quadratures sur des points déterministes.

Proposition 15 *La moyenne et la covariance du filtre prédit sont définies par :*

$$\begin{cases} \hat{X}_n^- = \mathbb{E}[X_n | Y_1, \dots, Y_{n-1}] = \int_{\mathbb{R}^d} f_n(x) \mu_{n-1}(dx) \\ P_n^- = \int_{\mathbb{R}^d} \left(f_n(x) - \hat{X}_n^- \right) \left(f_n(x) - \hat{X}_n^- \right)^* \mu_{n-1}(dx) + Q_n^W \end{cases}$$

Preuve

En effet W_n est un bruit blanc indépendant de (Y_1, \dots, Y_{n-1}) ,

$$\begin{aligned} \hat{X}_n^- &= \mathbb{E}[X_n | Y_1, \dots, Y_{n-1}] = \mathbb{E}[f_n(X_{n-1}) + W_n | Y_1, \dots, Y_{n-1}] \\ &= \mathbb{E}[f_n(X_{n-1}) | Y_1, \dots, Y_{n-1}] + \mathbb{E}[W_n | Y_1, \dots, Y_{n-1}] \end{aligned}$$

Comme $X_n - \hat{X}_n^- = (f_n(X_{n-1}) - \hat{X}_n^-) + W_n$,

$$\begin{aligned} P_n^- &= \mathbb{E} \left[(X_n - \hat{X}_n^-)(X_n - \hat{X}_n^-)^* | Y_1, \dots, Y_{n-1} \right] \\ &= \mathbb{E} \left[\left(f_n(X_{n-1}) - \hat{X}_n^- + W_n \right) \left(f_n(X_{n-1}) - \hat{X}_n^- + W_n \right)^* | Y_1, \dots, Y_{n-1} \right] \\ &= \mathbb{E} \left[\left(f_n(X_{n-1}) - \hat{X}_n^- \right) \left(f_n(X_{n-1}) - \hat{X}_n^- \right)^* | Y_1, \dots, Y_{n-1} \right] \\ &\quad + \mathbb{E} \left[\left(f_n(X_{n-1}) - \hat{X}_n^- \right) W_n^* | Y_1, \dots, Y_{n-1} \right] \\ &\quad + \mathbb{E} \left[W_n \cdot \left(f_n(X_{n-1}) - \hat{X}_n^- \right)^* | Y_1, \dots, Y_{n-1} \right] \\ &\quad + \mathbb{E} [W_n \cdot W_n^* | Y_1, \dots, Y_{n-1}] \\ &= \int_{\mathbb{R}^d} \left(f_n(x) - \hat{X}_n^- \right) \left(f_n(x) - \hat{X}_n^- \right)^* \mu_{n-1}(dx) + Q_n^W \end{aligned}$$

Proposition 16 *La moyenne et la covariance de la distribution conditionnelle de l'observation Y_n sachant les observations Y_1, \dots, Y_{n-1} sont définies par :*

$$\begin{cases} \hat{Y}_n^- = \mathbb{E}[X_n | Y_1, \dots, Y_{n-1}] = \int_{\mathbb{R}^d} h_n(x) \mu_{n-1}(dx) \\ Q_n^- = \int_{\mathbb{R}^d} \left(h_n(x) - \hat{Y}_n^- \right) \left(h_n(x) - \hat{Y}_n^- \right)^* \mu_{n-1}(dx) + Q_n^V \end{cases}$$

Preuve

En effet V_n est un bruit blanc indépendant de (Y_1, \dots, Y_{n-1}) ,

$$\begin{aligned} \hat{Y}_n^- &= \mathbb{E}[Y_n | Y_1, \dots, Y_{n-1}] = \mathbb{E}[h_n(X_n) + V_n | Y_1, \dots, Y_{n-1}] \\ &= \mathbb{E}[h_n(X_n) | Y_1, \dots, Y_{n-1}] + \mathbb{E}[V_n | Y_1, \dots, Y_{n-1}] \end{aligned}$$

On a aussi, $Y_n - \hat{Y}_n^- = (h_n(X_n) - \hat{Y}_n^-) + V_n$

$$\begin{aligned} Q_n^- &= \mathbb{E} \left[(Y_n - \hat{Y}_n^-)(Y_n - \hat{Y}_n^-)^* | Y_1, \dots, Y_{n-1} \right] \\ &= \mathbb{E} \left[\left(h_n(X_n) - \hat{Y}_n^- + V_n \right) \left(h_n(X_n) - \hat{Y}_n^- + V_n \right)^* | Y_1, \dots, Y_{n-1} \right] \\ &= \mathbb{E} \left[\left(h_n(X_n) - \hat{Y}_n^- \right) \left(h_n(X_n) - \hat{Y}_n^- \right)^* | Y_1, \dots, Y_{n-1} \right] \\ &\quad + \mathbb{E} \left[\left(h_n(X_n) - \hat{Y}_n^- \right) V_n^* | Y_1, \dots, Y_{n-1} \right] \\ &\quad + \mathbb{E} \left[V_n \cdot \left(h_n(X_n) - \hat{Y}_n^- \right)^* | Y_1, \dots, Y_{n-1} \right] \\ &\quad + \mathbb{E} \left[V_n \cdot V_n^* | Y_1, \dots, Y_{n-1} \right] \\ &= \int_{\mathbb{R}^d} \left(h_n(x) - \hat{Y}_n^- \right) \left(h_n(x) - \hat{Y}_n^- \right)^* \mu_{n-1}(dx) + Q_n^V \end{aligned}$$

en tenant compte de l'indépendance de (Y_1, \dots, Y_{n-1}) et V_n

$$\begin{aligned} &\mathbb{E} \left[V_n \cdot \left(h_n(X_n) - \hat{Y}_n^- \right)^* | Y_1, \dots, Y_{n-1} \right] = \\ &= \mathbb{E} \left[\mathbb{E} \left(V_n \cdot \left(h_n(X_n) - \hat{Y}_n^- \right)^* | X_n, Y_1, \dots, Y_{n-1} \right) | Y_1, \dots, Y_{n-1} \right] \\ &= \mathbb{E} \left[\underbrace{\mathbb{E} (V_n \cdot | X_n, Y_1, \dots, Y_{n-1})}_{=0} \left(h_n(X_n) - \hat{Y}_n^- \right)^* | Y_1, \dots, Y_{n-1} \right] = 0 \end{aligned}$$

De même la covariance conditionnelle de la loi jointe de (X_n, Y_n) sachant Y_1, \dots, Y_{n-1} ,

$$\begin{aligned} C_n &= \mathbb{E} \left[(X_n - \hat{X}_n^-)(Y_n - \hat{Y}_n^-)^* | Y_1, \dots, Y_{n-1} \right] \\ &= \mathbb{E} \left[(X_n - \hat{X}_n^-)(h_n(X_n) - \hat{Y}_n^-)^* | Y_1, \dots, Y_{n-1} \right] + \mathbb{E} \left[(X_n - \hat{X}_n^-) V_n^* | Y_1, \dots, Y_{n-1} \right] \\ &= \int_{\mathbb{R}^d} \left(x - \hat{X}_n^- \right) \left(h_n(x) - \hat{Y}_n^- \right)^* \mu_n^-(dx) \end{aligned}$$

La distribution conditionnelle jointe de (X_n, Y_n) sachant Y_1, \dots, Y_{n-1} est considérée comme une distribution gaussienne de moyenne $\begin{pmatrix} \hat{X}_n^- \\ \hat{Y}_n^- \end{pmatrix}$ et de matrice de covariance $\begin{pmatrix} P_n^- & C_n \\ C_n^* & Q_n^- \end{pmatrix}$

Lemme 17 $Z = (X, Y)$ est un vecteur aléatoire gaussien de moyenne $\mathbb{E}Z = \begin{pmatrix} \bar{X} \\ \bar{Y} \end{pmatrix}$ et de matrice de covariance $Q_Z = \begin{pmatrix} Q_X & Q_{XY} \\ Q_{YX} & Q_Y \end{pmatrix}$. Alors si la matrice Q_Y est inversible, la distribution conditionnelle de X sachant $Y = y$ est gaussienne de moyenne

$$\mathbb{E}[\mathbb{E}(X|Y = y)] = \bar{X} + Q_{XY} \cdot Q_Y^{-1} \cdot (y - \bar{Y})$$

et matrice de covariance :

$$\text{cov}[\mathbb{E}(X|Y = y)] = Q_X - Q_{XY} \cdot Q_Y^{-1} \cdot Q_{YX}$$

D'après le lemme 17, on a la proposition suivante

Proposition 18 Si Q_n est inversible, on obtient les relations suivantes :

$$\begin{cases} \hat{X}_n = \hat{X}_n^- + C_n \cdot Q_n^{-1} \cdot (Y_n - \hat{Y}_n^-) \\ P_n = P_n^- - C_n \cdot Q_n^{-1} \cdot C_n^* \end{cases}$$

Posons $P_{n-1} = S_{n-1} \cdot S_{n-1}^*$ et $P_n^- = S_n^- \cdot S_n^{-*}$ et supposons que $\mu_{n-1} \sim \mathcal{N}(\hat{X}_{n-1}, S_{n-1})$,

$$\hat{X}_n^- \simeq \int_{\mathbb{R}^d} \tilde{f}_n(u) \exp \left[-\frac{1}{2} \langle x - \hat{X}_{n-1}, P_{n-1}^{-1} \cdot (x - \hat{X}_{n-1}) \rangle \right] \frac{du}{\det S_{n-1} (2\pi)^{d/2}}$$

avec un changement de variables $x = \hat{X}_{n-1} + S_{n-1} \cdot u$, notons,

$$\tilde{f}_n(u) = f_n(\hat{X}_{n-1} + S_{n-1} \cdot u)$$

on a les approximations gaussiennes,

$$\hat{X}_n^- \simeq \int_{\mathbb{R}^d} \tilde{f}_n(u) \exp \left[-\frac{1}{2} |u|^2 \right] \frac{du}{(2\pi)^{d/2}} \quad (3.8)$$

$$P_n^- \simeq \int_{\mathbb{R}^d} (\tilde{f}_n(u) - \hat{X}_n^-) (\tilde{f}_n(u) - \hat{X}_n^-)^* \exp \left[-\frac{1}{2} |u|^2 \right] \frac{du}{(2\pi)^{d/2}} + Q_n^W \quad (3.9)$$

De même, supposons que, $\mu_n^- \sim \mathcal{N}(\hat{X}_n^-, S_n^-)$, et posons

$$\tilde{h}_n(u) = h_n(\hat{X}_n^- + S_n^- \cdot u)$$

On obtient les approximations gaussiennes suivantes,

$$\hat{Y}_n^- \simeq \int_{\mathbb{R}^d} \tilde{h}_n(u) \exp \left[-\frac{1}{2} |u|^2 \right] \frac{du}{(2\pi)^{d/2}} \quad (3.10)$$

$$Q_n^- \simeq \int_{\mathbb{R}^d} (\tilde{h}_n(u) - \hat{Y}_n^-) (\tilde{h}_n(u) - \hat{Y}_n^-)^* \exp \left[-\frac{1}{2} |u|^2 \right] \frac{du}{(2\pi)^{d/2}} + Q_n^V \quad (3.11)$$

$$C_n \simeq S_n^- \int_{\mathbb{R}^d} u (\tilde{h}_n(u) - \hat{Y}_n^-)^* \exp \left[-\frac{1}{2} |u|^2 \right] \frac{du}{(2\pi)^{d/2}} \quad (3.12)$$

3.5.1 Approximation de la loi gaussienne reduite

En supposant que X est de dimension d de loi gaussienne centrée reduite, on introduit une formule de quadrature à partir de $2d + 1$ σ -points ξ_{-d}, \dots, ξ_d tels que :

$$\begin{aligned} \xi_0 = 0, \quad \xi_i = \sqrt{d + \kappa} \cdot e_i \quad \text{et} \quad \xi_{-i} = -\xi_i \text{ munis des poids} \\ \omega_0 = \frac{\kappa}{d + \kappa} \text{ et pour tout } 1 < i < d, \quad \omega_{-i} = \omega_i = \frac{1}{2(d + \kappa)} \end{aligned} \quad (3.13)$$

où e_i est le i -ème vecteur de la base canonique de \mathbb{R}^d ,
 κ est un paramètre contrôlant la dispersion des σ -points

$$\text{avec} \quad \sum_{i=-d}^{+d} \omega_i = 1, \quad \sum_{i=-d}^{+d} \omega_i \cdot \xi_i = 0 \quad \text{et} \quad \sum_{i=-d}^{+d} \omega_i \cdot \xi_i \cdot \xi_i^* = \sum_{i=1}^d e_i \cdot e_i^* = I_d$$

Dans l'approximation de la distribution de X suivant la loi normale en dimension 1, de moyenne m et d'écart-type σ , on peut avoir κ à partir de la probabilité

$$\mathbb{P}[m - \kappa \cdot \sigma < X < m + \kappa \cdot \sigma] = 1 - \alpha$$

α est l'imprécision égale à 0.1, 0.05 ou 0.01. Pour $\alpha = 0.05$, $\kappa = 1.96$.

Pour toute fonction test φ , mesurable bornée sr \mathbb{R}^d ,

$$\mathbb{E}\varphi(X) = \int_{\mathbb{R}^d} \varphi(x) \exp\left[-\frac{1}{2}|x|^2\right] \frac{dx}{(2\pi)^{d/2}} \text{ est approximée par } \sum_{i=-d}^{+d} \omega_i \varphi(\xi_i).$$

Pour une variable gaussienne, $X = \mu + Q \cdot I_d$, de moyenne μ et de matrice de covariance Q , les σ -points x_{-i}, \dots, x_i associées sont définies par

$$\begin{aligned} x_0 = \mu, \quad x_i = \mu + Q^{1/2} \cdot e_i \cdot \sqrt{d + \kappa} \quad \text{et} \quad x_{-i} = \mu - Q^{1/2} \cdot e_i \cdot \sqrt{d + \kappa} \\ \text{avec les propriétés, } \sum_{i=-d}^{+d} \omega_i \cdot x_i = \mu, \quad \sum_{i=-d}^{+d} \omega_i \cdot (x_i - \mu)(x_i - \mu)^* = \sum_{i=1}^d Q^{1/2} e_i ((Q^{1/2} \cdot e_i)^*)^* = Q \end{aligned}$$

3.5.2 Approximation des paramètres des filtres

Dans le filtrage de Kalman inodore, à partir de $(2d + 1)$ σ -points munis des poids dans l'équation (3.13) avec $P_{n-1} = S_{n-1} \cdot S_{n-1}^*$ la matrice de covariance du filtre à l'étape $n - 1$ et $P_n^- = S_n^- \cdot S_n^{-*}$ la matrice de covariance du filtre prédit à l'étape n ,

$$\text{les } \sigma\text{-points sont, pour } 1 < i < d, \quad \begin{cases} x_0 = \hat{X}_{n-1} \\ x_i = \hat{X}_{n-1} + S_{n-1} \cdot e_i \cdot \sqrt{d + \kappa} \\ x_{-i} = \hat{X}_{n-1} - S_{n-1} \cdot e_i \cdot \sqrt{d + \kappa} \end{cases}$$

$$\tilde{f}_n(u) = f_n(\hat{X}_{n-1} + S_{n-1}u) \quad (3.14)$$

on obtient la moyenne du filtre prédit

$$\hat{X}_n^- = \sum_{i=-d}^{+d} \omega_i \cdot f_n(\hat{X}_{n-1} + S_{n-1} \cdot x_i)$$

$$\hat{X}_n^- = \sum_{i=-d}^{+d} \omega_i \cdot \tilde{f}_n(x_i) \quad (3.15)$$

et de matrice de covariance :

$$P_n^- = \sum_{i=-d}^{+d} \omega_i \cdot (f_n(\hat{X}_{n-1} + S_{n-1} \cdot x_i) - \hat{X}_n^-) \cdot (f_n(\hat{X}_{n-1} + S_{n-1} \cdot x_i) - \hat{X}_n^-)^* + Q_n^W$$

$$P_n^- = \sum_{i=-d}^{+d} \omega_i \cdot (\tilde{f}_n(x_i) - \hat{X}_n^-) \cdot (\tilde{f}_n(x_i) - \hat{X}_n^-)^* + Q_n^W \quad (3.16)$$

$$\tilde{h}_n(u) = h_n(\hat{X}_n^- + S_n^- u) \quad (3.17)$$

$$\hat{Y}_n^- = \sum_{i=-d}^{+d} \omega_i \cdot \tilde{h}_n(x_i) = \sum_{i=-d}^{+d} \omega_i \cdot h_n(\hat{X}_n^- + S_n^- \cdot x_i) \quad (3.18)$$

$$Q_n = \sum_{i=-d}^{+d} \omega_i \cdot (h_n(\hat{X}_n^- + S_n^- \cdot x_i) - \hat{Y}_n^-) \cdot (h_n(\hat{X}_n^- + S_n^- \cdot x_i) - \hat{Y}_n^-)^* + Q_n^V$$

$$Q_n = \sum_{i=-d}^{+d} \omega_i \cdot (\tilde{h}_n(x_i) - \hat{Y}_n^-) \cdot (\tilde{h}_n(x_i) - \hat{Y}_n^-)^* + Q_n^V \quad (3.19)$$

de matrice de corrélation,

$$C_n = \sum_{i=-d}^{+d} \omega_i \cdot (f_n(\hat{X}_{n-1} + S_{n-1} \cdot x_i) - \hat{Y}_n^-) \cdot (h_n(\hat{X}_n^- + S_n^- \cdot x_i) - \hat{Y}_n^-)^*$$

$$C_n = \sum_{i=-d}^{+d} \omega_i \cdot (\tilde{f}_n(x_i) - \hat{X}_n^-) \cdot (\tilde{h}_n(x_i) - \hat{Y}_n^-)^* \quad (3.20)$$

La moyenne du filtre corrigé est :

$$\hat{X}_n = \hat{X}_n^- + C_n \cdot Q_n^{-1} \cdot (Y_n - \hat{Y}_n^-) \quad (3.21)$$

de matrice de covariance

$$P_n = P_n^- - C_n \cdot Q_n^{-1} \cdot C_n^* = S_n \cdot S_n^* \quad (3.22)$$

Les décompositions des matrices de covariance sont obtenues par le lemme suivant :

Lemme 19 (Décomposition de Cholesky) *Pour toute matrice définie positive Q , il existe une matrice triangulaire orthogonale telle que $Q = S \cdot S^*$. $S_{1,1} = \sqrt{Q_{1,1}}$*

$$\text{Pour } 2 \leq i \leq d \text{ alors, } S_{i,1} = \frac{Q_{1,i}}{A_{1,1}} \text{ et } S_{i,i} = \sqrt{Q_{i,i} - \sum_{1 \leq k \leq i-1} |Q_{i,k}|^2}$$

$$\text{Et pour } i \leq j \leq d, \text{ alors } S_{j,i} = \frac{Q_{i,j} - \sum_{k=1}^{i-1} S_{i,j} \cdot S_{j,k}}{S_{i,i}} \text{ et } A_{i,j} = 0.$$

En particulier si $d = 2$,

$$Q = \begin{pmatrix} \sigma_1^2 & \rho \cdot \sigma_1 \cdot \sigma_2 \\ \rho \cdot \sigma_1 \cdot \sigma_2 & \sigma_2^2 \end{pmatrix}$$

$$S = \begin{pmatrix} \sigma_1 & 0 \\ \rho \cdot \sigma_2 & \sigma_2 \cdot \sqrt{1 - \rho^2} \end{pmatrix}$$

3.6 Approximation du bruit d'observation par un mélange gaussien

Toute densité de probabilité p_V d'un vecteur aléatoire X de moyenne m_X et de matrice de covariance P_X peut être approximée par une somme finie pondérée de N densités élémentaires.

En particulier, la décomposition est justifiée par le lemme suivant,

Lemme 20 (Décomposition de Karhunen-Lowe) X est un vecteur aléatoire gaussien de dimension d , de moyenne m et de matrice de covariance Q .

Si $k = \text{rang}(Q)$, alors il existe U_1, \dots, U_k , i.i.d suivant $\mathcal{N}(0, 1)$ et de nombres réels $(\alpha_{i,j})$ tels que :

$$X_i = \sum_{j=1}^k \alpha_{i,j} \cdot U_j, \quad i = 1, \dots, d$$

En posant $A = (\alpha_{i,j})_{i,j}$ la matrice $d \times k$, $X = A \cdot U$ où U est le vecteur aléatoire gaussien de dimension k centré de matrice de covariance I_k .

Dans le cas plus général, pour toute densité p , on a :

$$p(x) = \sum_{i=1}^N \bar{\omega}_i \cdot p_i(x) \quad \text{avec } p_i \sim \mathcal{N}(\bar{m}_i, \bar{P}_i)$$

Dans cette approximation, on a $3N$ paramètres à déterminer : les moyennes \bar{m}_i , les matrices de covariance \bar{P}_i et les poids initiaux $\bar{\omega}_i$, avec les conditions pour tout $1 \leq i \leq N$,

$$\sum_{i=1}^N \bar{\omega}_i = 1, \quad \sum_{i=1}^N \bar{\omega}_i \cdot \bar{P}_i = P, \quad \sum_{i=1}^N \bar{\omega}_i \cdot \bar{m}_i = m$$

Dans l'application avec l'équation d'observation, $Y_n = h(X_n) + V_n$ où V_n est un bruit quelconque, on a la proposition suivante,

Proposition 21 *Tout bruit d'observation peut être approximé par un mélange de bruits blancs gaussiens .*

Preuve

Considérons la densité f_V du bruit d'observation V de support I dans \mathbb{R}^d et une grille de discretisation de I . $\xi_i, i = 1 \dots N$ sont les points de la grille.

La décomposition en mélange gaussien peut être définie par une formule d'interpolation :

$$f_V(x) \simeq \sum_{i=1}^N \omega_i \cdot p_i(x)$$

où p_i est la densité gaussienne de moyenne ξ_i et de matrice de covariance $\bar{P}_i = \frac{1}{N} \cdot I_d$, matrice unité d'ordre d .

Les poids ω_i sont obtenus par le système d'équations linéaires tel que : pour tout $j = 1$ à N ,

$$\sum_{i=1}^N \omega_i \cdot p_i(\xi_j) = f_V(\xi_j)$$

Lemme 22 *L'équation d'observation, $Y_n = h(X_n) + V_n$ où $V_n \sim \mathcal{N}(m, Q^V)$ peut être remplacée par $Y_n = \tilde{h}(X_n) + V_n^0$ où $\tilde{h}(X_n) = h(X_n) + m$ et $V_n^0 \sim \mathcal{N}(0, Q^V)$.*

Lemme 23 *Si $V_1 \sim \mathcal{N}(m_1, Q^{V_1})$ et $V_2 \sim \mathcal{N}(m_2, Q^{V_2})$, alors $V_1 + V_2 \sim \mathcal{N}(m_1 + m_2, Q^{V_1} + Q^{V_2})$ si V_1 et V_2 sont indépendantes.*

Posons alors,

$$V_n \simeq \sum_{i=1}^N \bar{\omega}_i \cdot V_{i,n} \text{ avec } V_{i,n} \sim \mathcal{N}(\xi_i, \bar{P}_i).$$

$$Y_n = h(X_n) + V_n \text{ devient } Y_n = \tilde{h}(X_n) + \tilde{V}_n$$

où d'après le lemme 22

$$\tilde{h}(X_n) = h(X_n) + \sum_{i=1}^N \bar{\omega}_i \cdot \xi_i \text{ et } \tilde{V}_n \sim \mathcal{N}(0, \bar{\omega}_i \cdot \bar{P}_i)$$

3.7 Conclusion

Dans ce chapitre, nous avons développé une théorie sur l'existence du filtre optimal à partir du théorème de Girsanov. Cette théorie est applicable dans le cas des bruits gaussiens. Elle demande une connaissance à priori de la structure markovienne de l'état du signal à étudier, en particulier de la loi de transition homogène. On obtient une relation de récurrence entre le filtre optimal à un instant précis et le filtre optimal à l'instant antérieur. Une décomposition en phases de prédiction et correction ne pose pas trop de problème mais l'application numérique introduit des simulations basées sur la méthode de Monte Carlo.

Le cas de bruit d'observation uniforme est étudié dans le filtrage particulière. L'application de ce filtrage dans ce cas est immédiate et assez simple. La discrétisation de bruits quelconques en bruits uniformes est une solution acceptable.

Pour les systèmes paramétrés et commandés, le choix du filtrage particulière par noyau de convolution est justifié par sa facilité d'adaptation. Pour un système de dimension deux ou trois, la régularisation se fait sur une grille fixe.

Chapitre 4

Études de stabilité en filtrage

Sommaire

4.1 Introduction	83
4.2 Stabilité du filtrage de chaîne de Markov	84
4.3 Oubli dans le cas de bruits gaussien	90
4.4 Oubli dans le cas de bruits d'observation uniformes	91
4.5 Stabilité de filtre de Kalman	91
4.6 Stabilité de filtres particulières	98
4.7 Stabilité du filtrage par noyau de convolution	103
4.8 Conclusion	105

4.1 Introduction

La propriété d'oubli de la condition initiale mesure l'efficacité d'une méthode de filtrage. Intuitivement, le filtre optimal dépend de la condition initiale, définie par la loi initiale de X_0 . Cette propriété est vérifiée en considérant deux lois initiales μ_0 et $\bar{\mu}_0$ différentes, on obtient les filtres optimaux déduits μ_n et $\bar{\mu}_n$ dans un horizon assez lointain, possédant pratiquement les mêmes caractéristiques.

Cette stabilité ou cet oubli de la condition initiale est donc une des qualités des filtres optimaux déduits des différentes méthodes de filtrage mais selon la situation, les normes utilisées dans les calculs sont différentes.

Supposant que l'on utilise pour le filtre une loi initiale $\bar{\mu}_0(dx)$ alors que la véritable loi initiale est $\mu_0(dx)$.

Avec $\bar{\mu}_0(dx)$, on met en œuvre le filtre prédit $\bar{\mu}_n^-(dx)$ et le filtre corrigé $\bar{\mu}_n(dx)$, alors que le vrai filtre prédit est $\mu_n^-(dx)$, et le vrai filtre corrigé est $\mu_n(dx)$.

En général, la norme dans L^1 de la différence des densités est utilisée pour les filtres à densité.

Pour que le filtre oublie sa condition initiale ou pour qu'il y a stabilité, il suffit que

$$\lim_{n \rightarrow \infty} \varepsilon_n = 0$$

$$\text{où } \varepsilon_n = \int_{\mathbb{R}^d} |\mu_n(dx) - \bar{\mu}_n(dx)| = \int_{\mathbb{R}^d} |g_n(x) - \bar{g}_n(x)| dx$$

Une autre approche de l'étude de l'oubli de la condition initiale est l'utilisation de la norme de la variation totale.

Par définition, la norme en variation totale d'une mesure μ dans la tribu borelienne $\mathcal{B}(\mathbb{R}^d)$ est telle que :

$$\|\mu\|_{VT} = \sup_{\|f\|_\infty=1} \left| \int_{\mathbb{R}^d} f(x) \cdot \mu(dx) \right| = \sup_{\|f\|_\infty=1} |\langle \mu, f \rangle|$$

$$\|\mu\|_{VT} = 2 \sup [|\mu(A)|, A \in \mathcal{B}(\mathbb{R}^d)] = \sup_{A \in \mathcal{B}(\mathbb{R}^d)} \mu(A) - \inf_{A \in \mathcal{B}(\mathbb{R}^d)} \mu(A)$$

Il y a stabilité exponentielle [RUBENTHALER,2010], [GUYON,1999] si

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log[\|\mu_n - \bar{\mu}_n\|_{VT}] < 0$$

L'étude de stabilité est étroitement liée à l'étude de convergence dans l'approximation.

4.2 Stabilité du filtrage de chaîne de Markov

Considérons une chaîne de Markov non observable $\{X_n, n \in \mathbb{N}\}$ de loi initiale μ_0 et noyau de transition $\mathbb{P}(X_n \in dx' | X_{n-1} = x) = P_n(x, dx')$ telle que

$$\begin{aligned} \mathbb{P}\left(\bigcap_{k=0}^n X_k \in dx_k\right) &= \mathbb{P}(X_0 \in dx_0, X_1 \in dx_1, \dots, X_n \in dx_n) = \\ &= \mathbb{P}(X_0 \in dx_0) \mathbb{P}(X_1 \in dx_1 | X_0 = x_0) \dots \mathbb{P}(X_n \in dx_n | X_0 = x_0, \dots, X_{n-1} = x_{n-1}) = \\ &= \mathbb{P}(X_0 \in dx_0) \mathbb{P}(X_1 \in dx_1 | X_0 = x_0) \dots \mathbb{P}(X_n \in dx_n | X_{n-1} = x_{n-1}) = \\ &= \mu_0(dx_0) P_1(x_0, dx_1) \dots P_n(x_{n-1}, dx_n) = \mu_0(dx_0) \prod_{k=1}^n P_k(x_{k-1}, dx_k). \end{aligned}$$

$E_0, E_1, \dots, E_n, \dots$ une suite d'ensembles mesurables d'états dans \mathbb{R}^d .

$F_1, F_2, \dots, F_n, \dots$ une suite d'ensembles mesurables d'observations dans \mathbb{R}^p .

Supposons que le noyau P_n admet une densité telle que : $P_n(x, dy) = p_n(x, y) dy$.

D'après le théorème Chapman-Kolmogorov pour des états continus, la loi de l'état à

l'instant n est définie récursivement par :

$$\begin{aligned}
\mathbb{P}(X_n \in dx_n) &= \int_{E_{n-1}} \mathbb{P}(X_{n-1} \in dx_{n-1}, X_n \in dx_n) \\
&= \int_{E_{n-1}} \mathbb{P}(X_n \in dx_n | X_{n-1} = x) \mathbb{P}(X_{n-1} \in dx_{n-1}) \\
&= \int_{E_{n-1}} P_n(x_{n-1}, dx_n) \cdot \mathbb{P}(X_{n-1} \in dx_{n-1}) \\
&= \int_{E_{n-1}} \mathbb{P}(X_n \in dx_n | X_{n-1} = x) \mathbb{P}(X_{n-1} \in dx_{n-1}) \\
&= \int_{E_{n-1}} p_n(x_{n-1}, x_n) \cdot dx_n \mathbb{P}(X_{n-1} \in dx_{n-1})
\end{aligned}$$

Plus précisément, pour toute fonction test ϕ mesurable bornée, avec $\mu_0(dx_0) = p_0(x_0)dx_0$,

$$\begin{aligned}
\mathbb{E}(\phi(X_n)) &= \int_{E_n} \phi(x_n) \cdot \mathbb{P}(X_n \in dx_n) \\
&= \int_{E_n} \phi(x_n) \cdot \int_{E_{n-1}} P_n(x_{n-1}, dx_n) \cdot \mathbb{P}(X_{n-1} \in dx_{n-1}) \\
&= \int_{E_n} \phi(x_n) \cdot \int_{E_{n-1}} P_n(x_{n-1}, dx_n) \cdot \int_{E_{n-2}} P_{n-1}(x_{n-2}, dx_{n-1}) \cdot \mathbb{P}(X_{n-2} \in dx_{n-2}) \\
&= \int_{E_n} \dots \int_{E_0} \phi(x_n) \cdot \mu_0(dx_0) \prod_{k=1}^n P_k(x_{k-1}, dx_k) \\
&= \int_{E_n} \dots \int_{E_0} \phi(x_n) \cdot p_0(x_0) dx_0 \prod_{k=1}^n p_k(x_{k-1}, x_k) dx_k
\end{aligned}$$

Le processus $\{Y_n, n \in \mathbb{N}^*\}$ à valeurs dans \mathbb{R}^p , est mutuellement indépendant conditionnellement au processus X_n et que chaque observation Y_n ne dépend que de X_n .

Q_n est la loi d'émission, loi reliant l'état et l'observation, telle que

$$\mathbb{P}(Y_n \in dy | X_n = x) = Q_n(x, dy)$$

Le modèle d'observation est $Y_n = h(X_n) + V_n$ où V_n est un bruit blanc additif perturbateur de l'observation de loi de densité g_V .

Pour tout fonction mesurable bornée ϕ , et pour $F_1 = \dots = F_n = \dots = \mathbb{R}^p$

$$\begin{aligned}
\mathbb{E}(\phi(Y_n) | X_n = x) &= \int_{\mathbb{R}^p} \phi(y) \mathbb{P}(Y_n \in dy | X_n = x) = \int_{\mathbb{R}^p} \phi(y) Q_n(x, dy) \\
\mathbb{E}(\phi(Y_n) | X_n = x) &= \mathbb{E}(\phi(h(X_n) + V_n) | X_n = x) = \mathbb{E}(\phi(h(x) + V_n)) = \\
&= \int_{\mathbb{R}^p} \phi(h(x) + v) \mathbb{P}(V_n \in dv) = \int_{\mathbb{R}^p} \phi(h(x) + v) \cdot g_V(v) dv = \int_{\mathbb{R}^p} \phi(y) \cdot g_V(y - h(x)) dy
\end{aligned}$$

La densité de la loi d'émission est définie telle que $Q_n(x, dy) = g_V(y - h(x)) dy$.

On définit de même la densité de la loi conjointe des états et des observations par,

$$\begin{aligned}
\mathbb{P} \left[\bigcap_{k=0}^n X_k \in dx_k, \bigcap_{k=1}^n Y_k \in dy_k \right] &= \mathbb{P} \left[X_0 \in dx_0, \bigcap_{k=1}^n (X_k \in dx_k, Y_k \in dy_k) \right] \\
&= \mu_0(dx_0) \prod_{k=1}^n P_k(x_{k-1}, dx_k) \cdot Q_k(x_k, dy_k) \\
&= p_0(x_0) dx_0 \prod_{k=1}^n p_k(x_{k-1}, x_k) \cdot g_V(y_k - h(x_k)) dx_k \cdot dy_k
\end{aligned}$$

Et par marginalisation, pour toute fonction test mesurable bornée ϕ ,

$$\begin{aligned}
\mathbb{E}(\phi(X_n)) &= \int_{E_n} \phi(x_n) \mathbb{P}(X_n \in dx_n) = \int_{E_n} \phi(x_n) \int_{E_{n-1}} \mathbb{P}(X_{n-1} \in dx_{n-1}, X_n \in dx_n) \\
&= \int_{E_n} \phi(x_n) \int_{E_{n-1}} \int_{F_n} \mathbb{P}(X_{n-1} \in dx_{n-1}, X_n \in dx_n, Y_n \in dy_n) \\
&= \int_{E_n} \phi(x_n) \int_{E_{n-1}} \int_{F_n} \dots \int_{F_1} \int_{E_0} \mathbb{P} \left[X_0 \in dx_0, \bigcap_{k=1}^n (X_k \in dx_k, Y_k \in dy_k) \right] \\
&= \int_{E_n} \phi(x_n) \int_{E_{n-1}} \int_{F_n} \dots \int_{F_1} \int_{E_0} \mu_0(dx_0) \prod_{k=1}^n P_k(x_{k-1}, dx_k) \cdot g_V(y_k - h(x_k)) dy_k \\
&= \int_{E_n} \phi(x_n) \int_{E_{n-1}} \int_{F_n} \dots \int_{F_1} \int_{E_0} p_0(x_0) dx_0 \prod_{k=1}^n p_k(x_{k-1}, x_k) \cdot g_V(y_k - h(x_k)) dx_k \cdot dy_k
\end{aligned}$$

La loi conjointe des observations est définie telle que :

$$\begin{aligned}
\mathbb{P} \left[\bigcap_{k=1}^n Y_k \in dy_k \right] &= \int_{E_n} \mathbb{P} \left[X_n \in dx_n, \bigcap_{k=1}^n Y_k \in dy_k \right] \\
&= \int_{E_n} \dots \int_{E_0} \mathbb{P} \left[\bigcap_{k=0}^n X_k \in dx_k, \bigcap_{k=1}^n Y_k \in dy_k \right] \\
&= \int_{E_n} \dots \int_{E_0} \mu_0(dx_0) \prod_{k=1}^n P_k(x_{k-1}, dx_k) \cdot g_V(y_k - h(x_k)) dy_k \\
&= \int_{E_n} \dots \int_{E_0} p_0(x_0) dx_0 \prod_{k=1}^n p_k(x_{k-1}, x_k) \cdot g_V(y_k - h(x_k)) dy_k \cdot dx_k
\end{aligned}$$

La loi conjointe de X_n avec les observations est :

$$\begin{aligned}
\mathbb{P} \left[X_n \in dx_n, \bigcap_{k=1}^n Y_k \in dy_k \right] &= \int_{E_{n-1}} \mathbb{P} \left[X_{n-1} \in x_{n-1}, X_n \in dx_n, \bigcap_{k=1}^n Y_k \in dy_k \right] \\
&= \int_{E_{n-1}} \dots \int_{E_0} \mathbb{P} \left[X_0 \in dx_0, \bigcap_{k=1}^n (X_k \in dx_k, Y_k \in dy_k) \right] \\
&= \int_{E_{n-1}} \dots \int_{E_0} \mu_0(dx_0) \prod_{k=1}^n P_k(x_{k-1}, dx_k) \cdot Q_k(x_k, dy_k) \\
&= \int_{E_{n-1}} \dots \int_{E_0} \mu_0(dx_0) \prod_{k=1}^n P_k(x_{k-1}, dx_k) \cdot g_V(y_k - h(x_k)) dy_k \\
&= \int_{E_{n-1}} \dots \int_{E_0} p_0(x_0) dx_0 \prod_{k=1}^n p_k(x_{k-1}, x_k) \cdot g_V(y_k - h(x_k)) dy_k \cdot dx_k
\end{aligned}$$

D'après la formule de Bayes sur les densités conditionnelles,

$$p_{X_n|Y_1=y_1, \dots, Y_n=y_n}(x_n) = \frac{p_{X_n, Y_1, \dots, Y_n}(x_n, y_1, \dots, y_n)}{p_{Y_1, \dots, Y_n}(y_1, \dots, y_n)}$$

On obtient la mesure de Feynman-Kac normalisée exprimant le filtre optimal μ_n :

$$\begin{aligned}
\langle \mu_n, \phi \rangle &= \mathbb{E}(\phi(X_n) | Y_1 = y_1, \dots, Y_n = y_n) \\
&= \int_{E_n} \phi(x_n) \mathbb{P}(X_n \in dx_n | Y_1 = y_1, \dots, Y_n = y_n) \\
&= \int_{E_n} \phi(x_n) \frac{p_{X_n, Y_1, \dots, Y_n}(x_n, y_1, \dots, y_n)}{p_{Y_1, \dots, Y_n}(y_1, \dots, y_n)} \cdot dx_n
\end{aligned}$$

Plus précisément d'après les lois conjointes,

$$\langle \mu_n, \phi \rangle = \frac{\int_{E_n} \phi(x_n) \int_{E_{n-1}} \dots \int_{E_0} \mu_0(dx_0) \prod_{k=1}^n g_V(y_k - h(x_k)) \cdot P_k(x_{k-1}, dx_k)}{\int_{E_n} \dots \int_{E_0} \mu_0(dx_0) \prod_{k=1}^n g_V(y_k - h(x_k)) \cdot P_k(x_{k-1}, dx_k)} \quad (4.1)$$

$$\langle \mu_n, \phi \rangle = \frac{\int_{E_n} \phi(x_n) \int_{E_{n-1}} \dots \int_{E_0} p_0(x_0) dx_0 \prod_{k=1}^n g_V(y_k - h(x_k)) \cdot p_k(x_{k-1}, x_k) dx_k}{\int_{E_n} \dots \int_{E_0} p_0(x_0) dx_0 \prod_{k=1}^n g_V(y_k - h(x_k)) \cdot p_k(x_{k-1}, x_k) dx_k} \quad (4.2)$$

Notons, en considérant $E_0 = \dots = E_n = \dots = \mathbb{R}^d$,

$$\begin{aligned}
\gamma_n(\phi) &= \int_{E_n} \phi(x_n) \int_{E_{n-1}} \dots \int_{E_0} \mu_0(dx_0) \prod_{k=1}^n g_V(y_k - h(x_k)) \cdot P_k(x_{k-1}, dx_k) \\
&= \int_{E_n} \phi(x_n) \int_{E_{n-1}} \dots \int_{E_0} p_0(x_0) dx_0 \prod_{k=1}^n g_V(y_k - h(x_k)) \cdot p_k(x_{k-1}, x_k) \cdot dx_k
\end{aligned}$$

alors on exprime la mesure telle que $\langle \mu_n, \phi \rangle = \frac{\gamma_n(\phi)}{\gamma_n(1)}$.

Posons, $q_0(x_0) = p_0(x_0)$

$$q_1(x_1) = \int_{E_0} q_0(x_0) p_1(x_0, x_1) dx_0,$$

$$\gamma_1(1) = \int_{E_1} q_1(x_1) g_V(y_1 - h(x_1)) dx_1$$

$$q_2(x_2) = \int_{E_1} q_1(x_1) \cdot g_V(y_1 - h(x_1)) \cdot p_1(x_1, x_2) dx_1,$$

$$\gamma_2(1) = \int_{E_2} q_2(x_2) g_V(y_2 - h(x_2)) dx_2$$

.....

$$q_n(x_n) = \int_{E_{n-1}} q_{n-1}(x_{n-1}) \cdot g_V(y_{n-1} - h(x_{n-1})) \cdot p_{n-1}(x_{n-1}, x_n) dx_{n-1},$$

$$\gamma_n(1) = \int_{E_n} q_n(x_n) g_V(y_n - h(x_n)) dx_n$$

$$\gamma_n(\phi) = \int_{E_n} \phi(x_n) \cdot q_n(x_n) g_V(y_n - h(x_n)) dx_n$$

$$\text{et } \langle \mu_n, \phi \rangle = \frac{\int_{E_n} \phi(x_n) \cdot q_n(x_n) g_V(y_n - h(x_n)) dx_n}{\int_{E_n} q_n(x_n) g_V(y_n - h(x_n)) dx_n}$$

La densité du filtre optimal est :

$$f_n(x_n) = \frac{q_n(x_n) g_V(y_n - h(x_n))}{\int_{E_n} q_n(x_n) g_V(y_n - h(x_n))} = \frac{q_n(x_n) g_V(y_n - h(x_n))}{\gamma_n(1)}$$

$$\begin{aligned} f_n(x_n) &= \frac{\int_{E_{n-1}} q_{n-1}(x_{n-1}) \cdot g_V(y_{n-1} - h(x_{n-1})) \cdot p_{n-1}(x_{n-1}, x_n) dx_{n-1}}{\gamma_n(1)} \cdot g_V(y_n - h(x_n)) \\ &= \frac{\int_{E_{n-1}} f_{n-1}(x_{n-1}) \gamma_{n-1}(1) \cdot p_{n-1}(x_{n-1}, x_n) dx_{n-1}}{\gamma_n(1)} \cdot g_V(y_n - h(x_n)) \end{aligned}$$

Pour simplifier, considérons $E_0 = E_1 = \dots = E$.

On a la forme recursive sur les densités des filtres optimaux :

$$f_{n+1}(x) = \frac{\gamma_n(1)}{\gamma_{n+1}(1)} \cdot \left[\int_E f_n(z) \cdot p_n(z, x) dz \right] \cdot g_V(y_{n+1} - h(x)) \quad (4.3)$$

$$\begin{aligned}
\langle \mu_{n+1}, \phi \rangle &= \int_E \phi(x) \cdot f_{n+1}(x) dx \\
&= \int_E \phi(x) \frac{\gamma_n(1)}{\gamma_{n+1}(1)} \cdot \left[\int_E f_n(z) \cdot p_n(z, x) dz \right] \cdot g_V(y_{n+1} - h(x)) dx \\
&= \frac{\gamma_n(1)}{\gamma_{n+1}(1)} \cdot \int_E \phi(x) \left[\int_E f_n(z) \cdot p_n(z, x) dz \right] \cdot g_V(y_{n+1} - h(x)) dx \\
&= \frac{\gamma_{n+1}(\phi)}{\gamma_{n+1}(1)}
\end{aligned}$$

Et

$$\gamma_{n+1}(\phi) = \gamma_n(1) \cdot \int_E \phi(x) \left[\int_E f_n(z) \cdot p_n(z, x) dz \right] \cdot g_V(y_{n+1} - h(x)) dx$$

Cette relation de récurrence est pratique dans les simulations, mais pour l'étude de l'oubli de la condition initiale, reprenons les définitions générales.

Étant données deux lois initiales μ_0 et $\bar{\mu}_0$ différentes et considérons μ_n et $\bar{\mu}_n$ les filtres optimaux déduits à l'instant n .

Supposons de plus que la chaîne est faiblement ergodique c'est-à-dire que .

$$\lim_{n \rightarrow \infty} \|P_{X_n} - \bar{P}_{X_n}\|_{VT} = 0.$$

où P_{X_n} et \bar{P}_{X_n} les lois à priori de X_n .

Soit ϕ une fonction mesurable bornée telle que $\|\phi\|_\infty = 1$ et pour un $\varepsilon > 0$, à partir d'un certain rang n_0 , $|\langle P_{X_n}, \phi \rangle - \langle \bar{P}_{X_n}, \phi \rangle| < \varepsilon$, c'est à dire,

$$\begin{aligned}
|\langle P_{X_n}, \phi \rangle - \langle \bar{P}_{X_n}, \phi \rangle| &= \left| \int_{E_n} \dots \int_{E_0} \phi(x_n) (\mu_0(dx_0) - \bar{\mu}_0(dx_0)) \prod_{k=1}^n P_k(x_{k-1}, dx_k) \right| \\
&= \left| \int_{E_n} \dots \int_{E_0} \phi(x_n) (p_0(x_0) - \bar{p}_0(x_0)) dx_0 \prod_{k=1}^n p_k(x_{k-1}, x_k) dx_k \right| < \varepsilon
\end{aligned}$$

Notons les mesures suivantes,

$$\nu_0(dx_0) = \frac{\mu_0(dx_0)}{\int_{E_n} \dots \int_{E_0} \mu_0(dx_0) \prod_{k=1}^n P_k(x_{k-1}, dx_k) \cdot g_V(y_k - h(x_k))}$$

et

$$\bar{\nu}_0(dx_0) = \frac{\bar{\mu}_0(dx_0)}{\int_{E_n} \dots \int_{E_0} \bar{\mu}_0(dx_0) \prod_{k=1}^n P_k(x_{k-1}, dx_k) \cdot g_V(y_k - h(x_k))}$$

, Ces nouvelles mesures redéfinissent les lois des filtres optimaux déduits, et en introduisant

la faible érgodicité de X_n ,

$$\begin{aligned} | \langle \mu_n, \phi \rangle - \langle \bar{\mu}_n, \phi \rangle | &= \left| \int_{E_n} \dots \int_{E_0} \phi(x_n) (\nu_0(dx_0) - \bar{\nu}_0(dx_0)) \prod_{k=1}^n P_k(x_{k-1}, dx_k) \cdot g_V(y_k - h(x_k)) \right| \\ &= \left| \int_{E_n} \dots \int_{E_0} \phi(x_n) (p_0(x_0) - \bar{p}_0(x_0)) dx_0 \prod_{k=1}^n p_k(x_{k-1}, x_k) \cdot g_V(y_k - h(x_k)) dx_k \right| \\ &\leq \left| \int_{E_n} \dots \int_{E_0} \phi(x_n) (p_0(x_0) - \bar{p}_0(x_0)) dx_0 \prod_{k=1}^n p_k(x_{k-1}, x_k) dx_k \right| < \varepsilon \end{aligned}$$

Finalement, la faible ergodicité de la chaîne est une hypothèse utile dans la vérification de l'oubli de la condition initiale. Cette hypothèse n'est que suffisante mais dans des cas plus précis, on peut opérer autrement.

4.3 Oubli dans le cas de bruits gaussiens

Dans ce qui suit, la chaîne est supposée homogène, c'est-à-dire que pour tout instant n ,

$$\mathbb{P}(X_{n+1} \in dx_{n+1} | X_n = x_n) = \pi(x_n; x_{n+1})$$

π est le noyau de transition de la chaîne de Markov homogène.

Considérons les mesures de la section 3.2, avec les lois initiales $\sigma_0(dx)$ et $\bar{\sigma}_0(dx)$, on obtient les lois déduites à l'instant n par :

$$\begin{aligned} \sigma_n(dx) &= Z(x, Y_n) \int_{\mathbb{R}^d} \pi(x'; dx) \sigma_{n-1}(dx') \\ \bar{\sigma}_n(dx) &= Z(x, Y_n) \int_{\mathbb{R}^d} \pi(x'; dx) \bar{\sigma}_{n-1}(dx'). \end{aligned}$$

On a,

$$\begin{aligned} |\sigma_n(dx) - \bar{\sigma}_n(dx)| &= \left| Z(x, Y_n) \left[\int_{\mathbb{R}^d} \pi(x'; dx) \sigma_{n-1}(dx') - \int_{\mathbb{R}^d} \pi(x'; dx) \bar{\sigma}_{n-1}(dx') \right] \right| \\ &\leq Z(x, Y_n) \left[\int_{\mathbb{R}^d} \pi(x'; dx) |\sigma_{n-1}(dx') - \bar{\sigma}_{n-1}(dx')| \right] \\ &\text{avec } Z(x, Y_n) = \exp \left[\langle Q_V^{-1} h(x), Y_n \rangle - \frac{1}{2} |Q_V^{-1/2} h(x)|^2 \right] \end{aligned}$$

Q_V est la matrice de covariance du bruit d'observation gaussien.

$$\text{et notons l'erreur } \tilde{\varepsilon}_n = \int_{\mathbb{R}^d} |\sigma_n(dx) - \bar{\sigma}_n(dx)|$$

Avec une hypothèse supplémentaire sur Y_n telle que,

$$\langle Q^{-1} h(x), Y_n - \frac{1}{2} Q^{-1} h(x) \rangle \leq 0$$

, on obtient $\exp \left[\langle Q^{-1} h(x), Y_n \rangle - \frac{1}{2} |Q^{-1/2} h(x)|^2 \right] \leq 1$ et il existe une suite k_n telle que :

$$0 < Z(x, Y_n) = k_n < 1, \forall x \in \mathbb{R}^k.$$

Remarque

Dans \mathbb{R} , l'hypothèse sur Y_n est telle que : $Y_n \leq \frac{1}{2} h(x), \forall x \in \mathbb{R}$.

$$\begin{aligned} \tilde{\varepsilon}_n &= \int_{\mathbb{R}^d} |\sigma_n(dx) - \bar{\sigma}_n(dx)| \\ &\leq \int_{\mathbb{R}^d} Z(x, Y_n) \left[\int_{\mathbb{R}^d} \pi(x'; dx) |\sigma_{n-1}(dx') - \bar{\sigma}_{n-1}(dx')| \right] dx \\ &\leq \int_{\mathbb{R}^d} k_n \left[\int_{\mathbb{R}^d} \pi(x'; dx) |\sigma_{n-1}(dx') - \bar{\sigma}_{n-1}(dx')| \right] dx \end{aligned}$$

Finalement, $\|\sigma_n - \bar{\sigma}_n\|_{VT} \leq k_n \dots k_1 \cdot \|\sigma_0 - \bar{\sigma}_0\|_{VT}$ et à la limite, $\|\sigma_n - \bar{\sigma}_n\|_{VT}$ est proche de 0.

Avec les filtres optimaux $\mu_n(dx) = \frac{\sigma_n}{\int_{\mathbb{R}^d} \sigma_n(dx)}$ et $\bar{\mu}_n(dx) = \frac{\bar{\sigma}_n}{\int_{\mathbb{R}^d} \bar{\sigma}_n(dx)}$, des mesures normalisées, toujours avec la majoration de la quantité $Z(x, Y_n)$, on obtient le même résultat et l'oubli de la condition initiale est vérifié.

Dans le cas de bruit d'observation non gaussien, en particulier uniforme, les calculs se font sur les densités. On a les densités suivantes :

$$p_n^-(x) = \int_{\mathbb{R}^d} f_W(x - f(x')) p_{n-1}(x') dx'$$

et pour le filtre corrigé,

$$p_n(x) = \frac{f_V(Y_n - h(x)) p_n^-(x)}{\int_{\mathbb{R}^d} f_V(Y_n - h(x)) p_n^-(x) dx'}$$

4.4 Oubli dans le cas de bruits d'observation uniformes

Dans cette section, les lois initiales sont à densité et nécessairement les filtres optimaux sont à densité.

On suppose donc que l'on utilise pour le filtre une densité $\bar{p}_0(x)$ alors que la véritable loi initiale est $p_0(x)$.

Avec $\bar{p}_0(x)$, on met en œuvre la densité du filtre prédit $\bar{p}_n^-(x)$ et la densité du filtre corrigé $\bar{p}_n(x)$, alors que la vraie densité du filtre prédit est $p_n^-(x)$, et la vraie densité filtre corrigé est $p_n(x)$ et notons

$$\varepsilon_n = \int_{\mathbb{R}^d} |p_n(x) - \bar{p}_n(x)| dx$$

D'après l'équation (1.18), les filtres prédits sont :

$$p_n^-(x) = \int_{\mathbb{R}^d} f_W(x - f(x')) p_{n-1}(x') dx' \text{ et } \bar{p}_n^-(x) = \int_{\mathbb{R}^d} f_W(x - f(x')) \bar{p}_{n-1}(x') dx'.$$

Et d'après l'équation (1.19), les filtres corrigés sont :

$$p_n(x) = \frac{f_V(Y_n - h(x)) p_n^-(x)}{\int_{\mathbb{R}^d} f_V(Y_n - h(x)) p_n^-(x) dx} \text{ et } \bar{p}_n(x) = \frac{f_V(Y_n - h(x)) \bar{p}_n^-(x)}{\int_{\mathbb{R}^d} f_V(Y_n - h(x)) \bar{p}_n^-(x) dx}.$$

Le filtre oublie sa condition initiale ou qu'il y a stabilité si

$$\lim_{n \rightarrow \infty} \varepsilon_n = 0$$

Le bruit de mesure est supposé gaussien centré de matrice de covariance diagonale Q_W et de densité :

$$\begin{aligned} f_W(w) &= \frac{1}{(2\pi)^{k/2} |\det Q_W|^{1/2}} \exp \left[-\frac{1}{2} \langle w, Q_W^{-1} \cdot w \rangle \right] \\ &= \frac{1}{(2\pi \cdot \sigma_W)^{d/2}} \exp \left[-\frac{1}{2\sigma_W^2} |w|^2 \right] \end{aligned}$$

$$\text{avec } Q_W = \begin{bmatrix} \sigma_W & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_W \end{bmatrix} \text{ et } Q_W^{-1} = \begin{bmatrix} \frac{1}{\sigma_W} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sigma_W} \end{bmatrix}$$

et le bruit d'observation est uniforme sur $\mathcal{D} = [-a; a]^k$ de densité :

$$f_V(v) = \frac{1}{m(\mathcal{D})} \mathbb{I}_{\mathcal{D}}(v), \quad a > 0.$$

En notant $\mathcal{D}_n = \{x \in \mathbb{R}^k : h(x) \in \mathcal{D} + Y_n\}$, pour $x \in \mathcal{D}_n$, on définit :

les filtres prédits à l'instant n , déduits de $p_0(x)$ et $\bar{p}_0(x)$

$$\begin{aligned} p_n^-(x) &= \int_{\mathbb{R}^d} f_W(x - f(x')) p_{n-1}(x') dx' = \int_{\mathbb{R}^d} \frac{1}{(2\pi \cdot \sigma_W)^{d/2}} \exp \left[-\frac{1}{2\sigma_W^2} |x - f(x')|^2 \right] p_{n-1}(x') dx' \\ \bar{p}_n^-(x) &= \int_{\mathbb{R}^d} f_W(x - f(x')) \bar{p}_{n-1}(x') dx' = \int_{\mathbb{R}^d} \frac{1}{(2\pi \cdot \sigma_W)^{d/2}} \exp \left[-\frac{1}{2\sigma_W^2} |x - f(x')|^2 \right] \bar{p}_{n-1}(x') dx' \end{aligned}$$

Comme $f_V(Y_n - h(x)) = \frac{1}{m(\mathcal{D})}$ si $Y_n - h(x) \in \mathcal{D}$, ou $x \in \mathcal{D}_n$

les filtres corrigés à cet instant sont :

$$p_n(x) = \frac{f_V(Y_n - h(x)) p_n^-(x)}{\int_{\mathbb{R}^d} f_V(Y_n - h(x)) p_n^-(x) dx} = \frac{p_n^-(x)}{\int_{\mathcal{D}_n} p_n^-(x) dx} \text{ et}$$

$$\bar{p}_n(x) = \frac{f_V(Y_n - h(x))\bar{p}_n^-(x)}{\int_{\mathbb{R}^d} f_V(Y_n - h(x))\bar{p}_n^-(x)dx} = \frac{\bar{p}_n^-(x)}{\int_{\mathcal{D}_n} \bar{p}_n^-(x)dx}$$

On a donc,

$$\begin{aligned} |p_n(x) - \bar{p}_n(x)| &= \left| \frac{p_n^-(x)}{\int_{\mathcal{D}_n} p_n^-(x)dx} - \frac{\bar{p}_n^-(x)}{\int_{\mathcal{D}_n} \bar{p}_n^-(x)dx} \right| \\ &\leq \int_{\mathbb{R}^d} \exp \left[-\frac{1}{2\sigma_W^2} |x - f(x')|^2 \right] |C_n \cdot p_{n-1}(x') - \bar{C}_n \cdot \bar{p}_{n-1}(x')| dx' \end{aligned}$$

Avec les constantes de normalisation,

$$\begin{aligned} C_n &= \frac{1}{\int_{\mathcal{D}_n} \left[\int_{\mathbb{R}^d} \exp \left[-\frac{1}{2\sigma_W^2} |x - f(x')|^2 \right] p_{n-1}(x') dx' \right] dx} \\ \bar{C}_n &= \frac{1}{\int_{\mathcal{D}_n} \left[\int_{\mathbb{R}^d} f \exp \left[-\frac{1}{2\sigma_W^2} |x - f(x')|^2 \right] \bar{p}_{n-1}(x') dx' \right] dx} \end{aligned}$$

Posons

$$\begin{aligned} k_n &= \int_{\mathcal{D}_n} \int_{\mathbb{R}^d} \exp \left[-\frac{1}{2\sigma_W^2} |x - f(x')|^2 \right] p_{n-1}(x') dx' dx \\ \bar{k}_n &= \int_{\mathcal{D}_n} \int_{\mathbb{R}^d} \exp \left[-\frac{1}{2\sigma_W^2} |x - f(x')|^2 \right] \bar{p}_{n-1}(x') dx' dx \end{aligned}$$

$$\begin{aligned} \varepsilon_n &= \int_{\mathcal{D}^n} |p_n(x) - \bar{p}_n(x)| dx \\ &\leq \int_{\mathcal{D}^n} \int_{\mathbb{R}^d} \left| \frac{p_{n-1}(x')}{k_n} - \frac{\bar{p}_{n-1}(x')}{\bar{k}_n} \right| \cdot \exp \left[-\frac{1}{2\sigma_W^2} |x - f(x')|^2 \right] dx' dx \\ &\leq \int_{\mathcal{D}^n} \int_{\mathbb{R}^d} \left(\left| \frac{p_{n-1}(x')}{k_n} - \frac{\bar{p}_{n-1}(x')}{k_n} \right| + \left| \frac{\bar{p}_{n-1}(x')}{k_n} - \frac{\bar{p}_{n-1}(x')}{\bar{k}_n} \right| \right) \cdot \exp \left[-\frac{1}{2\sigma_W^2} |x - f(x')|^2 \right] dx' dx \\ &\leq \frac{1}{k_n} \int_{\mathcal{D}^n} \int_{\mathbb{R}^d} |p_{n-1}(x') - \bar{p}_{n-1}(x')| \cdot \exp \left[-\frac{1}{2\sigma_W^2} |x - f(x')|^2 \right] dx' dx + \\ &+ \int_{\mathcal{D}^n} \int_{\mathbb{R}^d} \left| \frac{\bar{p}_{n-1}(x')}{k_n} - \frac{\bar{p}_{n-1}(x')}{\bar{k}_n} \right| \cdot \exp \left[-\frac{1}{2\sigma_W^2} |x - f(x')|^2 \right] dx' dx \\ &\leq \frac{1}{k_n} \int_{\mathcal{D}^n} \int_{\mathbb{R}^d} |p_{n-1}(x') - \bar{p}_{n-1}(x')| \cdot \exp \left[-\frac{1}{2\sigma_W^2} |x - f(x')|^2 \right] dx' dx + \\ &+ \left| \frac{1}{k_n} - \frac{1}{\bar{k}_n} \right| \cdot \int_{\mathcal{D}^n} \int_{\mathbb{R}^d} \bar{p}_{n-1}(x') \cdot \exp \left[-\frac{1}{2\sigma_W^2} |x - f(x')|^2 \right] dx' dx \end{aligned}$$

Pour $x \in \mathcal{D}^n$ fixé, comme la fonction $x' \mapsto \exp \left[-\frac{1}{2\sigma_W^2} |x - f(x')|^2 \right]$ est continue bornée dans \mathbb{R}^d , d'après le théorème des valeurs intermédiaires, il existe $\alpha \in \mathbb{R}^d$, tel que

$$\begin{aligned} \sup_{x' \in \mathbb{R}^d} \left(\exp \left[-\frac{1}{2\sigma_W^2} |x - f(x')|^2 \right] \right) &= \exp \left[-\frac{1}{2\sigma_W^2} |x - f(\alpha)|^2 \right] \\ |k_n - \bar{k}_n| &\leq \int_{\mathcal{D}_n} \int_{\mathbb{R}^d} \exp \left[-\frac{1}{2\sigma_W^2} |x - f(x')|^2 \right] |p_{n-1}(x') - \bar{p}_{n-1}(x')| dx' . dx \\ &\leq \int_{\mathcal{D}_n} \exp \left[-\frac{1}{2\sigma_W^2} |x - f(\alpha)|^2 \right] \left[\int_{\mathbb{R}^d} |p_{n-1}(x') - \bar{p}_{n-1}(x')| dx' \right] . dx \\ &\leq \varepsilon_{n-1} \int_{\mathcal{D}_n} \exp \left[-\frac{1}{2\sigma_W^2} |x - f(\alpha)|^2 \right] . dx \end{aligned}$$

On a donc

$$\begin{aligned} \varepsilon_n &\leq \frac{\varepsilon_{n-1}}{k_n} \int_{\mathcal{D}_n} \exp \left[-\frac{1}{2\sigma_W^2} |x - f(\alpha)|^2 \right] . dx + \frac{\varepsilon_{n-1}}{k_n \cdot \bar{k}_n} \int_{\mathcal{D}_n} \exp \left[-\frac{1}{2\sigma_W^2} |x - f(\alpha)|^2 \right] . dx \times \\ &\quad \times \int_{\mathcal{D}_n} \exp \left[-\frac{1}{2\sigma_W^2} |x - f(\alpha)|^2 \right] . dx \\ \varepsilon_n &\leq K_n \cdot \varepsilon_{n-1} \text{ avec } K_n = \frac{\int_{\mathcal{D}_n} \exp \left[-\frac{1}{2\sigma_W^2} |x - f(\alpha)|^2 \right] . dx}{k_n} \left(1 + \frac{\int_{\mathcal{D}_n} \exp \left[-\frac{1}{2\sigma_W^2} |x - f(\alpha)|^2 \right] . dx}{\bar{k}_n} \right) \end{aligned}$$

Finalement comme $K_n < 1$, $\lim_{n \rightarrow \infty} \varepsilon_n = 0$ et le filtre oublie sa condition initiale.

4.5 Stabilité de filtres de Kalman

Considérons deux conditions initiales, $X_{1,0}$ et $X_{2,0}$ de lois gaussiennes respectives $\mathcal{N}(m_{1,0}, P_{1,0})$ et $\mathcal{N}(m_{2,0}, P_{2,0})$. Les filtres optimaux déduits à l'horizon n sont des gaussiens déterminés par leurs paramètres respectifs $(\hat{X}_{1,n}, P_{1,n})$ et $(\hat{X}_{2,n}, P_{2,n})$ et notons,

$$\begin{aligned} \hat{X}_{1,n} &= \mathbb{E}[X_{1,n} | \mathcal{Y}_n] \text{ et } P_{1,n} = \mathbb{E} \left[(X_{1,n} - \hat{X}_{1,n}) \cdot (X_{1,n} - \hat{X}_{1,n})^t | \mathcal{Y}_n \right] \\ \hat{X}_{2,n} &= \mathbb{E}[X_{2,n} | \mathcal{Y}_n] \text{ et } P_{2,n} = \mathbb{E} \left[(X_{2,n} - \hat{X}_{2,n}) \cdot (X_{2,n} - \hat{X}_{2,n})^t | \mathcal{Y}_n \right] \end{aligned}$$

$X_{1,n}$ et $X_{2,n}$ sont les états indépendants à l'instant n .

La solution par approximation par filtrage de Kalman dépend de la condition initiale.

D'après les relations de récurrence (2.14) et (2.15), on a pour $i = 1$ ou 2 :

$$\begin{aligned} \hat{X}_{i,n} &= F_n \hat{X}_{i,n-1} + f_n + K_{i,n} [Y_n - H_n \cdot F_n \cdot \hat{X}_{i,n-1} - H_n \cdot f_n - h_n] \\ &= (F_n - K_{i,n} \cdot H_n \cdot F_n) \hat{X}_{i,n-1} + (f_n + K_{i,n} \cdot Y_n - K_{i,n} \cdot H_n \cdot f_n - K_{i,n} \cdot h_n) \end{aligned}$$

où les estimations des covariances des filtres prédits sont :

$$P_{i,n}^- = F_n P_{i,n-1} F_n^t + G_n Q_n^W G_n^t$$

les covariances des innovations dans l'équation (2.10)

$$Q_{i,n}^Z = H_n P_{i,n}^- H_n^t + Q_n^V$$

avec les gains dans l'équation (2.11) ,

$$K_{i,n} = P_{i,n}^- H_n^t [Q_{i,n}^Z]^{-1}$$

et les estimations des covariances des filtres corrigés sont :

$$\begin{aligned} P_{i,n} &= [I - K_{i,n} H_n] [F_n P_{i,n-1} F_n^t + G_n Q_n^W G_n^t] \\ &= [I - K_{i,n} H_n] F_n P_{i,n-1} F_n^t + [I - K_{i,n} H_n] G_n Q_n^W G_n^t \end{aligned}$$

On a les expressions explicites de $\hat{X}_{i,n}$ et $P_{i,n}$ en fonction des paramètres de $X_{i,0}$ et on peut montrer que, tout en restant dans le cas gaussien, qu'avec une condition initiale erronée, on peut avoir le même résultat qu'avec le véritable filtre.

On a,

$$\begin{aligned} \|\hat{X}_{1,n} - \hat{X}_{2,n}\| &= \|F_n(\hat{X}_{1,n-1} - \hat{X}_{2,n-1}) + (K_{1,n} - K_{2,n})[Y_n - H_n \cdot f_n - h_n] + \\ &\quad + K_{2,n} H_n \cdot F_n \cdot \hat{X}_{2,n-1} - K_{1,n} H_n \cdot F_n \cdot \hat{X}_{1,n-1}\| \\ &\leq \|F_n(\hat{X}_{1,n-1} - \hat{X}_{2,n-1})\| + \|(K_{1,n} - K_{2,n})[Y_n - H_n \cdot f_n - h_n]\| + \\ &\quad + \|K_{2,n} H_n \cdot F_n \cdot \hat{X}_{2,n-1} - K_{1,n} H_n \cdot F_n \cdot \hat{X}_{1,n-1}\| \\ &\leq \|F_n(\hat{X}_{1,n-1} - \hat{X}_{2,n-1})\| + \|(K_{1,n} - K_{2,n})[Y_n - H_n \cdot f_n - h_n]\| + \\ &\quad + \|K_{2,n} H_n \cdot F_n \cdot (\hat{X}_{2,n-1} - \hat{X}_{1,n-1})\| + \|(K_{1,n} - K_{2,n}) H_n \cdot F_n \cdot \hat{X}_{2,n-1}\| \end{aligned}$$

$$\begin{aligned} \|K_{1,n} - K_{2,n}\| &= \|P_{1,n}^- H_n^t [Q_{1,n}^Z]^{-1} - P_{2,n}^- H_n^t [Q_{2,n}^Z]^{-1}\| \\ &\leq \|(P_{1,n}^- - P_{2,n}^-) H_n^t [Q_{1,n}^Z]^{-1}\| + \|P_{2,n}^- H_n^t ([Q_{1,n}^Z]^{-1} - [Q_{2,n}^Z]^{-1})\| \end{aligned}$$

$$\begin{aligned} \|P_{1,n}^- - P_{2,n}^-\| &= \|F_n P_{i,n-1} F_n^t - F_n P_{i,n-1} F_n^t\| \\ &= \|F_n (P_{i,n-1} - P_{i,n-1}) F_n^t\| \end{aligned}$$

La densité du filtre optimal déduit de la condition initiale $p_{i,0}$ est gaussienne telle que :

$$p_{i,n}(x) = \frac{1}{(2\pi)^{d/2} |\det P_{i,n}|^{1/2}} \exp \left[-\frac{1}{2} \langle x - \hat{X}_{i,n}, (P_{i,n})^{-1} \cdot (x - \hat{X}_{i,n}) \rangle \right]$$

Sans perte de généralités, prenons $d = 1$. On a :

les estimations des variances des filtres prédits,

$$\sigma_{i,n}^{2-} = F_n^2 \cdot \sigma_{i,n-1}^2 + G_n^2 \cdot \sigma_{W,n}^2$$

les variances des innovations et les gains de Kalman,

$$\sigma_{Z,i,n}^2 = H_n^2 \cdot \sigma_{i,n}^{2-} + \sigma_{V,n}^2 \text{ et } K_{i,n} = \frac{\sigma_{i,n}^{2-} \cdot H_n}{\sigma_{Z,i,n}^2} = \frac{\sigma_{i,n}^{2-} \cdot H_n}{H_n^2 \cdot \sigma_{i,n}^{2-} + \sigma_{V,n}^2}.$$

les estimations des variances des filtres corrigés,

$$\sigma_{i,n}^2 = (1 - K_{i,n}H_n)F_n^2\sigma_{i,n-1}^2 + (1 - K_{i,n}H_n)G_n^2\sigma_{W,n}^2$$

$$|\hat{X}_{1,n} - \hat{X}_{2,n}| \leq \underbrace{|F_n + K_{2,n}H_nF_n|}_{A_n} |\hat{X}_{1,n-1} - \hat{X}_{2,n-1}| + \underbrace{|Y_n - H_nf_n - h_n + H_nF_n\hat{X}_{2,n-1}|}_{C_n} |K_{1,n} - K_{2,n}|$$

$$|\sigma_{1,n}^{2-} - \sigma_{2,n}^{2-}| = F_n^2 |\sigma_{1,n-1}^2 - \sigma_{2,n-1}^2|$$

$$\begin{aligned} |K_{1,n} - K_{2,n}| &\leq \left| \frac{(\sigma_{1,n}^{2-} - \sigma_{2,n}^{2-})H_n}{\sigma_{Z,1,n}^2} \right| + \left| \sigma_{2,n}^{2-} H_n \left(\frac{1}{\sigma_{Z,1,n}^2} - \frac{1}{\sigma_{Z,2,n}^2} \right) \right| \\ &\leq \frac{|H_n|}{\sigma_{Z,1,n}^2} |\sigma_{1,n}^{2-} - \sigma_{2,n}^{2-}| + \frac{|\sigma_{2,n}^{2-} H_n|}{\sigma_{Z,1,n}^2 \cdot \sigma_{Z,2,n}^2} |\sigma_{Z,2,n}^2 - \sigma_{Z,1,n}^2| \\ &\leq \frac{|H_n| F_n^2}{\sigma_{Z,1,n}^2} |\sigma_{1,n-1}^2 - \sigma_{2,n-1}^2| + \frac{|\sigma_{2,n}^{2-} H_n|}{\sigma_{Z,1,n}^2 \cdot \sigma_{Z,2,n}^2} |\sigma_{Z,2,n}^2 - \sigma_{Z,1,n}^2| \\ &\leq \frac{|H_n| F_n^2}{H_n^2 \sigma_{1,n}^{2-} + \sigma_{V,n}^2} |\sigma_{1,n-1}^2 - \sigma_{2,n-1}^2| + \frac{|\sigma_{2,n}^{2-} H_n| H_n^2}{(H_n^2 \sigma_{1,n}^{2-} + \sigma_{V,n}^2) \cdot (H_n^2 \sigma_{2,n}^{2-} + \sigma_{V,n}^2)} |\sigma_{2,n}^{2-} - \sigma_{1,n}^{2-}| \\ &\leq \underbrace{\left[\frac{|H_n| F_n^2}{H_n^2 \sigma_{1,n}^{2-} + \sigma_{V,n}^2} + \frac{|\sigma_{2,n}^{2-} H_n| H_n^2 F_n^2}{(H_n^2 \sigma_{1,n}^{2-} + \sigma_{V,n}^2) \cdot (H_n^2 \sigma_{2,n}^{2-} + \sigma_{V,n}^2)} \right]}_{D_n} |\sigma_{2,n-1}^2 - \sigma_{1,n-1}^2| \end{aligned}$$

On a donc, $|\hat{X}_{1,n} - \hat{X}_{2,n}| \leq A_n \cdot |\hat{X}_{1,n-1} - \hat{X}_{2,n-1}| + B_n \cdot |\sigma_{2,n-1}^2 - \sigma_{1,n-1}^2|$

Avec,

$$\begin{cases} A_n &= |F_n + K_{2,n}H_nF_n| \\ B_n &= C_n \cdot D_n \\ C_n &= |Y_n - H_nf_n - h_n + H_nF_n\hat{X}_{2,n-1}| \\ D_n &= \frac{|H_n| F_n^2}{H_n^2 \sigma_{1,n}^{2-} + \sigma_{V,n}^2} + \frac{|\sigma_{2,n}^{2-} H_n| H_n^2 F_n^2}{(H_n^2 \sigma_{1,n}^{2-} + \sigma_{V,n}^2) \cdot (H_n^2 \sigma_{2,n}^{2-} + \sigma_{V,n}^2)} \\ \sigma_{1,n}^{2-} &= F_n^2 \cdot \sigma_{1,n-1}^2 + G_n^2 \cdot \sigma_{W,n}^2 \\ \sigma_{2,n}^{2-} &= F_n^2 \cdot \sigma_{2,n-1}^2 + G_n^2 \cdot \sigma_{W,n}^2 \end{cases}$$

$$\begin{aligned} |\sigma_{1,n}^2 - \sigma_{2,n}^2| &= |(1 - K_{1,n}H_n)F_n^2\sigma_{1,n-1}^2 + (K_{1,n} - K_{2,n})H_n \cdot G_n^2\sigma_{W,n}^2 - (1 - K_{2,n}H_n)F_n^2\sigma_{2,n-1}^2| \\ &\leq F_n^2 |\sigma_{2,n-1}^2 - \sigma_{1,n-1}^2| + |(K_{1,n} - K_{2,n})H_n \cdot G_n^2\sigma_{W,n}^2| \\ &\quad + |K_{1,n}H_nF_n^2\sigma_{1,n-1}^2 - K_{2,n}H_nF_n^2\sigma_{2,n-1}^2| \\ &\leq F_n^2 |\sigma_{2,n-1}^2 - \sigma_{1,n-1}^2| + |(K_{1,n} - K_{2,n})H_n \cdot G_n^2\sigma_{W,n}^2| \\ &\quad + |K_{1,n} - K_{2,n}| |H_nF_n^2\sigma_{2,n-1}^2| + |K_{1,n}H_nF_n^2| |\sigma_{1,n-1}^2 - \sigma_{2,n-1}^2| \\ &\leq \underbrace{F_n^2(1 - |K_{1,n}H_n|)}_{a_n} |\sigma_{2,n-1}^2 - \sigma_{1,n-1}^2| + \underbrace{|H_n(G_n^2\sigma_{W,n}^2 + F_n^2\sigma_{2,n-1}^2)|}_{b_n} |K_{1,n} - K_{2,n}| \\ &\leq (a_n + b_n \cdot D_n) |\sigma_{2,n-1}^2 - \sigma_{1,n-1}^2| \end{aligned}$$

Les densités des filtres optimaux sont telles que :

$$p_{1,n}(x) = \frac{1}{\sigma_{1,n}\sqrt{2\pi}} \exp \left[-\frac{(x - \hat{X}_{1,n})^2}{2\sigma_{1,n}^2} \right]$$

$$p_{2,n}(x) = \frac{1}{\sigma_{2,n}\sqrt{2\pi}} \exp \left[-\frac{(x - \hat{X}_{2,n})^2}{2\sigma_{2,n}^2} \right]$$

Lemme 24 (Théorème des accroissements finis) *Si f est continue sur un domaine fermé de \mathbb{R}^2 et si elle admet des dérivées partielles à l'intérieur du domaine, alors :*

$$f(x_0 + h, y_0 + k) - f(x_0, y_0) = h \cdot \frac{\partial f}{\partial x}(x_0 + \theta h, y_0 + \theta k) + k \cdot \frac{\partial f}{\partial y}(x_0 + \theta h, y_0 + \theta k), \quad 0 < \theta < 1.$$

Notons pour x fixé, $f(x_1, y_1^2) = \frac{1}{\sqrt{2\pi y_1^2}} \exp \left[-\frac{(x - x_1)^2}{2y_1^2} \right]$. D'après le Lemme 24,

$$f(\hat{X}_{1,n}, \sigma_{1,n}^2) - f(\hat{X}_{2,n}, \sigma_{2,n}^2) = (\hat{X}_{1,n} - \hat{X}_{2,n}) \cdot \frac{\partial f}{\partial \hat{X}}(\hat{X}_0, \sigma_0^2) + (\sigma_{1,n}^2 - \sigma_{2,n}^2) \cdot \frac{\partial f}{\partial \sigma^2}(\hat{X}_0, \sigma_0^2).$$

Avec $\hat{X}_0 \in [\min(\hat{X}_{1,n}; \hat{X}_{2,n}), \max(\hat{X}_{1,n}; \hat{X}_{2,n})]$ et $\sigma_0 \in [\min(\sigma_{1,n}^2; \sigma_{2,n}^2), \max(\sigma_{1,n}^2; \sigma_{2,n}^2)]$.

$$\begin{cases} \frac{\partial f}{\partial \hat{X}}(\hat{X}_0, \sigma_0^2) = \frac{x - \hat{X}_0}{\sigma_0^2} \cdot \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp \left[-\frac{(x - \hat{X}_0)^2}{2\sigma_0^2} \right] \\ \frac{\partial f}{\partial \sigma^2}(\hat{X}_0, \sigma_0^2) = \frac{1}{2\sigma_0^2 \sqrt{2\pi\sigma_0^2}} \left(\frac{(x - \hat{X}_0)^2}{\sigma_0^2} - 1 \right) \exp \left[-\frac{(x - \hat{X}_0)^2}{2\sigma_0^2} \right] \end{cases}$$

$$\begin{aligned} |p_{1,n}(x) - p_{2,n}(x)| &= \left| \frac{1}{\sqrt{2\pi^2\sigma_{1,n}^2}} \exp \left[-\frac{(x - \hat{X}_{1,n})^2}{2\sigma_{1,n}^2} \right] - \frac{1}{\sqrt{2\pi\sigma_{2,n}^2}} \exp \left[-\frac{(x - \hat{X}_{2,n})^2}{2\sigma_{2,n}^2} \right] \right| \\ &\leq |\hat{X}_{1,n} - \hat{X}_{2,n}| \cdot \left| \frac{\partial f}{\partial \hat{X}}(\hat{X}_0, \sigma_0^2) \right| + |\sigma_{1,n}^2 - \sigma_{2,n}^2| \cdot \left| \frac{\partial f}{\partial \sigma^2}(\hat{X}_0, \sigma_0^2) \right| \\ &\leq (A_n \cdot |\hat{X}_{1,n-1} - \hat{X}_{2,n-1}| + B_n \cdot |\sigma_{2,n-1}^2 - \sigma_{1,n-1}^2|) \cdot \left| \frac{\partial f}{\partial \hat{X}}(\hat{X}_0, \sigma_0^2) \right| \\ &\quad + |a_n + b_n \cdot D_n| \cdot |\sigma_{2,n-1}^2 - \sigma_{1,n-1}^2| \cdot \left| \frac{\partial f}{\partial \sigma^2}(\hat{X}_0, \sigma_0^2) \right| \\ &\leq |\hat{X}_{1,n-1} - \hat{X}_{2,n-1}| \left(A_n \left| \frac{\partial f}{\partial \hat{X}}(\hat{X}_0, \sigma_0^2) \right| \right) \\ &\quad + |\sigma_{2,n-1}^2 - \sigma_{1,n-1}^2| \left(B_n \left| \frac{\partial f}{\partial \sigma^2}(\hat{X}_0, \sigma_0^2) \right| + |a_n + b_n \cdot D_n| \right). \end{aligned}$$

Posons $\varphi_n(x) = \left| \frac{\partial f}{\partial \hat{X}}(\hat{X}_0, \sigma_0^2) \right|$, $\psi_n(x) = \left| \frac{\partial f}{\partial \sigma^2}(\hat{X}_0, \sigma_0^2) \right|$ et aussi,

$$\varphi_{n-1}(x) = \left(A_n \left| \frac{\partial f}{\partial \hat{X}}(\hat{X}_0, \sigma_0^2) \right| \right), \quad \psi_{n-1}(x) = \left(B_n \left| \frac{\partial f}{\partial \sigma^2}(\hat{X}_0, \sigma_0^2) \right| + |a_n + b_n \cdot D_n| \right)$$

$$\begin{aligned}
|p_{1,n}(x) - p_{2,n}(x)| &\leq \varphi_n(x) \cdot |\hat{X}_{1,n} - \hat{X}_{2,n}| + \psi_n(x) \cdot |\sigma_{1,n}^2 - \sigma_{2,n}^2| \\
&\leq \varphi_{n-1}(x) |\hat{X}_{1,n-1} - \hat{X}_{2,n-1}| + \psi_{n-1}(x) \cdot |\sigma_{1,n-1}^2 - \sigma_{2,n-1}^2| \\
&\dots\dots\dots \\
&\leq \varphi_0(x) |\hat{X}_{1,0} - \hat{X}_{2,0}| + \psi_0(x) |\sigma_{1,0}^2 - \sigma_{2,0}^2|
\end{aligned}$$

On a construit donc une suite recurrente de couples de fonctions $(\varphi_n(x), \psi_n(x))$ majorant la différences des densités des filtres optimaux déduits des conditions initiales différentes.

Cette différence est donc majorée en fonction des différences des paramètres des conditions initiales.

Par définition de l'optimalité des filtres, les estimations des paramètres des filtres s'approchent le mieux les valeurs réelles du signal non observé.

$$\begin{aligned}
\int_{\mathbb{R}} |p_{1,n}(x) - p_{2,n}(x)| dx &\leq |\hat{X}_{1,n} - \hat{X}_{2,n}| \cdot \int_{\mathbb{R}} \left| \frac{\partial f}{\partial \hat{X}}(\hat{X}_0, \sigma_0^2) \right| dx + |\sigma_{1,n}^2 - \sigma_{2,n}^2| \cdot \int_{\mathbb{R}} \left| \frac{\partial f}{\partial \sigma^2}(\hat{X}_0, \sigma_0^2) \right| dx \\
\lim_{n \rightarrow \infty} \int_{\mathbb{R}} |p_{1,n}(x) - p_{2,n}(x)| dx &= 0
\end{aligned}$$

Il y a donc une stabilité du filtre par rapport à la mesure initiale.

4.6 Stabilité de filtres particulières

La stabilité du filtrage particulière est étroitement liée à sa convergence basée sur la loi des grands nombres. Si on s'est trompé de condition initiale, c'est à dire si on a utilisé une mauvaise loi pour l'état initiale, les particules dans l'ensemble, restent dans la region des particules du vrai filtre. Ce resultat est acceptable si on a un nombre assez grand de particules.

$\{X_n, n \in \mathbb{N}\}$ est supposée une chaîne de Markov homogène .

Étant donné deux conditions initiales X_0 et \bar{X}_0 différentes de lois de probabilité respectives p_0 et \bar{p}_0 , le filtre particulière est stable si pour tout fonction mesurable bornée

$$\lim_{n \rightarrow \infty} \|p_n(\phi) - \bar{p}_n(\phi)\|_1 = 0.$$

p_n et \bar{p}_n sont les filtres déduits au vu de l'observation Y_n .

On simule pour chaque loi, N -particules ξ_0^1, \dots, ξ_0^N et $\bar{\xi}_0^1, \dots, \bar{\xi}_0^N$.

Le poids de chacune des particule est considéré égal à $\frac{1}{N}$.

À la première itération,

les filtres prédits déduits sont sont approximées par :

$$p_1^{N^-}(x) = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_0^i}(x)$$

$$\bar{p}_1^{N^-}(x) = \frac{1}{N} \sum_{i=1}^N \delta_{\bar{\xi}_0^i}(x)$$

On obtient les nouvelles particules en mutation par :

$$\xi_1^i = f(\xi_0^i) + W_1^i \text{ où } W_1^i \sim p_W(w)dw$$

$$\bar{\xi}_1^i = f(\bar{\xi}_0^i) + \bar{W}_1^i \text{ où } \bar{W}_1^i \sim p_W(w)dw$$

Et calculons les pondérations suivantes, compte tenu de l'observation Y_1 :

$$\omega_1^i = \frac{p_V(Y_1 - h(\xi_1^i))}{\sum_{j=1}^N p_V(Y_1 - h(\xi_1^j))}$$

$$\bar{\omega}_1^i = \frac{p_V(Y_1 - h(\bar{\xi}_1^i))}{\sum_{j=1}^N p_V(Y_1 - h(\bar{\xi}_1^j))}$$

Le nombre efficace des particules pour chaque filtre prédit est :

$$N_1^{eff} = \frac{1}{N \sum_{i=1}^N (\omega_1^i)^2} \quad \text{et} \quad \bar{N}_1^{eff} = \frac{1}{N \sum_{i=1}^N (\bar{\omega}_1^i)^2}$$

On trie les poids par ordre décroissant et on redistribue suivant la loi multinomiale les N_n^{eff} premières particules rangées déduites de la loi initiale p_0 et les \bar{N}_n^{eff} premières particules rangées déduites de la loi initiale \bar{p}_0 pour la correction.

On obtient alors les approximations des filtres corrigés suivants :

$$p_n^N(x) = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_n^i}(x)$$

$$\bar{p}_n^N(x) = \frac{1}{N} \sum_{i=1}^N \delta_{\bar{\xi}_n^i}(x)$$

À la n -ième itération,

on reprend les particules ξ_{n-1}^i et $\bar{\xi}_{n-1}^i$ sélectionnées dans :

$$p_{n-1}^N(x) = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_{n-1}^i}(x) \quad \text{et} \quad \bar{p}_{n-1}^N(x) = \frac{1}{N} \sum_{i=1}^N \delta_{\bar{\xi}_{n-1}^i}(x)$$

Dans la phase de mutation,

$$\xi_n^i = f(\xi_{n-1}^i) + W_n^i \text{ où } W_n^i \sim p_W(w)dw$$

$$\bar{\xi}_n^i = f(\bar{\xi}_{n-1}^i) + \bar{W}_n^i \text{ où } \bar{W}_n^i \sim p_W(w)dw$$

Dans la phase de pondération, compte tenu de l'observation Y_n :

$$\omega_n^i = \frac{p_V(Y_n - h(\xi_n^i))}{\sum_{j=1}^N p_V(Y_n - h(\xi_n^j))}$$

$$\bar{\omega}_n^i = \frac{p_V(Y_n - h(\bar{\xi}_n^i))}{\sum_{j=1}^N p_V(Y_n - h(\bar{\xi}_n^j))}$$

Les nombres efficaces de particules sont :

$$N_n^{eff} = \frac{1}{\sum_{i=1}^N (\omega_n^i)^2} \quad \text{et} \quad \bar{N}_n^{eff} = \frac{1}{\sum_{i=1}^N (\bar{\omega}_n^i)^2}$$

La résolution particulière repose sur l'adaptation de la loi des grands nombres pour construire $\mu_n(dx)$.

Lemme 25 (Loi forte des grands nombres) $(X_n)_{n \in \mathbb{N}^*}$ est une suite de variables aléatoires dans \mathbb{R}^d , indépendantes et identiquement distribuées de même loi que X . Alors pour presque tout $\omega \in \Omega$:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N [X_i - \mathbb{E}(X)] = 0$$

Lemme 26 (Théorème central limite) Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a i.i.d de moyenne μ et de variance σ^2 . Alors la distribution de \bar{X}_n approche celle de la loi $\mathcal{N}(0, 1)$ telle que pour tout $a < b$, on a :

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[a < \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} < b \right] = \mathbb{P}(a < Z < b)$$

où $Z \sim \mathcal{N}(0, 1)$ ou si $S_n = X_1 + \dots + X_n$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[a < \frac{S_n - n\mu}{\sigma\sqrt{n}} < b \right] = \mathbb{P}(a < Z < b)$$

Avec $\text{Var}(S_n) = n\sigma^2$ et $\mathbb{E}(S_n) = n\mu$, on a donc

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} \simeq Z.$$

$=\Delta$

Lemme 27 (Loi faible des grands nombres) Si \mathbb{P}_0 est une probabilité sur \mathbb{R}^d , il existe un espace $(\Omega, \mathcal{F}, \mathbb{P})$ sur lequel on peut définir une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires dans \mathbb{R}^d , indépendantes et identiquement distribuées, telles que $\mathbb{P}_{X_n} = \mathbb{P}_0$, pour tout n .

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires i.i.d de moyenne μ et de variance σ^2 . À un échantillon de taille n , $(X_1(\omega), \dots, X_n(\omega))$, on associe la moyenne empirique

$$\bar{X}_n(\omega) := \frac{1}{n} \sum_{i=1}^n X_i(\omega)$$

et la mesure empirique,

$$\mu_n^\omega := \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)}.$$

Alors pour tout $\varepsilon > 0$, on a :

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) = 0$$

ou encore dans L^2 ,

$$\lim_{n \rightarrow \infty} \|\bar{X}_n - \mu\|_2 = 0$$

La vitesse de convergence est telle que :

$$\mathbb{P}(|\bar{X}_n - \mu| \geq b) \leq \frac{\sigma^2}{nb^2}$$

Par définition, $(X^N)_{N \geq 1}$, une suite de processus converge uniformément vers X , un processus à valeurs dans \mathbb{R}^d sur $(\Omega, \mathcal{F}, \mathbb{P})$, X^N dans $L^p(\Omega, \mathcal{F}, \mathbb{P})$ et on note $X^N \xrightarrow[N \rightarrow +\infty]{L^p(\Omega, \mathcal{F}, \mathbb{P})} X$

$$\text{si } (\forall p > 0), \|X^N - X\|_p \stackrel{\Delta}{=} \sup_{t \geq 0} \|X_t^N - X_t\|_p \xrightarrow[N \rightarrow +\infty]{} 0$$

$$\lim_{N \rightarrow +\infty} \sup_{t \geq 0} \mathbb{E}(|X^N - X|^p) = \lim_{N \rightarrow +\infty} \sup_{t \geq 0} \int_{\Omega} |X^N(\omega) - X(\omega)|^p \cdot \mathbb{P}(d\omega) = 0$$

$$\text{et si } (p = 0), (\forall \varepsilon > 0), \sup_{t \geq 0} \mathbb{P}(\|X_t^N - X_t\|_{\mathbb{R}^d} > \varepsilon) \stackrel{\Delta}{=} \mathbb{P}(\|X^N - X\|_{\mathbb{R}^d} > \varepsilon) \xrightarrow[N \rightarrow +\infty]{} 0$$

Pour ξ_n variable aléatoire distribuée selon la loi $\mathbb{P}(X_{0:n} \in dx_{0:n} | Y_{1:n})$ [DEL MORAL et al., 2006], et ξ_n^i , N -variables aléatoires indépendantes conditionnellement à $(Y_{1:n})_{1 \leq i \leq N}$, de même loi que ξ_n , on a

$$\mathbb{E}(\xi_n | Y_{1:n}) = \mathbb{E}(X_n | Y_{1:n}) \mathbb{P} - ps$$

et d'après le lemme 27

$$\left\| \frac{1}{N} \sum_{i=1}^N \xi_n^i - \mathbb{E}(\xi_n | Y_{1:n}) \right\|_{L^2(\Omega, \mathcal{F}, \mathbb{P})} = \frac{1}{\sqrt{N}} \|X_n - \mathbb{E}(X_n | Y_{1:n})\|_{L^2(\Omega, \mathcal{F}, \mathbb{P})} \quad (4.4)$$

$$\lim_{N \rightarrow +\infty} \left\| \frac{1}{N} \sum_{i=1}^N \xi_n^i - \mathbb{E}(X_n | Y_{1:n}) \right\|_{L^2(\Omega, \mathcal{F}, \mathbb{P})} = 0$$

$$\lim_{N \rightarrow +\infty} \left\| \int_{\mathbb{R}^d} x_n \cdot \frac{1}{N} \sum_{i=1}^N \delta_{\xi_n^i}(dx_n) - \int_{\mathbb{R}^d} x_n \cdot \mathbb{P}(X_n \in dx_n | Y_{1:n}) \right\|_{L^2(\Omega, \mathcal{F}, \mathbb{P})} = 0 \quad (4.5)$$

Par une approximation de Monte-Carlo [CARON,2010] pour ϕ mesurable bornée :

$$\begin{aligned}\mathbb{E} [\phi(X_{0:n}|Y_{1:n})] &= \int_{\mathbb{R}^d} \phi(x_{0:n}) \mathbb{P}(X_{0:n} \in dx_{0:n} | Y_{1:n}) \\ &= \int_{\mathbb{R}^d} \phi(x_{0:n}) p(x_{0:n} | Y_{1:n}) dx_{0:n} \\ \int_{(\mathbb{R}^d)^t} x_t \frac{1}{N} \sum_{i=1}^N \delta_{\xi_t^i}(dx_t) &\xrightarrow[N \rightarrow +\infty]{L^p(\Omega, \mathcal{F}, \mathbb{P})} \int_{(\mathbb{R}^d)^t} x_t \mathbb{P}(X_{0:t} \in dx_{0:t} | Y_{1:t})\end{aligned}\quad (4.6)$$

La mesure de probabilité $\mathbb{P}(X_{0:t} \in dx_{0:t} | Y_{1:t})$ est approchée par un peigne de Dirac équi pondéré : $\mathbb{P}(X_{0:t} \in dx_{0:t} | Y_{1:t}) = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_t^i}(dx_t)$.

$\mathbb{P}(X_t \in dx_t | Y_{1:t})$ est la marginale de $\mathbb{P}(X_{0:t} \in dx_{0:t} | Y_{1:t})$ et X_t est l'extrémité de la trajectoire de $X_{0:t}$.

Plus précisément,

$$\frac{1}{N} \sum_{i=1}^N \xi_n^i \xrightarrow[N \rightarrow +\infty]{L^0(\Omega, \mathcal{F}, \mathbb{P})} \mathbb{E}(X_n | Y_{1:n})$$

D'où la proposition suivante,

Proposition 28 ([RUBENTHALER,2010], [DEL MORAL et al.,2006]) *Pour toute approximation particulière p_n^N du filtre optimal μ_n^N , il existe chaque n , C_n tel que :*

$$\sup_{\phi, \|\phi\|_\infty \leq 1} \mathbb{E} [|\mu_n(\phi) - p_n^N(\phi)|] \leq \frac{C_n}{\sqrt{N}} \quad (4.7)$$

Soit ϕ telle que $\|\phi\|_\infty \leq 1$, une fonction mesurable.

$$\begin{aligned}\mathbb{E} [|\mu_n(\phi) - p_n^N(\phi)|] &= \int_{\mathbb{R}^d} |\phi(x) \mu_n(dx) - \phi(x) p_n^N(x) dx| \\ &= \int_{\mathbb{R}^d} |\phi(x) [p_n(x) - p_n^N(x)]| dx \\ &= \int_{\mathbb{R}^d} |\phi(x) \left[p_n(x) - \frac{1}{N} \sum_{i=1}^N \delta_{\xi_n^i}(x) \right]| dx \\ &\leq \frac{C_n}{\sqrt{N}}\end{aligned}$$

Pour les approximations particulières des filtres déduits optimaux,

$$\begin{aligned}\|p_n^N(\phi) - \bar{p}_n^N(\phi)\|_1 &= \int_{\mathbb{R}^d} |\phi(x) p_n^N(x) - \phi(x) \bar{p}_n^N(x)| dx \\ &\leq \int_{\mathbb{R}^d} |\phi(x) p_n(x) - \phi(x) p_n^N(x)| dx + \int_{\mathbb{R}^d} |\phi(x) p_n(x) - \phi(x) \bar{p}_n^N(x)| dx \\ &\leq \mathbb{E} [|\mu_n(\phi) - p_n^N(\phi)|] + \mathbb{E} [|\mu_n(\phi) - \bar{p}_n^N(\phi)|] \\ &\leq \frac{C_n}{\sqrt{N}} + \frac{\bar{C}_n}{\sqrt{N}} = \frac{C_n + \bar{C}_n}{\sqrt{N}} = \frac{K_n}{\sqrt{N}}\end{aligned}$$

Pour un $\varepsilon > 0$, et pour n assez grand il existe $n_0 \in \mathbb{N}$ tel que pour tout $N \geq n_0$, $\frac{K_n}{\sqrt{N}} < \varepsilon$.

On a donc

$$\lim_{n \rightarrow \infty} \|p_n^N(\phi) - \bar{p}_n^N(\phi)\|_1 = 0$$

Ce qui montre la propriété d'oubli de la condition initiale du filtrage particulière.

4.7 Stabilité du filtrage par noyau de convolution

Le filtrage par noyau de convolution est une extension du filtrage particulière. Des régularisations interviennent soit pour les filtres prédits, le filtrage pré-régularisation, soit pour les filtres corrigés, le filtrage post-régularisation. En tout cas, le filtre optimal est régularisé.

Considérons les filtres optimaux déduits des lois initiales différentes p_0 et \bar{p}_0 par les approximations particulières.

$$p_n(x|y_1, \dots, y_n) = \frac{1}{N} \sum_{i=1}^N \delta_{\xi_i}(x)$$

$$\bar{p}_n(x|y_1, \dots, y_n) = \frac{1}{N} \sum_{i=1}^N \delta_{\bar{\xi}_i}(x)$$

Les filtres optimaux sont estimés par régularisation à partir de convolution par un noyau de Parzen-Rosenblatt :

Si p_n^N et \bar{p}_n^N les approximations particulières des filtres optimaux, on obtient par régularisation les approximations des densités des filtres optimaux $K_h * p_n^N$ et $K_h * \bar{p}_n^N$.

Pour ϕ une fonction mesurable bornée ou $\|\phi\|_\infty = 1$,

$$\begin{aligned} | \langle \phi, K_h * p_n^N \rangle - \langle \phi, K_h * \bar{p}_n^N \rangle | &= \left| \int_{\mathbb{R}^d} \phi(x) K_h * p_n^N(x) dx - \int_{\mathbb{R}^d} \phi(x) K_h * \bar{p}_n^N(x) dx \right| \\ &= \left| \int_{\mathbb{R}^d} \phi(x) [K_h * p_n^N(x) - K_h * \bar{p}_n^N(x)] dx \right| \\ &= \left| \int_{\mathbb{R}^d} \phi(x) K_h * (p_n^N(x) - \bar{p}_n^N(x)) dx \right| \\ &\leq \int_{\mathbb{R}^d} K_h(x) dx \left| \int_{\mathbb{R}^d} \phi(x) (p_n^N(x) - \bar{p}_n^N(x)) dx \right| \\ &\leq | \langle \phi, p_n^N \rangle - \langle \phi, \bar{p}_n^N \rangle | \end{aligned}$$

$$p_n^M(x|y_1, \dots, y_n) = \frac{p_{XY}^M(z)}{p_Y^M(y_1, \dots, y_n)} = \frac{\sum_{i=1}^M L_{h_M}(z - z_n^i)}{\sum_{i=1}^M K_{h_M}(y_1 - y_1^i, \dots, y_n - y_n^i)}$$

et

$$\bar{p}_n^M(x|y_1, \dots, y_n) = \frac{p_{XY}^M(z)}{p_Y^M(y_1, \dots, y_n)} = \frac{\sum_{i=1}^M L_{h_M}(z - \bar{z}_n^i)}{\sum_{i=1}^M K_{h_M}(y_1 - \bar{y}_1^i, \dots, y_n - \bar{y}_n^i)}$$

Avec les M -états dans la loi conjointe obtenus à partir de p_0 :

$$z_n^1, \dots, z_n^M, \text{ avec } z_n^i = (x_n^i, y_n^i).$$

où $z - z_n^i = (x - x_n^i, y_i - y_n^i)$.

et aussi avec les M -états dans la loi conjointe obtenus à partir de \bar{p}_0 :

$$\bar{z}_n^1, \dots, \bar{z}_n^M, \text{ avec } \bar{z}_n^i = (\bar{x}_n^i, \bar{y}_n^i).$$

où $z - \bar{z}_n^i = (x - \bar{x}_n^i, y_i - \bar{y}_n^i)$.

Pour toute fonction ϕ , mesurable bornée et telle que $\|\phi\|_\infty \leq 1$,

$$\begin{aligned} \|p_n^M(\phi) - \bar{p}_n^M(\phi)\|_1 &= \int_{\mathbb{R}^d} |\phi(x)(p_n^M(x|y_1, \dots, y_n) - \bar{p}_n^M(x|y_1, \dots, y_n))| dx = \\ &= \int_{\mathbb{R}^d} \left| \phi(x) \left(\frac{\sum_{i=1}^M L_{h_M}(z - z_n^i)}{\sum_{i=1}^M K_{h_M}(y_1 - y_1^i, \dots, y_n - y_n^i)} - \frac{\sum_{i=1}^M L_{h_M}(z - \bar{z}_n^i)}{\sum_{i=1}^M K_{h_M}(y_1 - \bar{y}_1^i, \dots, y_n - \bar{y}_n^i)} \right) \right| dx = \\ &= \int_{\mathbb{R}^d} \left| \phi(x) \left(\frac{\sum_{i=1}^M L_{h_M}((x - x_1^i, y_1 - y_1^i), \dots, (x - x_n^i, y_n - y_n^i))}{\sum_{i=1}^M K_{h_M}(y_1 - y_1^i, \dots, y_n - y_n^i)} - \right. \right. \\ &\quad \left. \left. - \frac{\sum_{i=1}^M L_{h_M}((x - \bar{x}_1^i, y_1 - \bar{y}_1^i), \dots, (x - \bar{x}_n^i, y_n - \bar{y}_n^i))}{\sum_{i=1}^M K_{h_M}(y_1 - \bar{y}_1^i, \dots, y_n - \bar{y}_n^i)} \right) \right| dx \end{aligned}$$

Par application dans la section précédente, on a bien un oubli de la condition initiale.

4.8 Conclusion

L'oubli de la condition initiale est une propriété importante dans les différentes méthodes de filtrage. Son étude permet de vérifier le bien-fondé du filtre optimal déduit à chaque instant.

Dans l'étude théorique du filtrage, la propriété markovienne du signal non observé est une caractéristique indispensable. Cette propriété est utilisée dans la majoration des mesures de Feynman-Kac construites pour déterminer les filtres optimaux. La faible ergodicité de la chaîne de Markov est une hypothèse introduite pour justifier l'oubli de la condition initiale.

Pour des cas particuliers comme le cas de bruits gaussiens, une solution analytique de filtrage construite à partir du théorème de Girsanov, la stabilité est vérifiée avec la norme de la variation totale; le cas de bruit d'observation uniforme donc quelconque, utilise plutôt la norme dans L^1 .

Dans le filtrage de Kalman, les expressions analytiques du filtre optimal permet de construire les filtres optimaux gaussiens. Comme la différence de deux variables gaussienne est une gaussienne, la stabilité est aisément vérifiée.

En poursuite de trajectoire, la correction du filtre prédit permet de s'approcher de la vraie trajectoires les trajectoires estimées par deux conditions initiales différentes.

En filtrage particulaire, il s'agit d'explorer l'espace d'état de façon aléatoire. Dans la stabilité du filtrage particulaire, la loi des grands nombres justifie la convergence dans l'échantillonnage des particules. Cette convergence est étroitement liée à l'oubli de la condition initiale. Une condition nécessaire dans l'étude est l'échantillonnage en très grands nombre de particules.

La stabilité du filtrage par noyau de convolution n'est qu'une conséquence immédiate de la stabilité du filtrage particulaire. La régularisation des filtres obtenus par le filtrage particulaire permet de mieux contrôler cette stabilité.

L'étude de l'oubli de la condition initiale est complétée par des applications dans lesquelles, l'échantillonnage des particules ne rencontre pas de difficultés particulières.

Conclusion de la deuxième partie

Dans cette deuxième partie, nous avons démontré l'existence d'une mesure avec laquelle on a construit le filtre. Cette existence de mesure tient son intérêt dans l'approche en théorie de probabilité. La théorie traite le cas de bruit d'observation gaussien dans un système markovien homogène. La solution analytique n'est pas explicitée convenablement en pratique mais une application numérique basée sur la méthode de Monte Carlo, permet une bonne approximation.

Dans les modèles à bruits quelconques, le filtrage particulaire est le mieux adapté. En particulier, avec un bruit d'observation uniforme, les particules de poids nuls sont automatiquement éliminées. Il suffit de redistribuer les particules restantes.

Un bruit d'observation quelconque peut être décomposé en mélange de bruits uniformes et on peut appliquer le filtrage particulaire mais la décomposition peut se faire en mélange de bruits gaussiens et dans ce cas on peut appliquer le filtrage de Kalman inodore.

L'oubli de la condition initiale est étudié pour les différents types de filtrage. En théorie, le filtrage particulaire introduit des théorèmes classiques de probabilités comme la loi des grands nombres ou le théorème de la limite centrale et filtrage de Kalman introduit le théorème des accroissements finis. Dans la pratique, l'oubli de la condition initiale est justifié dans les applications.

La résolution analytique est assez complexe dans le cas pratique. Des méthodes numériques de filtrage sont alors introduites. Chacune des méthodes a son inconvénient dans son application mais elle tire son avantage dans la facilité de mise en œuvre. Le filtrage de Kalman, le filtrage particulaire et le filtrage par noyau de convolution sont les principales méthodes appliquées.

Troisième partie
Applications et simulations

Introduction la troisième partie

Le filtrage d'un signal non observé est en général un problème ne possédant pas de solution analytique. Dans la théorie, des hypothèses sont introduites avec des théorèmes et des propositions. Chaque méthode de filtrage utilisée a ses particularités dans l'application. Des conditions d'application sur des exemples classiques sont nécessaires pour vérifier la propriété de filtrage comme l'oubli de la condition initiale.

Deux applications sont introduites dans cette partie :

D'une part la poursuite de trajectoire est proposée avec le filtrage de processus de Markov caché. En particulier, le filtrage de Kalman étendu et le filtrage de Uhlmann et Julier sont appliqués dans le problème classique de lancement de projectile.

D'autre part, le problème d'estimation paramétrique et de commande optimale qu'on rencontre dans le robotique. Ces problèmes sont étudiés avec l'application des filtres particuliers et les filtres par noyau de convolution.

Dans les deux applications, une modélisation de phénomènes en mécanique classique est nécessaire et la mise en œuvre des algorithmes correspondants sont exposés. Les filtres sont représentés soit par leurs densités dans la poursuite de la trajectoire, soit par des particules dans la commande optimale ou dans l'estimation paramétrique.

Pour chaque méthode de filtrage, l'oubli de la condition initiale est vérifié numériquement. Cette vérification de l'oubli de la condition initiale varie d'une méthode de filtrage à une autre. En particulier, la différence de densité entre deux filtres déduits de deux conditions initiales différentes peut être mesurée par la norme de L^1 . En filtrage particulier, cette différence est mesurée par la norme de la variation totale.

L'approximation des intégrales par la méthode de Monte Carlo est la base essentielle des simulations. Cette approximation est justifiée par la loi des grands nombres et du théorème de la limite centrale. Le nombre assez important des particules rend plus vraisemblable l'approximation mais ralentit l'exécution de l'algorithme.

Chapitre 5

Poursuite de trajectoire

Sommaire

5.1 Introduction	113
5.2 Modélisation	114
5.2.1 Modèle d'état	114
5.2.2 Modèle d'observation	117
5.3 Filtrage non linéaire	118
5.3.1 Algorithme	119
5.3.2 Oubli de la condition initiale	122
5.4 Filtrage de Kalman étendu	125
5.4.1 Algorithme	125
5.4.2 Oubli de la condition initiale	128
5.5 Filtrage de Uhlmann et Julier	130
5.5.2 Algorithme	131
5.5.3 Oubli de la condition initiale	132
5.6 Conclusion	134

5.1 Introduction

Considérons une trajectoire théorique définie par une loi de cinématique. Le projectile que suit cette trajectoire théoriquement, suit une autre trajectoire observée par des capteurs à des instants à intervalle régulier. On applique une méthode de filtrage pour approximer effectivement la trajectoire théorique et pouvoir redresser le projectile à partir d'une commande pour qu'il puisse suivre la trajectoire théorique.

En mécanique classique, un projectile suit une trajectoire nominale définie par :

$$\begin{cases} x(t) = V_{0x}t + x_0 \\ y(t) = -\frac{1}{2}g_0t^2 + V_{0y}t + y_0 \end{cases}$$

$\bar{x}(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}$ est la position du mobile à l'instant t .

$\bar{x}_0 = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$ est la position du mobile à l'instant 0.

$V(t) = \begin{bmatrix} V_x(t) \\ V_y(t) \end{bmatrix}$ est la vitesse du mobile à l'instant initiale t .

$\bar{V}_0 = \begin{bmatrix} V_{0x} \\ V_{0y} \end{bmatrix} = \begin{bmatrix} V_0 \cos \alpha \\ V_0 \sin \alpha \end{bmatrix}$ est la vitesse du mobile à l'instant initiale 0.
 g_0 est l'accélération de la pesanteur supposée constante.

Notons $X(t) = \begin{bmatrix} \bar{V}(t) \\ \bar{x}(t) \end{bmatrix} = \begin{bmatrix} V_x(t) \\ V_y(t) \\ x(t) \\ y(t) \end{bmatrix}$

le vecteur caractéristique du mouvement du mobile.

Valeurs numériques

L'espace de temps	: $t \in [-100; 100]$
Intervalle de temps	: $\tau = 10s$
La vitesse initiale	: $V_0 = 10ms^{-1}$
Accélération de la pesanteur	: $g_0 = 10m.s^{-2}$.
Inclinaison de la rampe de lancement	: $\alpha = \pi/8$.

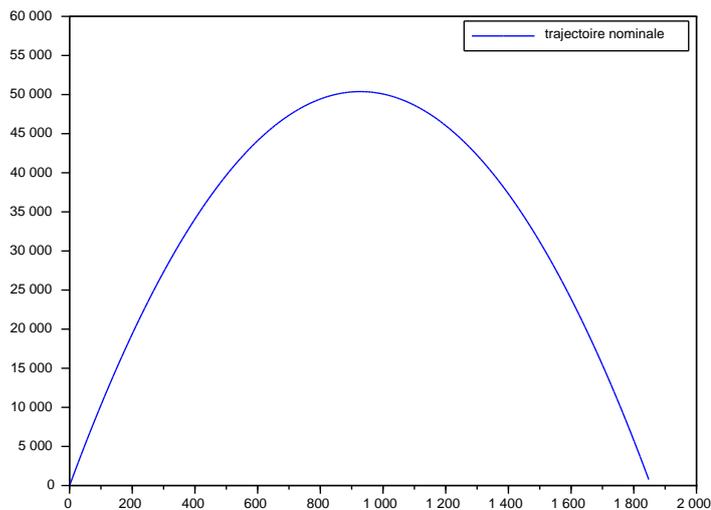


FIGURE 5.1 – Trajectoire nominale

En mécanique classique, le poids du projectile domine les autres forces comme la résistance de l'air. L'accélération de la pesanteur dépend de l'altitude mais on peut supposer qu'elle est constante et la trajectoire nominale est une parabole.

5.2 Modelisation

5.2.1 Modèle d'état

En réalité, cette trajectoire est perturbée pour des raisons inconnues par des bruits blancs gaussiens tels que :

$$\frac{d}{dt}X(t) = A.X(t) + B + \xi(t)$$

$$\text{où } A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \text{ avec la condition initiale } X_0 = \begin{bmatrix} V_{0x} \\ V_{0y} \\ x_0 \\ y_0 \end{bmatrix} \text{ et } B = \begin{bmatrix} 0 \\ -g \\ 0 \\ 0 \end{bmatrix}.$$

Par nécessité de discretisation, avec le schéma d'Euler, on peut considérer,

$$X_{n+1} = X_n + \int_{n\tau}^{(n+1)\tau} [A.X(t) + B(t)] dt + W_t$$

$1/\tau$ est la fréquence d'échantillonnage et $X_n = X(n\tau)$ et

$$X_{n+1} = X_n + \tau(A.X_n + B) + W_n$$

$$X_{n+1} = (I_4 + \tau.A)X_n + \tau.B + W_n$$

$$W_n \sim \mathcal{N}(0, \sigma_W^2.I_4) \text{ avec } \sigma_W > 0.$$

Dans la suite, supposons que la vitesse ne subit pas de perturbation.

La trajectoire est définie par le système d'équation horaire : pour $t \in [0; T]$

$$\begin{cases} x(t) &= V_0 \cos(\alpha).t \\ y(t) &= -\frac{g_0}{2}.t.(t - T) + V_0 \sin(\alpha).t \end{cases}$$

$$\text{À l'instant } t_n, \begin{cases} x(t_n) &= V_0 \cos(\alpha).t_n \\ y(t_n) &= -\frac{g_0}{2}.t_n.(t_n - T) + V_0 \sin(\alpha).t_n \end{cases}$$

En notant $x(t_n) = x_n$ et $y(t_n) = y_n$, on a

$t_n = n.\tau$ où $\tau = t_{n+1} - t_n$ est la période d'échantillonnage,

$$\begin{cases} x_{n+1} - x_n &= V_0 \cos(\alpha).(t_{n+1} - t_n) \\ y_{n+1} - y_n &= -\frac{g_0}{2}.[t_{n+1}(t_{n+1} - T) - t_n(t_n - T)] + V_0 \sin(\alpha).(t_{n+1} - t_n) \end{cases}$$

$$\begin{cases} x_{n+1} - x_n &= V_0 \cos(\alpha).(t_{n+1} - t_n) \\ y_{n+1} - y_n &= -\frac{g_0}{2}.[t_{n+1}^2 - t_n^2 - T.(t_{n+1} - t_n)] + V_0 \sin(\alpha).(t_{n+1} - t_n) \end{cases}$$

$$t_{n+1}^2 - t_n^2 = (t_{n+1} - t_n)(t_{n+1} + t_n) = \tau[(n+1)\tau + n\tau] = (2n+1)\tau^2$$

$$\begin{cases} x_{n+1} &= x_n + V_0 \cos(\alpha).\tau \\ y_{n+1} &= y_n - \frac{g_0}{2}.[(2n+1)\tau^2 - T.\tau] + V_0 \sin(\alpha).\tau \end{cases}$$

$$\begin{cases} x_{n+1} &= x_n + V_0 \cos(\alpha).\tau \\ y_{n+1} &= y_n - \frac{g_0.\tau}{2}.[(2n+1)\tau - T] + V_0 \sin(\alpha).\tau \end{cases}$$

Posons alors,

$$\begin{cases} x_{n+1} &= g_1(x_n, y_n) \\ y_{n+1} &= g_2(x_n, y_n) \end{cases} \text{ avec } \begin{cases} g_1(x_n, y_n) &= x_n + V_0 \cos(\alpha).\tau \\ g_2(x_n, y_n) &= y_n - \frac{g_0.\tau}{2}.[(2n+1)\tau - T] + V_0 \sin(\alpha).\tau \end{cases}$$

Avec l'équation matricielle,

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_n \\ y_n \end{bmatrix} + \begin{bmatrix} V_0 \cos(\alpha) \cdot \tau \\ -\frac{g_0 \cdot \tau}{2} \cdot [(2n+1)\tau - T] + V_0 \sin(\alpha) \cdot \tau \end{bmatrix}.$$

$$X_{n+1} = A \cdot X_n + B \text{ où } A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ et } B = \begin{bmatrix} V_0 \cos(\alpha) \cdot \tau \\ -\frac{g_0 \cdot \tau}{2} \cdot [(2n+1)\tau - T] + V_0 \sin(\alpha) \cdot \tau \end{bmatrix}.$$

$$\text{Avec la perturbation, } \begin{cases} x_{n+1} = x_n + V_0 \cos(\alpha) \cdot \tau + \sigma_W W_n \\ y_{n+1} = y_n - \frac{g_0 \cdot \tau}{2} \cdot [(2n+1)\tau - T] + V_0 \sin(\alpha) \cdot \tau + \sigma_W W_n \end{cases}$$

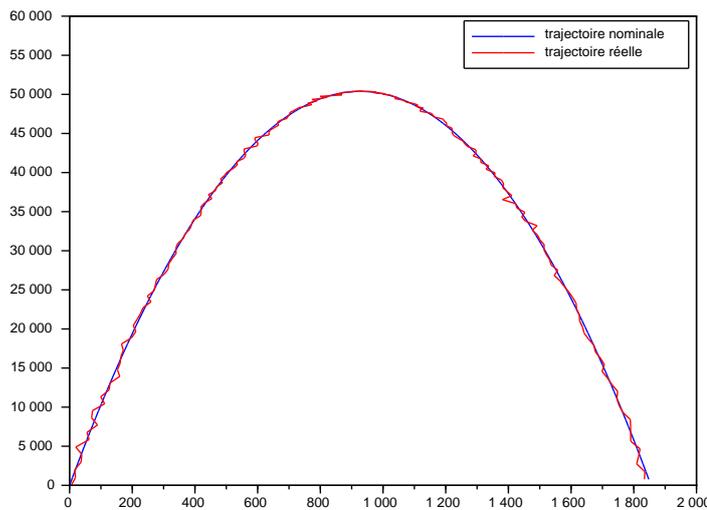
$$W_n \sim \mathcal{N}(0, 1) \text{ avec } \sigma_W > 0 \text{ et } X_{n+1} = A \cdot X_n + B + \sigma_W W_n.$$

$$\text{ou } X_{n+1} = G_n(X_n) + \sigma_W W, \text{ tel que } W \sim \mathcal{N}(0, I_2) \text{ avec } \sigma_W > 0.$$

$$G_n(x, y) = \begin{bmatrix} g_{1,n}(x, y) \\ g_{2,n}(x, y) \end{bmatrix} \text{ avec } \begin{cases} g_{1,n}(x, y) = x + V_0 \cos(\alpha) \cdot \tau \\ g_{2,n}(x, y) = y - \frac{g_0 \cdot \tau}{2} \cdot [(2n+1)\tau - T] + V_0 \sin(\alpha) \cdot \tau \end{cases}$$

Valeurs numériques

Vitesse initiale	: $V_0 = 10m.s^{-1}$.
Accélération de la pesanteur	: $g_0 = 10m.s^{-2}$.
Nombre de mesures	: $T = 65$
periode	: $\tau = 10s$.
Angle de lancement	: $\alpha = \pi/8$.
Écart type de perturbation	: $\sigma_W = 10$.



La trajectoire réelle est obtenue en perturbant la trajectoire nominale. Cette trajectoire est en réalité non observée. Cette trajectoire suit pratiquement une parabole.

FIGURE 5.2 – Trajectoire réelle non observée

5.2.2 Modèle d'observation

Un capteur positionné dans le plan, à la distance par rapport au point de lancement O du mobile telle $\vec{OA} = \begin{bmatrix} a \\ b \end{bmatrix}$, repère le mobile avec les coordonnées polaires.

Notons $Y_1(t) = \begin{bmatrix} x_1(t) \\ y_1(t) \end{bmatrix}$ le vecteur d'observation à chaque instant t .

$$\begin{cases} x_1(t) = x(t) - a = r(t) \cdot \cos(\theta(t)) \\ y_1(t) = y(t) - b = r(t) \cdot \sin(\theta(t)) \end{cases}$$

$$\text{ou } \begin{cases} r(t) = \sqrt{(x(t) - a)^2 + (y(t) - b)^2} \\ \theta(t) = \arctan \left[\frac{y(t) - b}{x(t) - a} \right] \end{cases}$$

L'équation d'observation est non linéaire $Y(t) = \begin{bmatrix} r(t) \\ \theta(t) \end{bmatrix} = h(X(t))$.

$$h(X(t)) = h(x(t), y(t)) = \begin{bmatrix} h_1(X(t)) \\ h_2(X(t)) \end{bmatrix}$$

$$\text{avec } \begin{cases} h_1(x(t), y(t)) = \sqrt{(x(t) - a)^2 + (y(t) - b)^2} \cdot \cos \left(\arctan \left[\frac{y(t) - b}{x(t) - a} \right] \right) + a \\ h_2(x(t), y(t)) = \sqrt{(x(t) - a)^2 + (y(t) - b)^2} \cdot \sin \left(\arctan \left[\frac{y(t) - b}{x(t) - a} \right] \right) + b \end{cases}$$

Les mesures d'observations sont perturbées par un bruit blanc non nécessairement gaussien.

$$Y(t) = h(X(t)) + V(t)$$

Les observations sont mesurées périodiquement en chaque instant t_n ,

$$X(t_n) = \begin{bmatrix} x(t_n) \\ y(t_n) \end{bmatrix}, Y(t_n) = \begin{bmatrix} r(t_n) \\ \theta(t_n) \end{bmatrix}, Y(t_n) = h(X(t_n)) + V(t_n)$$

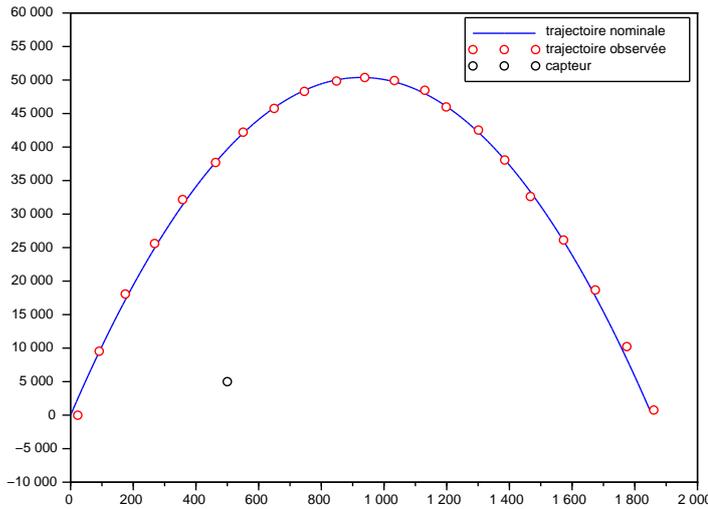
$$X_n = \begin{bmatrix} x_n \\ y_n \end{bmatrix}, Y_n = \begin{bmatrix} r_n \\ \theta_n \end{bmatrix}, Y_n = h(X_n) + V_n$$

$$\text{où } \begin{cases} r_n = \sqrt{(x_n - a)^2 + (y_n - b)^2} \\ \theta_n = \arctan \left[\frac{y_n - b}{x_n - a} \right] \end{cases}$$

$$V_n = \begin{bmatrix} V_n^1 \\ V_n^1 \end{bmatrix} \sim \mathcal{N}(0, \sigma_V^2 \cdot I_2) \text{ avec } \sigma_V > 0 \text{ ou } V_n \sim \mathcal{U}([-a; +a] \times [-a; +a]), a > 0.$$

Valeurs numériques

vitesse initiale	: $V_0 = 10m.s^{-1}$.
Accélération de la pesanteur	: $g_0 = 10m.s^{-2}$.
Nombre de mesures	: 21
periode	: $\tau = 10s$.
Angle de lancement	: $\alpha = \pi/8$.
Position du capteur	: $A(500; 5000)$.
Erreur de mesure d'observation	: $\mathcal{U}[-10; 10]$.



La trajectoire observée suit pratiquement la trajectoire nominale qui est une parabole. Les observations sont captées en des instants à intervalle régulier. La position du capteur est dans le plan de la trajectoire nominale.

FIGURE 5.3 – Trajectoire observée et position d'un capteur

5.3 Filtrage non linéaire

Supposons que tous les bruits sont gaussiens. On a le modèle,

$$\begin{cases} Y_n &= h(X_n) + V_n, & V_n \sim \mathcal{N}(0, Q_V) \\ X_{n+1} &= g(X_n) + W_n, & W_n \sim \mathcal{N}(0, Q_W) \end{cases}$$

$$g(X_n) = g(x_n, y_n) = \begin{cases} g_1(x_n, y_n) &= x_n + V_0 \cos(\alpha) \cdot \tau \\ g_2(x_n, y_n) &= y_n - \frac{g_0 \cdot \tau}{\cos(\alpha)} x_n - \frac{1}{2} g_0 \cdot \tau^2 + V_0 \sin(\alpha) \cdot \tau \end{cases}$$

$$h(X_n) = h(x_n, y_n) = \begin{cases} h_1(x_n, y_n) &= \sqrt{(x_n - a)^2 + (y_n - b)^2} \\ h_2(x_n, y_n) &= \arctan \left[\frac{y_n - b}{x_n - a} \right] \end{cases}$$

Les densités des bruits sont :

$$f_V(x, y) = \frac{1}{2\pi\sigma_V} \exp \left[-\frac{x^2 + y^2}{2\sigma_V^2} \right]$$

$$f_W(x, y) = \frac{1}{2\pi\sigma_W} \exp\left[-\frac{x^2 + y^2}{2\sigma_W^2}\right]$$

Avec les notations,

$$\mu_n^-(dx, dy) = \mathbb{P}(X_n \in (dx, dy) | \mathcal{Y}_{n-1}) = \mathbb{P}_{X_n}(dx, dy | \mathcal{Y}_{n-1}) = p_n^-(x, y) dx \cdot dy.$$

$$\mu_n(dx, dy) = \mathbb{P}(X_n \in (dx, dy) | \mathcal{Y}_n) = \mathbb{P}_{X_n}(dx, dy | \mathcal{Y}_n) = p_n(x, y) dx \cdot dy.$$

Avec la densité f_W de W_n , la prédiction est :

$$p_n^-(x, y) = \int_{\mathbb{R}} \int_{\mathbb{R}} f_W(x - g_1(x', y'), y - g_2(x', y')) p_{n-1}(x', y') dx' \cdot dy'$$

et la correction : $p_n(x, y) = C_n Z(x, y, Y_n^1, Y_n^2) \cdot p_n^-(x, y)$

$$C_n^{-1} := \int_{\mathbb{R}} \int_{\mathbb{R}} Z(x, y, Y_n^1, Y_n^2) \left[\int_{\mathbb{R}} \int_{\mathbb{R}} f_W(x - g_1(x', y'), y - g_2(x', y')) \cdot p_{n-1}(x', y') dx' \cdot dy' \right] dx \cdot dy$$

$$\text{avec } Z(x, y, Y_n^1, Y_n^2) = \exp\left[\frac{h_1(x, y) \cdot Y_n^1 + h_2(x, y) \cdot Y_n^2}{\sigma_V^2} - \frac{h_1(x, y)^2 + h_2(x, y)^2}{2\sigma_V^2}\right]$$

Avec les approximations de Monte-Carlo des intégrales, compte tenu de l'invariance des mesures de Lebesgue par translation, on a :

$$\int_{\mathbb{R}^d} f(x) \cdot g_X(x - a) \cdot dx \simeq \frac{1}{N} \sum_{i=1}^N f(X_i + a), \text{ telle que } X_i \sim g_X(x) \cdot dx$$

$$p_n^-(x, y) \simeq \frac{1}{N_1} \sum_{i=1}^{N_1} f_W(x - g_1(X_{i,n-1}^1, X_{i,n-1}^2), y - g_2(X_{i,n-1}^1, X_{i,n-1}^2))$$

$$(X_{i,n-1}^1, X_{i,n-1}^2) \sim p_{n-1}(x, y) dx \cdot dy \text{ et } (W_j^1, W_j^2) \sim f_W(x, y) dx \cdot dy$$

$$\frac{1}{C_n} \simeq \frac{1}{N_2} \sum_{j=1}^{N_2} \left[\frac{1}{N_1} \sum_{i=1}^{N_1} Z(W_j^1 + g_1(X_{i,n-1}), W_j^2 + g_2(X_{i,n-1}), Y_n^1, Y_n^2) \right].$$

5.3.1 Algorithme

$$\text{Posons } \begin{cases} h_1(x, y) &= \sqrt{(x-a)^2 + (y-b)^2} \\ h_2(x, y) &= \arctan\left[\frac{y-b}{x-a}\right] \end{cases}$$

$$h(x, y) Y_n = h_1(x, y) r_n + h_2(x, y) \theta_n, \quad (x, y) \in \mathbb{R}^2.$$

$$h^2(x, y) = h_1^2(x, y) + h_2^2(x, y), \quad (x, y) \in \mathbb{R}^2.$$

$$Z(U_i, Y_n) = \exp\left[\frac{h_1(U_i) Y_n^1 + h_2(U_i) Y_n^2}{\sigma_V^2} - \frac{h_1(U_i)^2 + h_2(U_i)^2}{2\sigma_V^2}\right]$$

Algorithme filtre optimal

Données : h_1, h_2, g_1, g_2 : fonctions ; N, N_1, N_2 : entiers ; f_v, f_W : densités ;
 $p_0(x, y)$ densité initiale (Y_n^1, Y_n^2) observations

Résultats : $q(x, y)$ le filtre prédit et $p(x, y)$ le filtre corrigé

DEBUT

-initialisation de $q(x, y)$ à $p_0(x, y)$
 -initialisation de $p(x, y)$ à $p_0(x, y)$

pour $n = 1$ à N faire

$$Z(x, y, Y_n^1, Y_n^2) = \exp \left[\frac{h_1(x, y)Y_n^1 + h_2(x, y)Y_n^2}{\sigma_V^2} - \frac{h_1^2(x, y) + h_2^2(x, y)}{2\sigma_V^2} \right]$$

pour $j = 1$ à N_2 faire

simuler (W_j^1, W_j^2) suivant $f_W(x, y).dxdy$

pour $i = 1$ à N_1 faire

simuler $(X_{i,n-1}^1, X_{i,n-1}^2)$ suivant $p_{n-1}(x, y)dx.dy$

$$\begin{aligned} u_i^1(x) &= x - g_1(X_{i,n-1}^1, X_{i,n-1}^2) \\ u_i^2(y) &= y - g_2(X_{i,n-1}^1, X_{i,n-1}^2) \\ v_{i,j}^1 &= W_j^1 + g_1(X_{i,n-1}^1, X_{i,n-1}^2) \\ v_{i,j}^2 &= W_j^2 + g_2(X_{i,n-1}^1, X_{i,n-1}^2) \end{aligned}$$

fin pour i

$$p_n^-(x, y) = \sum_{i=1}^{N_1} \frac{1}{N_1} f_W(u_i^1(x), u_i^2(y))$$

fin pour j

$$C_n^{-1} = \frac{1}{N_2} \sum_{j=1}^{N_2} \left[\frac{1}{N_1} \sum_{i=1}^{N_1} Z(v_{i,j}^1, v_{i,j}^2, Y_n^1, Y_n^2) \right]$$

$$p_n(x, y) = C_n \cdot Z(x, y, Y_n^1, Y_n^2) \cdot p_n^-(x, y)$$

fin pour n

$$p(x, y) = p_n(x, y)$$

FIN.

TABLE 5.1 – Algorithme filtre optimal

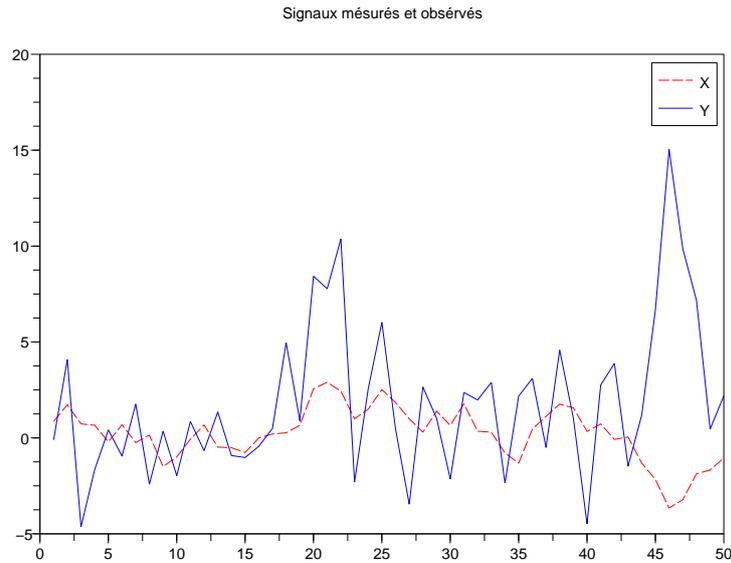


FIGURE 5.4 – Signaux mesurés et observés

$$\begin{cases} x_{n+1} = x_n + V_0 \cos(\alpha)\tau + \sigma_W W_n \\ y_{n+1} = y_n - \frac{1}{2}g_0\tau^2 + V_0 \sin(\alpha)\tau + \sigma_W W_n \end{cases}$$

$$W_n \sim \mathcal{N}(0, 1) \text{ avec } \sigma_W > 0.$$

$$f_W(x, y) = f_W(w) = \frac{1}{2\pi\sigma_W} \exp\left[-\frac{|w|^2}{2\sigma_W^2}\right] = \frac{1}{2\pi\sigma_W} \exp\left[-\frac{x^2 + y^2}{2\sigma_W^2}\right]$$

Valeurs numériques

La loi initiale est gaussienne centrée de matrice de covariance : $\sigma \cdot I_2$.

Les plages des valeurs sont telles que :

$$-4 \leq x \leq 4 \text{ et } -4 \leq y \leq 4.$$

Le paramètre est $\sigma = 1$.

Les densités sont dans \mathbb{R}^2 et elles sont représentées graphiquement. Les intégrales sont approximées par la méthode de Monte Carlo.

X_0 suit une loi gaussienne et sa fonction de repartition permet de simuler les valeurs dans p_0 pour déterminer le filtre prédit p_1^- .

Dans la Figure 5.6, la densité du filtre prédit p_1^- est gaussienne centrée de variance plus grande que pour la densité p_0 . La densité du filtre corrigé p_1 est gaussienne non centrée de même variance que p_1^- .

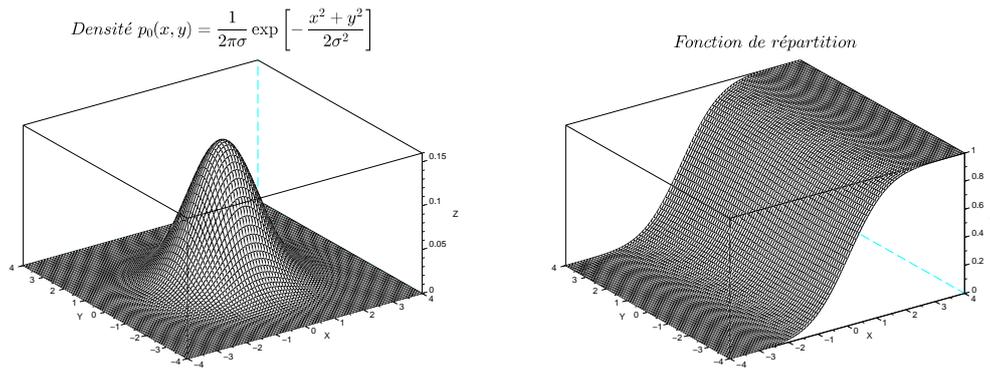


FIGURE 5.5 – Densité et fonction de répartition de la loi initiale

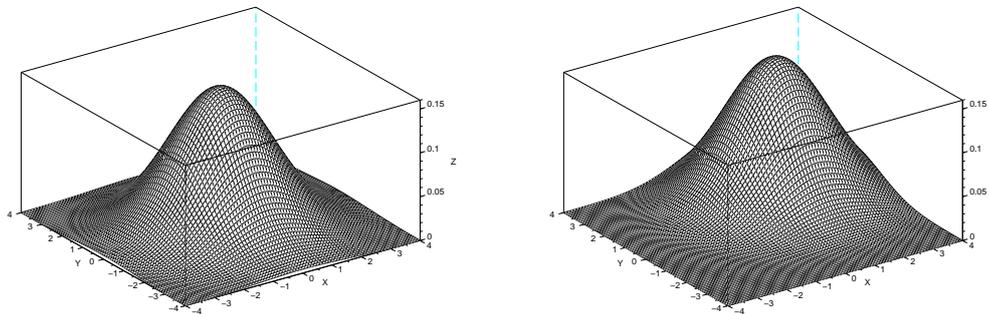


FIGURE 5.6 – Densité du filtre prédit et du filtre corrigé

Pour voir l'évolution des filtres prédits et corrigés, on projette les densités dans un plan.

Dans la Figure 5.7, les filtres prédits gardent la même espérance et s'aplatissent progressivement et les filtres corrigés ont des moyennes variables dues à l'introduction de l'observation et de variances de plus en plus grandes.

5.3.2 Oubli de la condition initiale

Considérons deux filtres optimaux déduits de deux différentes lois initiales p_0 et \bar{p}_0 et calculons l'erreur

$$\varepsilon_n = \int_{\mathbb{R}^d} |p_n(x) - \bar{p}_n(x)| dx.$$

Il y a oubli de la condition initiale si $\lim_{n \rightarrow \infty} \varepsilon_n = 0$. En général, les filtres optimaux ne sont pas explicites par des densités. Dans ce cas, il est préférable d'utiliser la norme de la variation totale telle que :

$$\lim_{n \rightarrow \infty} \sup_{\phi} \| \langle \mu_n, \phi \rangle - \langle \bar{\mu}_n, \phi \rangle \|_{VT} = 0$$

où ϕ est une fonction test mesurable telle que $\|\phi\|_{\infty} = 1$ et $\langle \mu_n, \phi \rangle = \int_{\mathbb{R}^d} \phi(x) \mu_n(dx)$ et $\langle \bar{\mu}_n, \phi \rangle = \int_{\mathbb{R}^d} \phi(x) \bar{\mu}_n(dx)$.

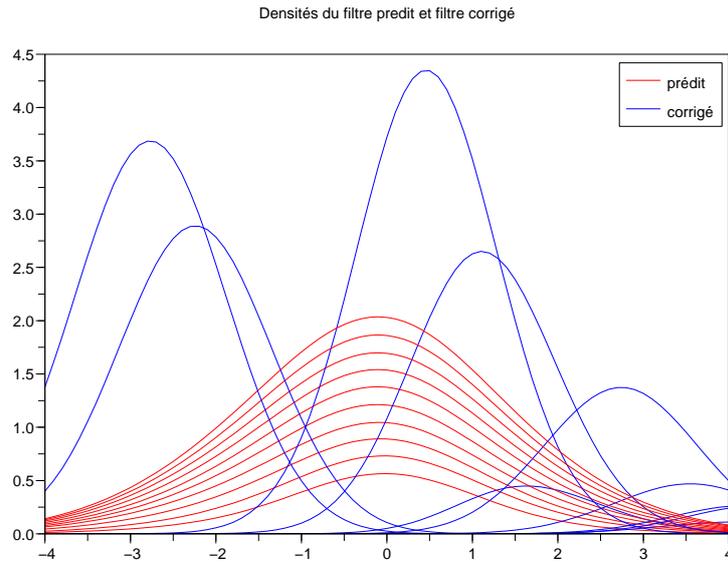
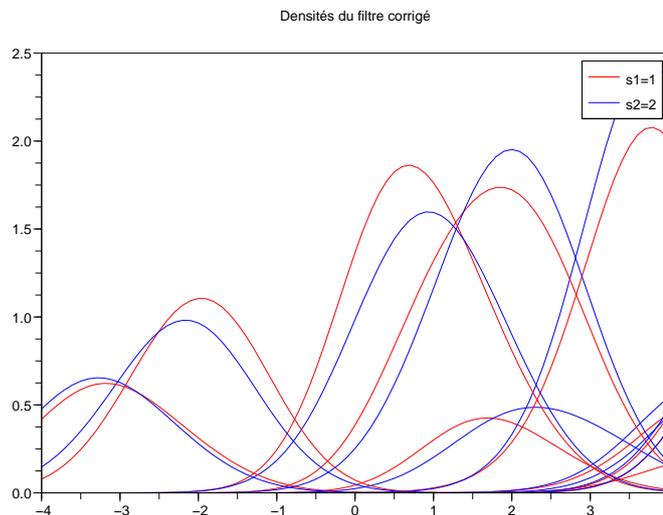


FIGURE 5.7 – Filtrage prédit et corrigé

La stabilité est évaluée par l'erreur ε_n à l'instant n . Cette erreur décroît rapidement au cours du temps jusqu'à s'annuler. Cette convergence vers 0 justifie la stabilité du filtre. La Table 5.2 donne l'algorithme de calcul de l'erreur.



On considère deux filtres optimaux issus de deux conditions initiales de variances 1 et 2. Les filtres optimaux à l'instant n sont gaussiens et ont la même allure et se rapprochent.

FIGURE 5.8 – filtres optimaux pour 2 conditions initiales différentes

Algorithme de l'erreur

Données : p_0 et \bar{p}_0 : lois initiales ; N, N_2 : entiers ; g : densité

Résultats : ε l'erreur

DEBUT

$\varepsilon(1) = 0$

pour $n = 1$ à N faire

calculer $p_n(x)$ et $\bar{p}_n(x)$

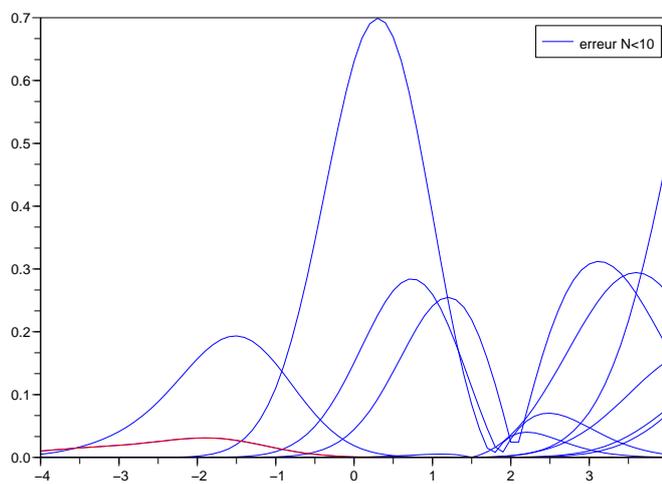
$\alpha_n(x) = |p_n(x) - \bar{p}_n(x)|$

$\varepsilon(n) = \int_{\mathbb{R}} \alpha_n(x) dx$

fin pour n

FIN

TABLE 5.2 – Algorithme de calcul de l'erreur



L'erreur linéaire représentant la norme dans L^1 de la différence des filtres est une gaussienne et s'aplatit très vite. L'aire définie par la courbe de la différence des filtres tend vers 0. La courbe en rouge marque l'aire de l'erreur après 10 itérations.

FIGURE 5.9 – Erreur linéaire à $N=10$

5.4 Filtrage de Kalman étendu

Considérons l'application sur la poursuite de trajectoire

On a le modèle d'état :

$$X_{n+1} = f_n(X_n) + W_n$$

$$\begin{cases} x_{n+1}^1 &= x_n^1 + V_0 \cos(\alpha) \cdot \tau + W_n^1 \\ x_{n+1}^2 &= x_n^2 - \frac{g_0 \cdot \tau}{2} \cdot [(2n+1)\tau - T] + V_0 \sin(\alpha) \cdot \tau + W_n^2 \\ f_{1,n}(x_n^1, x_n^2) &= x_n^1 + V_0 \cos(\alpha) \cdot \tau \\ f_{2,n}(x_n^1, x_n^2) &= x_n^2 - \frac{g_0 \cdot \tau}{2} \cdot [(2n+1)\tau - T] + V_0 \sin(\alpha) \cdot \tau \end{cases}$$

et le modèle d'observation :

$$Y_n = h_n(X_n) + V_n$$

$$\begin{cases} y_n^1 &= h_{1,n}(x_n^1, x_n^2) + V_n^1 = \sqrt{(x_n^1 - a)^2 + (x_n^2 - b)^2} + V_n^1 \\ y_n^2 &= h_{2,n}(x_n^1, x_n^2) + V_n^2 = \arctan \left[\frac{x_n^2 - b}{x_n^1 - a} \right] + V_n^2 \end{cases}$$

W est un bruit d'état gaussien de densité :

$$f_W(w) = \frac{1}{(2\pi)^{d/2} |\det Q^W|^{1/2}} \exp \left[-\frac{1}{2} \langle w, (Q^W)^{-1} \cdot w \rangle \right]$$

V est un bruit d'observation gaussien de densité :

$$f_V(v) = \frac{1}{(2\pi)^{d/2} |\det Q^V|^{1/2}} \exp \left[-\frac{1}{2} \langle v, (Q^V)^{-1} \cdot v \rangle \right]$$

La Table 5.3 donne les calculs dans l'algorithme de Kalman :

5.4.1 Algorithmes

$\bar{x}_n = (\bar{x}_n^1, \bar{x}_n^2)$ est une approximation de l'état à la n -itération. Les fonctions f_n et h_n sont dérivables. On linéarise f_n et h_n autour de \bar{x}_n tels que :

$$\begin{aligned} f_n(x) &\simeq f_n(\bar{x}_n) + \vec{\nabla} f_n(\bar{x}_n)(x - \bar{x}_n), \\ h_n(x) &\simeq h_n(\bar{x}_n) + \vec{\nabla} h_n(\bar{x}_n)(x - \bar{x}_n) \end{aligned}$$

On obtient le système linéarisé suivant :

$$\begin{cases} Y_n &= H_n(X_n - \bar{x}_n) + \tilde{h}_n + V_n \\ X_{n+1} &= F_n(X_n - \bar{x}_n) + \tilde{f}_n + W_n \end{cases}$$

où $f_n(\bar{x}_n) = \tilde{f}_n$, $F_n = \vec{\nabla} f_n(\bar{x}_n)$, $h_n(\bar{x}_n) = \tilde{h}_n$ et $H_n = \vec{\nabla} h_n(\bar{x}_n)$.

Algorithme du filtrage de Kalman :

Debut Q_n^V et Q_n^W : matrices de covariance des bruits, f_n , h_n

Résultat filtre prédit : \hat{X}_n^- , filtre corrigé : \hat{X}_n

DEBUT

initialisation : $\bar{X}_0^- = \hat{X}_0$ et $R_0^- = Q_0$

pour $n = 1 : N$

Linéarisations dans l'équation d'état

$$\begin{aligned}\tilde{f}_n &= f_n(\hat{X}_{n-1}) \\ F_n &= \vec{\nabla} f_n(\hat{X}_{n-1})\end{aligned}$$

Moyenne et covariance du filtre prédit

$$\begin{aligned}\hat{X}_n^- &= F_n \hat{X}_{n-1} + \tilde{f}_n \\ R_n^- &= F_n R_{n-1} F_n^t + Q_n^W\end{aligned}$$

Linéarisations dans l'équation d'observation

$$\begin{aligned}\tilde{h}_n &= h_n(\hat{X}_n^-) \\ H_n &= \vec{\nabla} h_n(\hat{X}_n^-)\end{aligned}$$

Covariance de l'innovation

$$Q_n^I = H_n R_n^- H_n^t + Q_n^V$$

Gain de Kalman

$$K_n = R_n^- H_n^t (Q_n^I)^{-1}$$

Moyenne et covariance d filtre corrigé

$$\begin{aligned}\hat{X}_n &= \hat{X}_n^- + K_n [Y_n - (H_n \hat{X}_n^- + \tilde{h}_n)] \\ R_n &= [I - K_n H_n] R_n^-\end{aligned}$$

fin pour n

FIN

TABLE 5.3 – Algorithme filtre de Kalman

Le gradient de f_n est : $\vec{\nabla} f_n(x^1, x^2) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ et

le gradient de h_n est :

$$\vec{\nabla} h_n(x^1, x^2) = \begin{bmatrix} \frac{x^1}{\sqrt{(x^1 - a)^2 + (x^2 - b)^2}} & \frac{x^2}{\sqrt{(x^1 - a)^2 + (x^2 - b)^2}} \\ \frac{-(x^2 - b)}{(x^1 - a)^2 + (x^2 - b)^2} & \frac{x^1 - a}{(x^1 - a)^2 + (x^2 - b)^2} \end{bmatrix}$$

En notant, $\tilde{X}_n = X_n - \bar{x}_n$, $\tilde{f}_n = f_n - \bar{x}_{n+1}$, on a un nouveau système linéaire :

$$\begin{cases} Y_n &= H_n \tilde{X}_n + \tilde{h}_n + V_n \\ \tilde{X}_{n+1} &= F_n \tilde{X}_n + \tilde{f}_n + \tilde{W}_n \end{cases}$$

$$\text{Avec } H_n = \begin{bmatrix} \frac{\bar{x}_n^1}{\sqrt{(\bar{x}_n^1 - a)^2 + (\bar{x}_n^2 - b)^2}} & \frac{\bar{x}_n^2}{\sqrt{(\bar{x}_n^1 - a)^2 + (\bar{x}_n^2 - b)^2}} \\ \frac{-(\bar{x}_n^2 - b)}{(\bar{x}_n^1 - a)^2 + (\bar{x}_n^2 - b)^2} & \frac{\bar{x}_n^1 - a}{(\bar{x}_n^1 - a)^2 + (\bar{x}_n^2 - b)^2} \end{bmatrix}$$

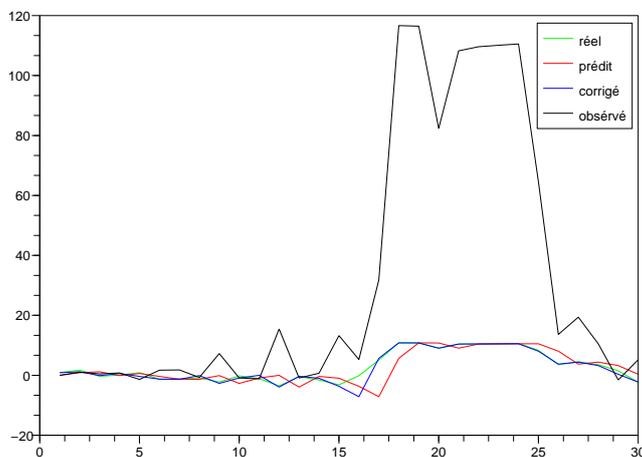
$$\tilde{h}_n = \begin{bmatrix} \sqrt{(\bar{x}_n^1 - a)^2 + (\bar{x}_n^2 - b)^2} \\ \arctan \left[\frac{\bar{x}_n^2 - b}{\bar{x}_n^1 - a} \right] \end{bmatrix}$$

$$F_n = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ et } \tilde{f}_n = \begin{bmatrix} \bar{x}_n^1 + V_0 \cos(\alpha) \cdot \tau - \bar{x}_{n+1}^1 \\ \bar{x}_n^2 - \frac{g_0 \cdot \tau}{2} \cdot [(2n+1)\tau - T] + V_0 \sin(\alpha) \cdot \tau - \tau - \bar{x}_{n+1}^2 \end{bmatrix}$$

$$Q_n^W = \begin{bmatrix} s_W^2 & 0 \\ 0 & s_W^2 \end{bmatrix} \text{ et } Q_n^V = \begin{bmatrix} s_V^2 & 0 \\ 0 & s_V^2 \end{bmatrix}$$

Valeurs numériques

vitesse initiale	: $V_0 = 10m \cdot s^{-1}$.
Accélération de la pesanteur	: $g_0 = 10m \cdot s^{-2}$.
Nombre de mesures	: $N = 21$, période : $\tau = 10s$.
Angle de lancement	: $\alpha = \pi/8$.
Position du capteur	: $A(500; 5000)$: $a = 500$ et $b = 5000$.
Erreur de mesure d'état	: $\mathcal{N}(0, \sigma_W \cdot I_2)$ avec $\sigma_W = 1.2$.
L'espace de temps est	: $t \in [-100; 100]$, $T = 200$
Erreur de mesure d'observation $\mathcal{N}(0, \sigma_V \cdot I_2)$: $\sigma_V = 0.8$.



Dans le filtrage de Kalman, l'approximation du signal théorique est immédiate. Le signal prédit s'écarte légèrement de la trajectoire nominale. L'observation est très éloignée de la trajectoire théorique mais avec cette observation, le signal corrigé est très proche de la trajectoire nominale.

FIGURE 5.10 – Filtrage de Kalman

5.4.2 Oubli de la condition initiale

Dans le filtrage de Kalman, l'oubli de la condition initiale est justifié en considérant deux conditions initiales différentes avec les paramètres, les valeurs estimées sont très proches des valeurs observées. À partir de la troisième itération, les valeurs estimées s'approchent rapidement des valeurs observées. Dans cette section, on considère deux cas :

Premier cas

y_1	y_2	x_1^1	x_2^1	x_1^2	x_2^2
- 7.5459809	0.4687459	0.	0.	100.	1000.
67.013525	4526.5359	75.725179	4527.411	106.2688	4529.4597
284.31387	8071.669	251.02856	8070.85	251.60959	8066.9498
297.75522	10604.622	292.2238	10595.565	292.12102	10595.434
343.0538	12148.095	351.63765	12146.714	351.6321	12146.726
144.37797	12703.307	248.27557	12714.522	248.27581	12714.523
501.43865	12236.603	515.70583	12265.868	515.70586	12265.868
629.5054	10765.128	635.64991	10771.164	635.64991	10771.164
651.40825	8317.6599	679.90396	8321.4983	679.90396	8321.4983
797.04113	4849.3746	807.69374	4858.141	807.69374	4858.141
928.08766	610.32518	920.11305	610.15184	920.11305	610.15184

TABLE 5.4 – Valeurs observées et estimées avec $X_0 = (0; 0)$ et $X_0 = (100; 1000)$

En prenant le même écart type $\sigma_0 = 1$ pour les valeurs initiales différentes des moyennes $X_0 = (0; 0)$ et $X_0 = (100; 1000)$, on obtient pratiquement les mêmes estimations à partir de quatrième itération.

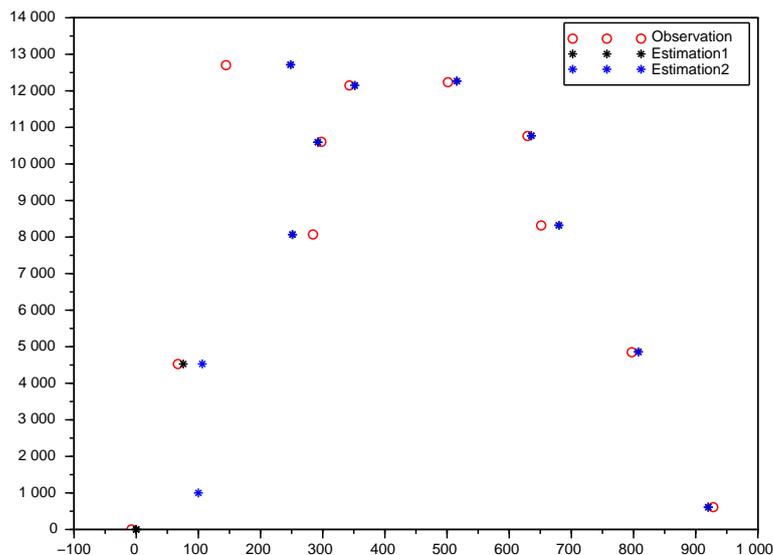


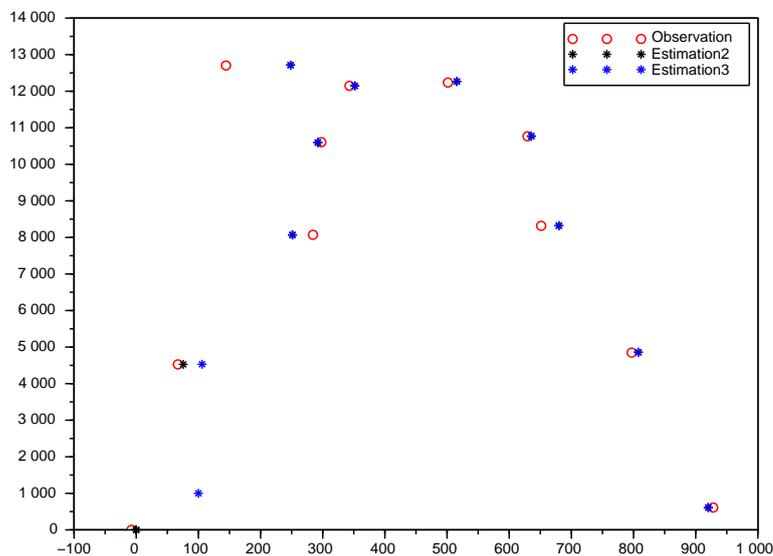
FIGURE 5.11 – Trajectoires avec $X_0 = (0; 0)$ et $X_0 = (100; 1000)$

Deuxième cas

y_1	y_2	x_1^2	x_2^2	x_1^3	x_2^3
- 7.5459809	0.4687459	100.	1000.	100.	1000.
67.013525	4526.5359	106.2688	4529.4597	106.89795	4529.9269
284.31387	8071.669	251.60959	8066.9498	251.61686	8067.0114
297.75522	10604.622	292.12102	10595.434	292.12293	10595.434
343.0538	12148.095	351.6321	12146.726	351.63213	12146.725
144.37797	12703.307	248.27581	12714.523	248.2758	12714.523
501.43865	12236.603	515.70583	12265.868	515.70586	12265.868
629.5054	10765.128	635.64991	10771.164	635.64991	10771.164
651.40825	8317.6599	679.90396	8321.4983	679.90396	8321.4983
797.04113	4849.3746	807.69374	4858.141	807.69374	4858.141
928.08766	610.32518	920.11305	610.15184	920.11305	610.15184

TABLE 5.5 – Valeurs observées et estimées avec $\sigma_0 = 1$ et $\sigma_0 = 2$

En prenant la même valeur de la moyenne à l'état initial $X_0 = (100; 1000)$ et avec deux écart-type différents $\sigma_0 = 1$ et $\sigma_0 = 2$, on obtient encore pratiquement les mêmes estimations après quatre itérations.

FIGURE 5.12 – Trajectoires avec $\sigma_0 = 1$ et $\sigma_0 = 2$

Finalement, les différentes valeurs des moyennes et des écarts-type n'ont pas d'impact dans l'estimation de la trajectoire réelle. Les valeurs observées permettent

5.5 Filtrage de Uhlmann et Julier

5.5.1 Modélisation

Dans la poursuite de trajectoire, considérons le système d'équations horaires

$$\begin{cases} x(t) = V_0 \cdot \cos(\alpha) \cdot t \\ y(t) = -\frac{g_0}{2} t \cdot (t - T) + V_0 \cdot \sin(\alpha) \cdot t \end{cases}$$

avec $t \in [0; T]$ et τ la période des mesures.

On a l'équation de la trajectoire avec $t = \frac{x}{V_0 \cdot \cos(\alpha)}$,

$$\begin{aligned} y &= -\frac{g_0}{2} \frac{x}{V_0 \cdot \cos(\alpha)} \cdot \left(\frac{x}{V_0 \cdot \cos(\alpha)} - T \right) + V_0 \cdot \sin(\alpha) \cdot \frac{x}{V_0 \cdot \cos(\alpha)} \\ y &= -\frac{g_0}{2(V_0 \cdot \cos(\alpha))^2} x^2 + \left(\frac{g_0 \cdot T}{2V_0 \cdot \cos(\alpha)} + \frac{V_0 \cdot \sin(\alpha)}{V_0 \cdot \cos(\alpha)} \right) \cdot x \\ y &= -\frac{g_0}{2(V_0 \cdot \cos(\alpha))^2} x^2 + \left(\frac{g_0 \cdot T + 2V_0 \cdot \sin(\alpha)}{2V_0 \cdot \cos(\alpha)} \right) \cdot x \end{aligned}$$

Notons à l'instant t_n , $x_n = x(t_n)$ et $y_n = y(t_n)$

$$x_{n+1} = x(t_n + \tau) = V_0 \cdot \cos(\alpha) \cdot (t_n + \tau) = x_n + V_0 \cdot \cos(\alpha) \cdot \tau. \quad (5.1)$$

$$y_{n+1} = -\frac{g_0}{2(V_0 \cdot \cos(\alpha))^2} x_{n+1}^2 + \left(\frac{g_0 \cdot T + 2V_0 \cdot \sin(\alpha)}{2V_0 \cdot \cos(\alpha)} \right) \cdot x_{n+1}$$

$$y_{n+1} = -\frac{g_0}{2(V_0 \cdot \cos(\alpha))^2} (x_n + V_0 \cdot \cos(\alpha) \cdot \tau)^2 + \left(\frac{g_0 \cdot T + 2V_0 \cdot \sin(\alpha)}{2V_0 \cdot \cos(\alpha)} \right) \cdot (x_n + V_0 \cdot \cos(\alpha) \cdot \tau)$$

$$\begin{aligned} y_{n+1} &= -\frac{g_0}{2(V_0 \cdot \cos(\alpha))^2} x_n^2 + \left(\frac{g_0 \cdot T + 2V_0 \cdot \sin(\alpha)}{2V_0 \cdot \cos(\alpha)} \right) \cdot x_n - \\ &\quad - \frac{2g_0 \cdot V_0 \cdot \cos(\alpha) \cdot \tau}{2(V_0 \cdot \cos(\alpha))^2} \cdot x_n + \left[-\frac{g_0 \cdot (V_0 \cdot \cos(\alpha) \cdot \tau)^2}{2(V_0 \cdot \cos(\alpha))^2} + \frac{(g_0 \cdot T + 2V_0 \cdot \sin(\alpha)) V_0 \cdot \cos(\alpha) \cdot \tau}{2V_0 \cdot \cos(\alpha)} \right] \\ y_{n+1} &= y_n - \frac{g_0 \cdot \tau}{V_0 \cdot \cos(\alpha)} \cdot x_n + \frac{g_0 \cdot \tau}{2} (T - \tau) + V_0 \cdot \sin(\alpha) \cdot \tau \end{aligned} \quad (5.2)$$

D'où l'équation d'état : $X_{n+1} = F_n \cdot X_n + f_n + W_n$

Avec $X_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix}$, $X_{n+1} = \begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix}$, $W_n \sim \mathcal{N}(0, \sigma_W \cdot I_2)$ où $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ et

$$F_n = \begin{pmatrix} 1 & 0 \\ -\frac{g_0 \cdot \tau}{V_0 \cdot \cos(\alpha)} & 1 \end{pmatrix}, f_n = \begin{pmatrix} V_0 \cdot \cos(\alpha) \cdot \tau \\ \frac{g_0 \tau}{2} (T - \tau) + V_0 \cdot \sin(\alpha) \cdot \tau \end{pmatrix}.$$

5.5.2 Algorithme

Algo filtre Uhlmann et Julier

Données : p_{X_0} densité de la loi initiale , h_n et f_n fonctions du modèle, p_W densité du bruit d'état, p_V densité du bruit d'observation.

Résultats : estimation des paramètres des filtres prédit et corrigé.

DEBUT

$$\omega_0 = \frac{\kappa}{d + \kappa}, \quad \text{pour } 1 \leq i \leq d, \omega_{-i} = \omega_i = \frac{1}{2(d + \kappa)}$$

Décomposition de Cholesky : $P_0 = S_0 \cdot S_0^*$

Initialisation des σ -points

$$x_0 = \hat{X}_0, \quad x_i = \hat{X}_0 + S_0 \cdot e_i \cdot \sqrt{d + \kappa}, \quad x_{-i} = \hat{X}_0 - S_0 \cdot e_i \cdot \sqrt{d + \kappa}$$

Pour n=1 à N

Décomposition de Cholesky : $P_{n-1} = S_{n-1} \cdot S_{n-1}^*$

Détermination des σ -points

$$x_0 = \hat{X}_{n-1}, \quad x_i = \hat{X}_{n-1} + S_{n-1} \cdot e_i \cdot \sqrt{d + \kappa}, \quad x_{-i} = \hat{X}_{n-1} - S_{n-1} \cdot e_i \cdot \sqrt{d + \kappa}$$

Evolution des σ -points : $u_i = \hat{X}_{n-1} - S_{n-1} \cdot x_i$

Moyenne et covariance du filtre prédit

$$\hat{X}_n^- = \sum_{i=-d}^d \omega_i \cdot f_n(u_i)$$

$$P_n^- = \sum_{i=-d}^d \omega_i (f_n(u_i) - \hat{X}_n^-) \cdot (f_n(u_i) - \hat{X}_n^-)^* + Q_n^W$$

Décomposition de Cholesky : $P_n'^- = S_n^- \cdot S_n'^{-*}$

Evolution des σ -points $v_i = \hat{X}_n^- - S_n^- \cdot x_i$

$$\hat{Y}_n^- = \sum_{i=-d}^d \omega_i \cdot h_n(v_i)$$

$$Q_n = \sum_{i=-d}^d \omega_i (h_n(v_i) - \hat{Y}_n^-) \cdot (h_n(v_i) - \hat{Y}_n^-)^* + Q_n^V$$

$$C_n = \sum_{i=-d}^d \omega_i (f_n(u_i) - \hat{X}_n^-) \cdot (h_n(v_i) - \hat{Y}_n^-)^*$$

Moyenne et covariance du filtre corrigé

$$\hat{X}_n = \hat{X}_n^- + C_n \cdot Q_n^{-1} \cdot (Y_n - \hat{Y}_n^-)$$

$$P_n = P_n^- - C_n \cdot Q_n \cdot C_n^*$$

Fin pour N.

FIN

TABLE 5.6 – Algorithme filtrage de Uhlmann et Julier

Approximation linéaire

Avec la loi gaussienne de moyenne m et d'écart type σ , considérons l'approximation linéaire par interpolation sur les sigma-points. $x_{-1} = m - \kappa \cdot \sigma$, $x_0 = m$ et $x_1 = m + \kappa \cdot \sigma$ avec $\kappa = 1.96$.

Si g est la densité gaussienne, les ordonnées sont définies par $g(x_{-1})$, $g(x_0)$ et $g(x_1)$.

Prenons $m = 2$, $\sigma = 1$ et on obtient les courbes brisées comme approximation linéaire de la densité gaussienne.

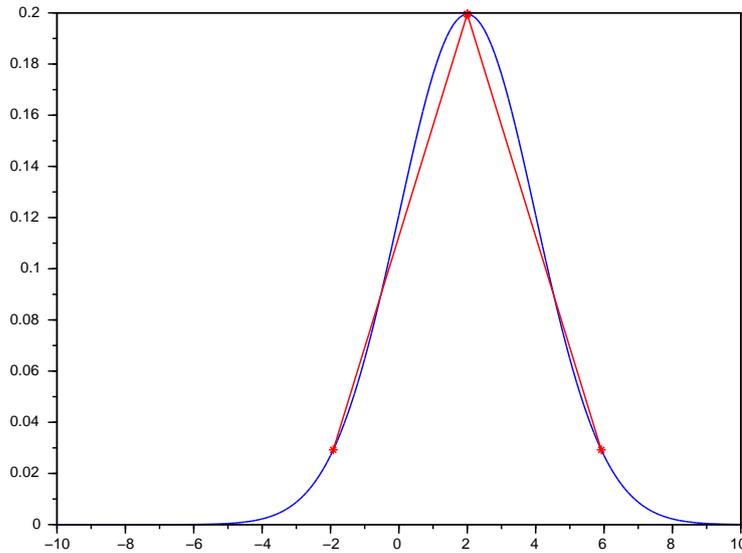


FIGURE 5.13 – Approximation linéaire de la densité gaussienne

5.5.3 Oubli de la condition initiale

Dans le filtrage de Kalman inodore, il y a effectivement oublié de la condition initiale. 4 cas sont à considérer :

- $X_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ et $Q_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
- $X_0 = \begin{pmatrix} 100 \\ 1000 \end{pmatrix}$ et $Q_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
- $X_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ et $Q_0 = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$
- $X_0 = \begin{pmatrix} 100 \\ 1000 \end{pmatrix}$ et $Q_0 = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$

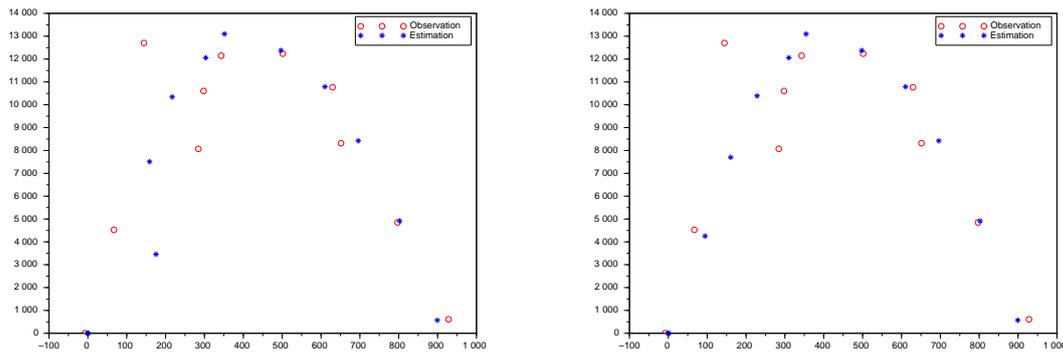
y_1 et y_2 sont les coordonnées des observations obtenues par les capteurs.

Le tableau suivant donne les trajectoires pour deux valeurs $X_0 = \begin{pmatrix} 100 \\ 1000 \end{pmatrix}$ et $X_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ avec $Q_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

y_1	y_2	x_1^1	x_2^1	x_1^2	x_2^2
- 7.5459809	0.4687459	100.	1000.	0.	0.
67.013525	4526.5359	175.37003	4411.8067	175.41208	3457.2727
284.31387	8071.669	199.69358	7694.3361	158.74927	7510.7
297.75522	10604.622	244.43169	10367.74	217.04032	10343.58
343.0538	12148.095	315.92076	12045.088	303.37072	12054.901
144.37797	12703.307	356.3824	13089.9	351.58818	13098.584
501.43865	12236.603	498.34045	12374.148	496.71915	12378.595
629.5054	10765.128	610.48533	10787.423	609.99987	10789.226
651.40825	8317.6599	695.95373	8422.1481	695.82799	8422.7835
797.04113	4849.3746	802.27755	4914.3471	802.25195	4914.5464
928.08766	610.32518	899.28444	570.04486	899.2821	570.09975

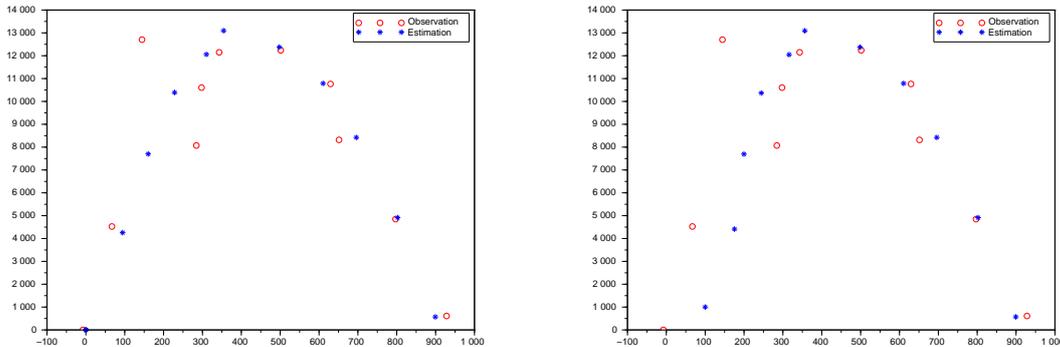
TABLE 5.7 – Valeurs observées et estimés pour X_0 différentes

Les moyennes estimées à l'état initial n'ont pas d'impact sur les valeurs estimées à la quatrième estimation.

FIGURE 5.14 – Trajectoires estimées pour X_0 différentes

Le tableau suivant donne les trajectoires pour deux valeurs $Q_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ et $Q_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ avec $X_0 = \begin{pmatrix} 100 \\ 1000 \end{pmatrix}$

y_1	y_2	x_1^3	x_2^3	x_1^4	x_2^4
- 7.5459809	0.4687459	100.	1000.	100.	1000.
67.013525	4526.5359	175.37003	4411.8067	260.63672	3591.5728
284.31387	8071.669	199.69358	7694.3361	199.61806	7482.7591
297.75522	10604.622	244.43169	10367.74	232.56823	10313.715
343.0538	12148.095	315.92076	12045.088	308.52086	12040.156
144.37797	12703.307	356.3824	13089.9	353.0985	13092.732
501.43865	12236.603	498.34045	12374.148	497.09921	12376.567
629.5054	10765.128	610.48533	10787.423	610.07263	10788.602
651.40825	8317.6599	695.95373	8422.1481	695.8322	8422.6156
797.04113	4849.3746	802.27755	4914.3471	802.24676	4914.5096
928.08766	610.32518	899.28444	570.04486	899.27846	570.09505

TABLE 5.8 – Valeurs observées et estimés pour Q_0 différentesFIGURE 5.15 – Trajectoires estimées pour Q_0 différentes

5.6 Conclusion

Dans la poursuite de trajectoire, les filtrages dans des cas de modèle d'état markovien sont appliqués. L'introduction des coordonnées polaires dans les mesures des observations entraîne la non linéarité du modèle d'observation. Les mesures d'observation sont en général, perturbées par des bruits blancs uniformes dûs aux imprécisions des mesures. L'oubli de la condition initiale est aisément vérifié dans ce cas.

De même, les filtrages de Kalman étendus et inodores sont appliqués. Pratiquement, la mise en œuvre du filtrage de Kalman étendu est plus simple que celle du filtrage de Kalman inodore mais la précision est apparemment meilleure pour le dernier. L'oubli de la condition inintiale est vérifié pour ces filtrages de Kalman.

Chapitre 6

Commande optimale et estimation paramétrique

Sommaire

6.1 Introduction	135
6.2 Filtrage particulaire	136
6.2.1 Modelisation de système commandé	136
6.2.2 Algorithme	137
6.2.3 Oubli de la condition initiale	139
6.3 Filtrage par noyau de convolution	143
6.3.1 Modelisation de système paramétrique	143
6.3.2 Estimation paramétrique	143
6.3.3 Estimation non paramétrique	144
6.3.4 Estimation par histogramme mobile	144
6.3.5 Estimation par noyau de convolution	146
6.3.6 Estimation des densités conditionnelles	147
6.3.7 Algorithme	150
6.3.8 Oubli de la condition initiale	153
6.4 Conclusion	158

6.1 Introduction

Dans ce chapitre d'applications, le filtrage particulaire est adopté pour un système commandé et le filtrage par noyau de convolution pour un système paramétré. Pour ces deux filtrage, l'échantillonnage des particules tient le rôle clef. Les methodes de filtrage sont appliquées dans un plan et les particules sont tirées dans l'espace de didmension 2.

6.2 Filtrage particulière

6.2.1 Modélisation de système commandé

Étant donnée la trajectoire nominale d'un avion volant à une vitesse théoriquement uniforme, le filtrage consiste à estimer la trajectoire réelle à chaque période de temps selon l'observation capté par un radar.

Le modèle d'état est tel que,

$$X_{n+1} = X_n + A_n + W_n$$

où $A_n = \begin{pmatrix} \omega_n \cdot \cos \alpha_n \cdot \tau \\ \omega_n \cdot \sin \alpha_n \cdot \tau \end{pmatrix}$ avec τ la période de mesure, ω_n est la vitesse commandée en fonction de la dernière estimation.

Plus précisément, on a $\omega_n = \frac{l_n}{\tau}$ avec $l_n = \sqrt{(x_{n+1} - \hat{X}_{1,n})^2 + (x_n - \hat{X}_{2,n})^2}$

$X_{n+1} = \begin{pmatrix} x_{1,n+1} \\ x_{2,n+1} \end{pmatrix}$, $X_n = \begin{pmatrix} x_{1,n} \\ x_{2,n} \end{pmatrix}$ et $W_n = \begin{pmatrix} W_{1,n} \\ W_{2,n} \end{pmatrix}$ est un bruit blanc perturbateur des mesures d'état.

$$\begin{cases} x_{1,n+1} &= x_{1,n} + \omega_n \cdot \cos \alpha_n \cdot \tau + W_{1,n} \\ x_{2,n+1} &= x_{2,n} + \omega_n \cdot \sin \alpha_n \cdot \tau + W_{2,n} \end{cases}$$

$$\alpha_n = \arctan \left(\frac{x_n - \hat{X}_{2,n}}{x_{n+1} - \hat{X}_{1,n}} \right)$$

Le modèle d'observation est tout simplement

$$Y_n = X_n + V_n$$

où $Y_n = \begin{pmatrix} y_{1,n} \\ y_{2,n} \end{pmatrix}$ et $V_n = \begin{pmatrix} V_{1,n} \\ V_{2,n} \end{pmatrix}$ est un bruit blanc perturbateur des mesures d'observation.

$$\begin{cases} y_{1,n} &= x_{1,n} + V_{1,n} \\ y_{2,n} &= x_{2,n} + V_{2,n} \end{cases}$$

6.2.2 Algorithme

Algo filtre particulière

Données : p_{X_0} densité de la loi initiale , h et f fonctions du modèle,
 $\sigma_W > 0$ écart type du bruit de mesure p_V densité du bruit d'observation.

Résultats : particules $(\hat{\xi}_n^i, 1 \leq i \leq N)$

DEBUT

Initialisation des particules : $(\hat{\xi}_1^0, \dots, \hat{\xi}_N^0)$

$$\xi_n^0 \sim p_{X_0} \text{ avec les poids } \omega_n^0 = \frac{1}{N}$$

Pour $i = 1 : N_1$ faire

 Pour $n = 1 : N$ faire

$$\text{mutation : } \xi_n^i = f(\hat{\xi}_n^{i-1}) + W_n^i$$

 pondération :

$$\omega_n^i = \frac{p_V(Y_i - h(\xi_n^i))}{\sum_{j=1}^N p_V(Y_i - h(\xi_n^j))}$$

 fin pour n

$$N^{eff} \leftarrow \left(\sum_{i=1}^N (\omega_n^i)^2 \right)^{-1}$$

 Test si $N^{eff}/N \leq 0.75$ alors,

 Selection : redistribution des particules :

$$(\hat{\xi}_1^i, \dots, \hat{\xi}_N^i) \text{ avec } \omega_n^i \leftarrow 1/N$$

 Sinon continuer

$$(\hat{\xi}_1^i, \dots, \hat{\xi}_N^i) = (\xi_1^i, \dots, \xi_N^i)$$

fin pour i

FIN

TABLE 6.1 – Algorithme filtre particulière

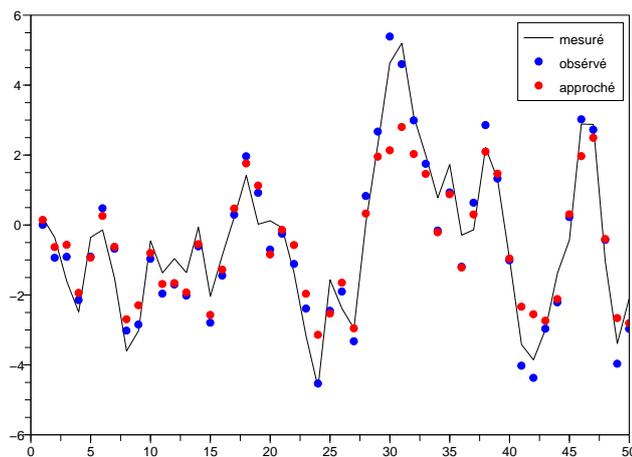


FIGURE 6.1 – Filtre particulaire

Dans le filtrage particulaire, la trajectoire théorique mesurée est approchée par les estimations obtenues à partir de la distribution de particules. L'observation s'écarte légèrement de la trajectoire théorique non observée. Les estimations sont proches de la trajectoire nominale sans perturbations.

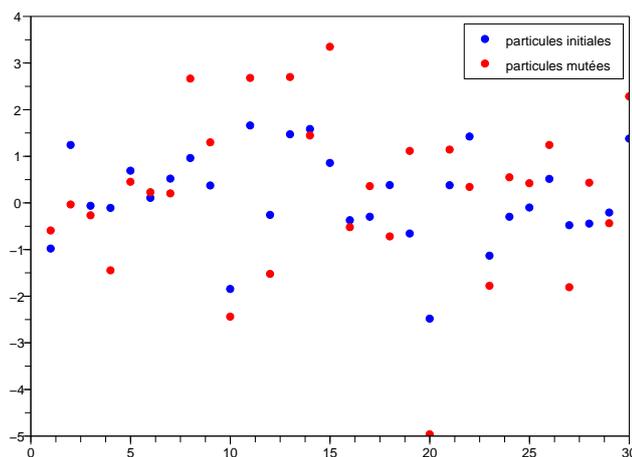


FIGURE 6.2 – Mutation des particules

Les particules initiales sont mutées à partir de l'équation d'état. Les poids des particules sont calculés à partir de l'équation d'observation. Le poids d'une particule caractérise sa contribution dans l'approximation. Le nombre de particules efficaces est obtenu de la formule de N_{eff} .

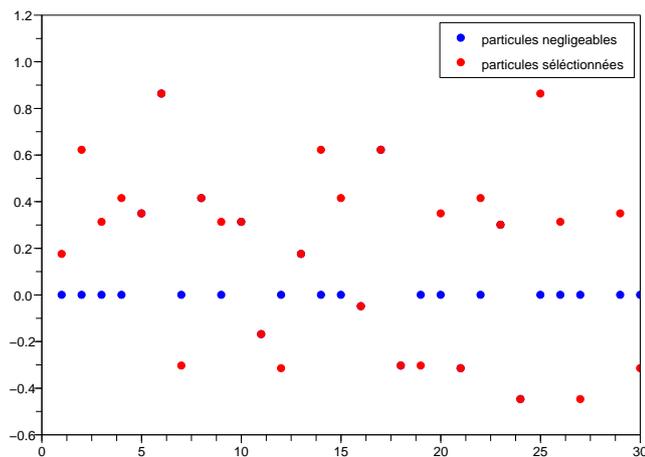


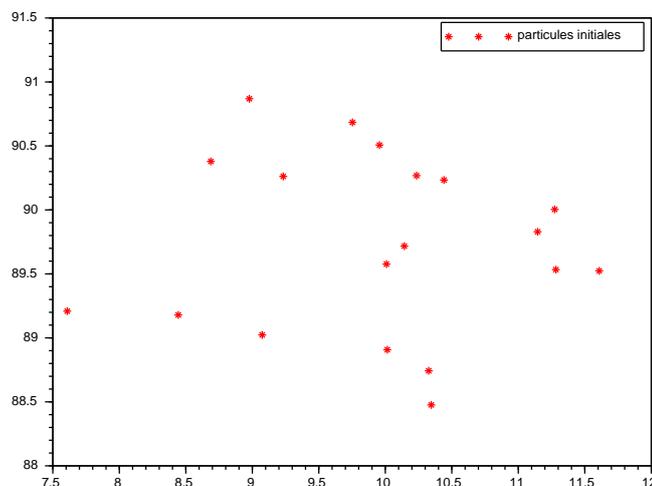
FIGURE 6.3 – Selection des particules

Les particules sont sélectionnées selon leurs poids. Les particules de poids négligeables sont remplacées par des particules obtenues par redistribution multinomiale des particules de poids non négligeables. Après la redistribution, les particules ont le même poids $1/N$.

6.2.3 Oubli de la condition initiale

Valeurs numériques

Condition initiale	: $X_0 \sim \mathcal{N}(\hat{X}_0, \Gamma_0)$
Periode	: $\tau = 10s$
La vitesse initiale	: $V_0 = 10ms^{-1}$
Nombre de mesures	:20
Ecart type	
Mesure d'état	: $s_W = 0.8$
Mesure d'observation	: $s_V = 1.4$.
Accélération de la pesanteur	: $g_0 = 10m.s^{-2}$.



Les particules initiales sont obtenues par échantillonnage de la loi initiale normale $\mathcal{N}(\hat{X}_0, s_0^2)$ avec la moyenne $\hat{X}_0 = \begin{pmatrix} 10 \\ 90 \end{pmatrix}$ et de variance $s_0^2 = 0.8$. Le nombre de particules est de 40.

FIGURE 6.4 – Particules de la loi initiale

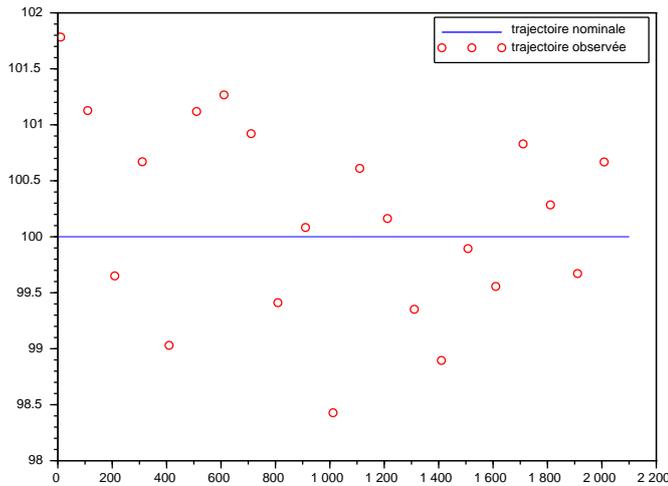
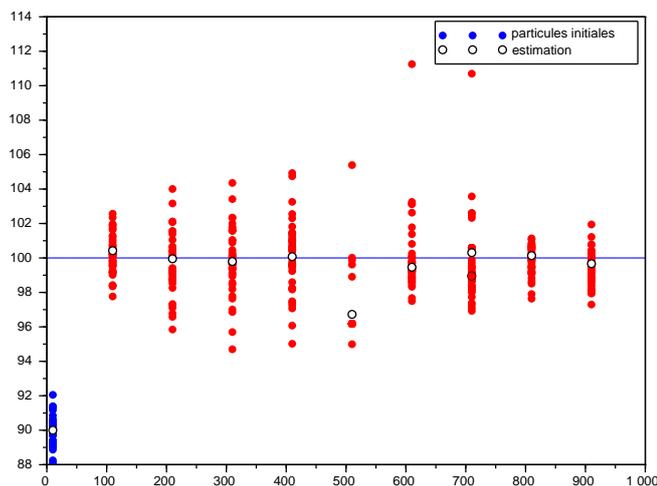


FIGURE 6.5 – Trajectoire nominale et observée

L'engin volant suit une trajectoire rectiligne. Les mesures sont prises systématiquement avec une période de 10s. Les mesures des observations sont perturbées par un bruit blanc de matrice de covariance l'unité.

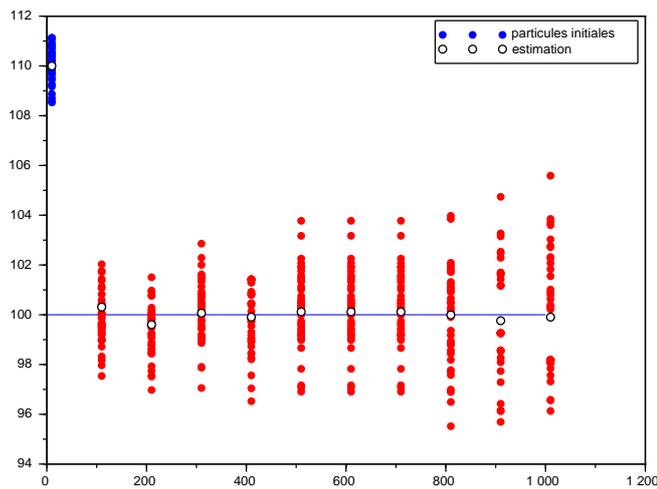


La condition initiale est $X_0 \sim \mathcal{N}(\hat{X}, s_0)$ avec $\hat{X}_0 = \begin{pmatrix} 10 \\ 90 \end{pmatrix}$ et $s_0 = 0.8$

Les particules sont dispersées sur des droites verticales à la trajectoire nominale. À la première itération, l'estimation est très proche de la trajectoire réelle.

FIGURE 6.6 – Évolution des particules avec $\hat{X}_0 = (10; 90)$ et $s_0 = 0.8$

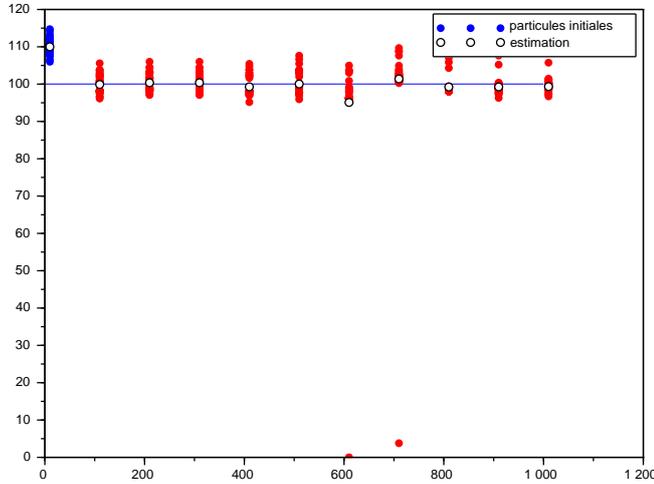
n	x	y	v	α
0	10	90	10.050876	1.4701377
1	110	100.42353	10.000137	1.57603165
2	210	99.953279	10.000011	1.5693291
3	310	99.794889	10.000047	1.57384745
4	410	100.07696	10.000016	1.5690268
5	510	96.72457	10.005695	1.5370558
6	610	99.463145	10.000203	1.57716475
7	710	100.30853	10.000083	1.566711
8	810	98.952991	10.000658	1.58226595
9	910	100.13502	10.000028	1.5684461
10	1010	99.66789		

TABLE 6.2 – Vitesse et angle pour $\hat{X}_0 = (10; 90)$ et $s_0 = 0.8$ 

La condition initiale est
 $X_0 \sim \mathcal{N}(\hat{X}, s_0)$ avec
 $\hat{X}_0 = \begin{pmatrix} 10 \\ 110 \end{pmatrix}$ et $s_0 = 0.8$

FIGURE 6.7 – Évolution des particules avec $\hat{X}_0 = (10; 110)$ et $s_0 = 0.8$

n	x	y	v	α
0	10	110	10.050876	1.67145495
1	110	100.30978	10.000084	1.57489405
2	210	99.603192	10.000123	1.5658283
3	310	100.0651	10.000014	1.57247145
4	410	99.907158	10.000019	1.688679
5	510	100.11962	10.000024	1.57299245
6	610	100.11962	10.000024	1.57299245
7	710	100.11962	10.000024	1.57299245
8	810	99.995124	10.000005	1.5697476
9	910	99.76132	10.000057	1.5674095
10	1010	99.904991		

TABLE 6.3 – Vitesse et angle pour $\hat{X}_0 = (10; 110)$ et $s_0 = 0.8$ 

La condition initiale est
 $X_0 \sim \mathcal{N}(\hat{X}, s_0)$ avec
 $\hat{X}_0 = \begin{pmatrix} 10 \\ 110 \end{pmatrix}$ et $s_0 = 2$

FIGURE 6.8 – Évolution des particules avec $\hat{X}_0 = (10; 110)$ et $s_0 = 2$

n	x	y	v	α
0	10	110	10.050876	1.67145495
1	110	99.904991	10.000019	1.5688462
2	210	100.37613	10.000113	1.58124165
3	310	100.37613	10.000113	1.57555755
4	410	99.273133	10.000342	1.5625278
5	510	100.04038	10.00001	1.57240015
6	610	95.096761	10.012508	1.5208056
7	710	101.39904	10.001123	1.58578565
8	810	99.225405	10.000382	1.5620506
9	910	99.252267	10.000359	1.5623192
10	1010	99.34405		

TABLE 6.4 – Vitesse et angle pour $\hat{X}_0 = (10; 110)$ et $s_0 = 2$

6.3 Filtrage par noyau de convolution

6.3.1 Modélisation de système paramétré

Une trajectoire théorique est suivi par un avion. Avec un radar, les contrôleurs aériens peuvent suivre ce déplacement en recevant un signal en temps à intervalle régulier. L'avion ne peut pas suivre exactement le plan de vol à cause des perturbations aériennes comme les trous d'air. De même, le tour de contrôle capte les signaux avec des imprecision du radar.

Notons X_n l'écart entre la trajectoire idéale et la position de l'avion au temps n et Y_n la mesure observée par le radar au temps n . On obtient le système suivant :

$$\begin{cases} X_0 &= W_0 \\ X_n &= aX_{n-1} + W_n \text{ où } W_n \text{ est un bruit blanc de variance } \sigma \text{ de la mesure de la} \\ Y_n &= X_n + V_n \end{cases}$$

position X et V_n est un bruit blanc de l'observation Y .

On souhaite estimer les coefficients a et σ .

6.3.2 Estimation paramétrique

Introduisons d'abord le lemme sur les lois conditionnelles gaussiennes suivant,

Lemme 29 Soit (X, Y) un vecteur gaussien dans \mathbb{R}^2 tel que $\mathcal{L} \sim \mathcal{N}(\mu, \gamma^2)$ et

$$\mathcal{L}(Y|X) \sim \mathcal{N}(\alpha X + \beta, \delta^2)$$

alors

$$\mathcal{L}(X|Y) \sim \mathcal{N}\left(\rho \left(\frac{\mu}{\gamma^2} + \frac{\alpha(Y - \beta)}{\delta^2}\right), \rho^2\right)$$

avec

$$\frac{1}{\rho^2} = \frac{1}{\gamma^2} + \frac{\alpha^2}{\delta^2}.$$

La vraisemblance du modèle est :

$$L(a, \sigma^2, X_1, \dots, X_n) = \frac{1}{\sqrt{2\pi\sigma^2}^n} \prod_{k=1}^n \exp\left[-\frac{(X_k - aX_{k-1})^2}{2\sigma^2}\right]$$

La Log-vraisemblance est :

$$\ln L(a, \sigma^2, X_1, \dots, X_n) = -\frac{n \ln(2\pi)}{2} - \frac{n}{2} \ln(\sigma^2) + \sum_{k=1}^n \left[-\frac{(X_k - aX_{k-1})^2}{2\sigma^2}\right]$$

Les estimateurs obtenus par la méthode du maximum de vraisemblance sont obtenus par

$$\frac{\partial \ln L(a, \sigma, X_1, \dots, X_n)}{\partial a} = 0 \text{ et } \frac{\partial \ln L(a, \sigma, X_1, \dots, X_n)}{\partial \sigma^2} = 0 \text{ et on a :}$$

$$\hat{a}_n = \frac{\sum_{k=1}^n X_{k-1} X_k}{\sum_{k=1}^n X_{k-1}^2} \quad (6.1)$$

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \hat{a}_n X_{k-1})^2. \quad (6.2)$$

$X_k = Y_n - V_n$ n'est pas observable.

6.3.3 Estimation non-paramétrique

À partir de n -observations x_1, \dots, x_n de la variable aléatoire X , on construit l'histogramme à pas régulier $h = \frac{\max x_i - \min x_i}{N}$, où N est le nombre de rectangles dans l'histogramme. Avec un nombre assez important d'observations et un pas tendant vers 0, la densité de X est approchée par l'histogramme.

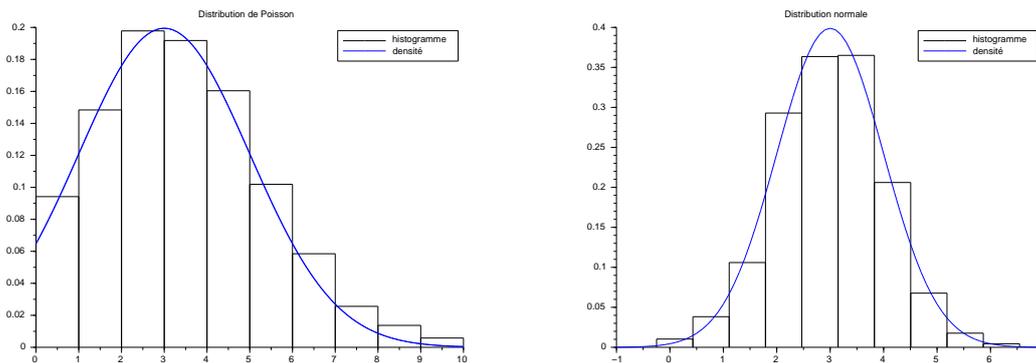


FIGURE 6.9 – Estimations de densités par histogramme

6.3.4 Estimation par histogramme mobile

En considérant un intervalle de longueur h appelé paramètre de lissage, centré en x , par définition, la relation entre la fonction densité f_X de X et sa fonction de répartition \mathbb{F}_X est telle que,

$$\begin{aligned} f_X(x) &= \lim_{h \rightarrow 0} \frac{\mathbb{F}_X\left(x + \frac{h}{2}\right) - \mathbb{F}_X\left(x - \frac{h}{2}\right)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\mathbb{P}\left(x - \frac{h}{2} \leq X \leq x + \frac{h}{2}\right)}{h} \end{aligned}$$

Avec n -observations x_1, \dots, x_n de X , on peut approximer la densité f_X par :

$$\begin{aligned}
 \tilde{f}_X^n(x) &= \frac{1}{h} \left[\frac{\text{card} \{x_i : x - h/2 \leq x_i \leq x + h/2\}}{n} \right] \\
 &= \frac{1}{n \cdot h} \sum_{i=1}^n \mathbb{I}_{[x-h/2, x+h/2]}(x_i) \\
 &= \frac{1}{n \cdot h} \sum_{i=1}^n \mathbb{I}_{[x_i-h/2, x_i+h/2]}(x) \\
 &= \frac{1}{n \cdot h} \sum_{i=1}^n \mathbb{I}_{[-h/2, h/2]}(x - x_i) \\
 &= \frac{1}{n \cdot h} \sum_{i=1}^n \mathbb{I}_{[-1/2, 1/2]} \left(\frac{x - x_i}{h} \right) \\
 &= \frac{1}{n \cdot h} \sum_{i=1}^n \delta_{|x| \leq 1/2} \left(\frac{x - x_i}{h} \right) \\
 &= \frac{1}{n} \sum_{i=1}^n \delta_{|x| \leq 1/2}^h(x - x_i) \quad \text{avec} \quad \delta_{|x| \leq 1/2}^h(u) = \frac{1}{h} \delta_{|x| \leq 1/2} \left(\frac{u}{h} \right)
 \end{aligned}$$

D'une part, l'approximation de la densité de X en un point x est la moyenne de la densité uniforme $u_{[-h/2, h/2]}$ sur l'intervalle $[-h/2, h/2]$ appliquée aux écarts par rapport à x , et d'autre part c'est le produit de convolution de cette densité uniforme par la mesure somme de mesure de Dirac sur les observations $\mu = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$.

$$\tilde{f}_X^n(x) = (u_{[-h/2, h/2]} * \mu)(x) = \int_{\mathbb{R}} u_{[-h/2, h/2]}(x - x_i) \mu(x_i) dx_i$$

Pour un paramètre de lissage $h \leq 1$, l'estimation de la densité est plus proche de la

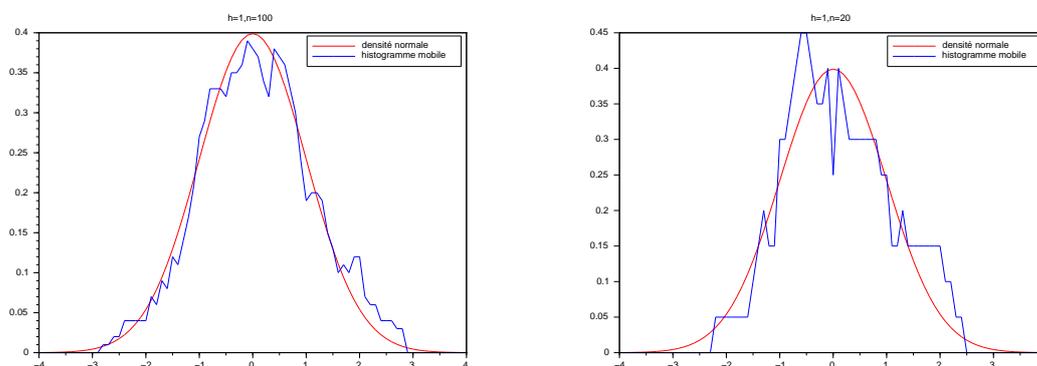


FIGURE 6.10 – Estimations par histogramme mobile pour $h=1$

densité exacte. Cette approximation est encore meilleure si le nombre d'observation est plus important.

Pour $h = 10$, la discontinuité de l'estimation est mise en évidence. L'approximation par noyau améliore l'estimation.

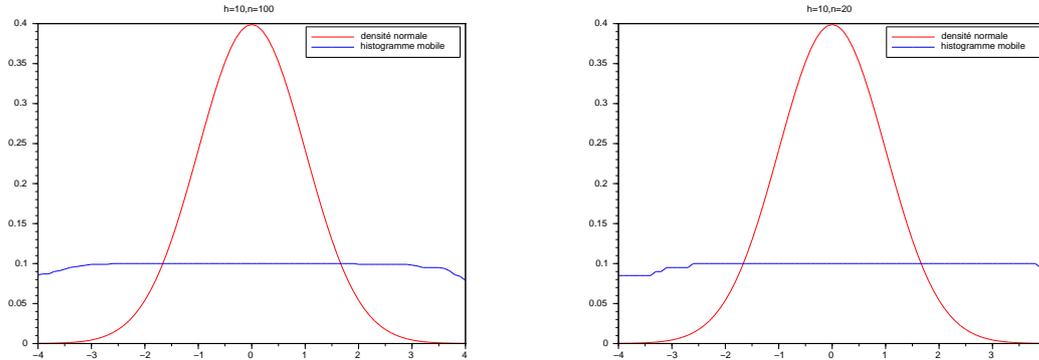


FIGURE 6.11 – Estimations par histogramme mobile pour $h=10$

6.3.5 Estimation par noyau de convolution

L'estimation par noyau de convolution [ABDOUS et al,1994] est une généralisation de l'estimation par histogramme mobile. L'approximation de la densité est telle que :

$$\tilde{f}_X^n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad x \in \mathbb{R}^d$$

K est un noyau de Parzen-Rosenblatt tel que : $\int_{\mathbb{R}^d} K(x)dx = 1$, $\lim_{|x| \rightarrow \infty} |x|^d K(x) = 0$ et le paramètre de lissage h est une fonction du nombre d'observations n .

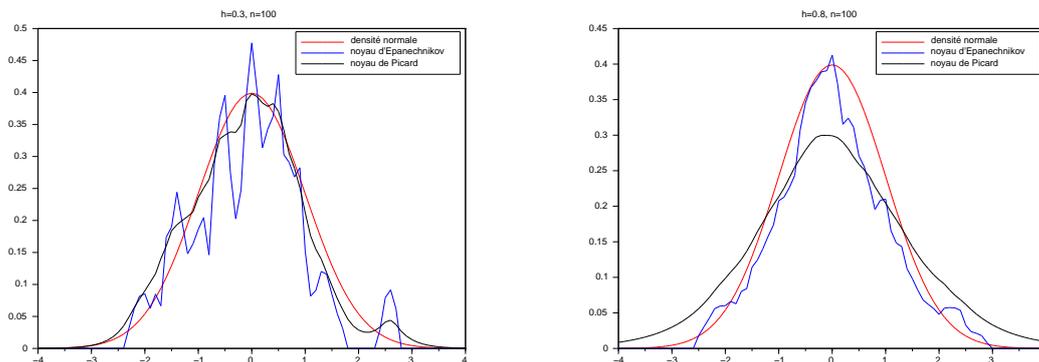
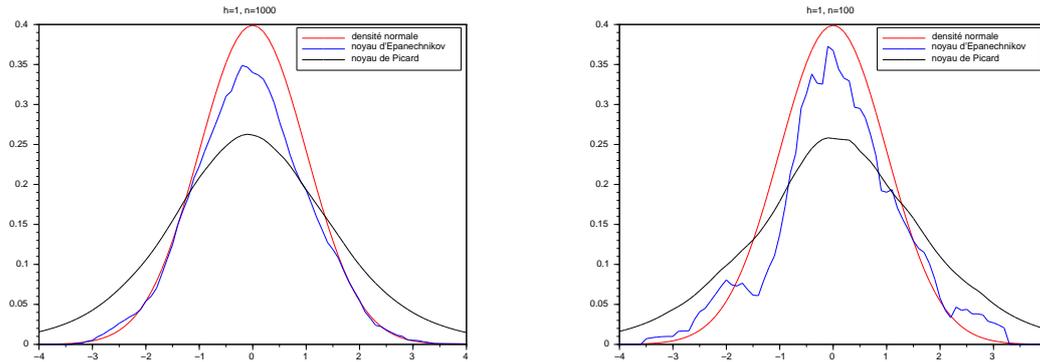


FIGURE 6.12 – Estimations par noyau pour $h=0.3$ et $h=0.8$

L'estimation de la densité par noyau est meilleure lorsque le paramètre de lissage est environ 0.3. Elle est apparemment meilleure avec un noyau de Picard qu'avec un noyau d'Epanechnikov. En tout cas, le nombre d'observation doit être assez important.

FIGURE 6.13 – Estimations par noyau pour $h=1$

Pour le paramètre de lissage égal à 1, l'estimation de la densité normale est meilleure avec un noyau d'Epanechnikov qu'avec un noyau de Picard. L'estimation est acceptable avec un nombre d'observé élevé à partir de 1000.

On a les mêmes constatations pour les estimations par noyau d'une densité exponentielle.

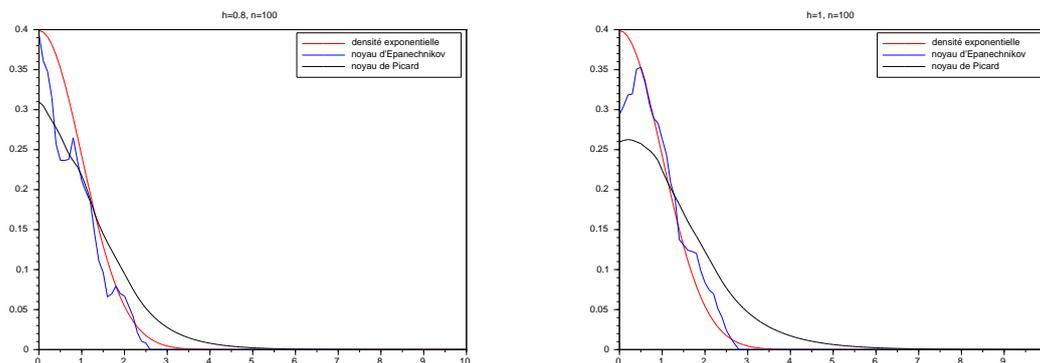


FIGURE 6.14 – Estimations par noyau de la densité exponentielle

6.3.6 Estimation des densités conditionnelles

La loi de Y est de densité p_Y et la loi conjointe de (X, Y) est de densité p_{XY} . Par la formule de Bayes, la loi conditionnelle de X sachant $Y = y$ est telle que : pour tout φ fonction test mesurable bornée,

$$\mathbb{E}[\varphi(X)|X = y] = \int_{\mathbb{R}^d} \varphi(x)p_{X|Y=y}(x)dx$$

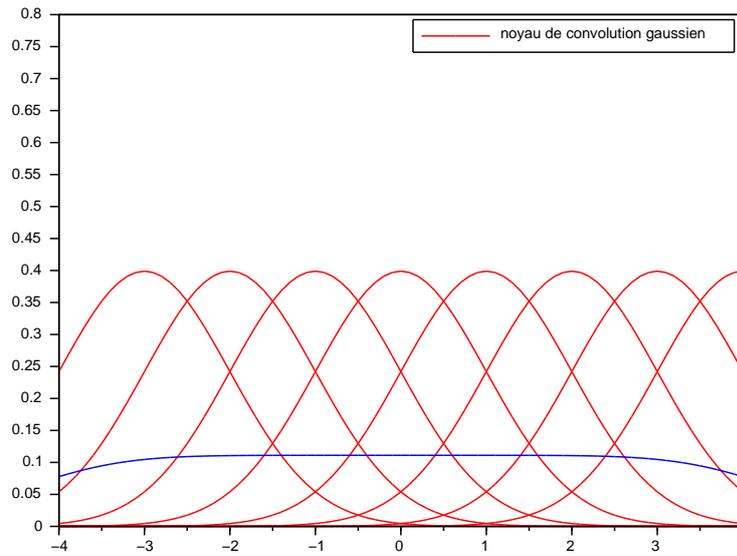


FIGURE 6.15 – Estimation par noyau de convolution gaussien

et avec des noyaux de convolution K_h^{XY} dans l'estimation de la densité conjointe de (X, Y) et K_h^Y dans l'estimation de la densité marginale de Y ,

$$\begin{aligned}
 p_{X|Y=y}(x) &= \frac{p_{XY}(x, y)}{p_Y(y)} = \frac{p_X(x) \cdot p_{Y|X=x}(y)}{p_Y(y)} \\
 &\approx \frac{\frac{1}{n \cdot h^2} \sum_{i=1}^n K_h^{XY} \left(\frac{x - X_i}{h}, \frac{y - Y_i}{h} \right)}{\frac{1}{n \cdot h} \sum_{i=1}^n K_h^Y \left(\frac{y - Y_i}{h} \right)} \\
 &\approx \frac{\frac{1}{n \cdot h^2} \sum_{i=1}^n K_h^X \left(\frac{x - X_i}{h} \right) \cdot K_h^{Y|X=x} \left(\frac{y - Y_i}{h} \right)}{\frac{1}{n \cdot h} \sum_{i=1}^n K_h^Y \left(\frac{y - Y_i}{h} \right)}
 \end{aligned}$$

$K_h^{Y|X=x}$ est le noyau de convolution dans l'estimation de la densité conditionnelle de Y sachant $X = x$.

Avec le noyau d'Epanechnikov, $K(y) = 1 - \frac{3}{2}y^2$ si $|y| \leq 1$ et le noyau de Picard

$K(y) = \frac{1}{2} \exp(-|y|)$, on construit les noyaux de Parzen-Rosenblatt en deux dimensions tels que,

$$K(x, y) = \left(1 - \frac{3}{2}x^2\right) \left(1 - \frac{3}{2}y^2\right)$$

défini sur le rectangle $] - 1, 1[\times] - 1, 1[$ et

$$K(x, y) = \frac{1}{4} \exp[-(|x| + |y|)]$$

défini sur \mathbb{R}^2 où $\lim_{\sqrt{x^2+y^2} \rightarrow \infty} (x^2 + y^2)K(x, y) = 0$.

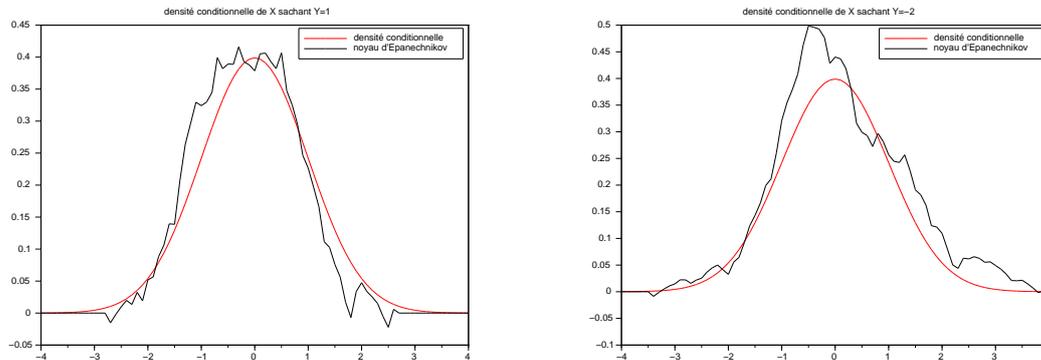


FIGURE 6.16 – Estimations par noyau d'Epanechnikov de la densité conditionnelle

La loi conjointe de (X, Y) est gaussienne et la loi conditionnelle de X sachant $Y = y$ est également gaussienne.

Dans les noyaux d'Epanechnikov, le paramètre de lissage est égal à 1 et dans les noyaux de Picard il est de 0.3. Le nombre d'observation est égal à 1000.

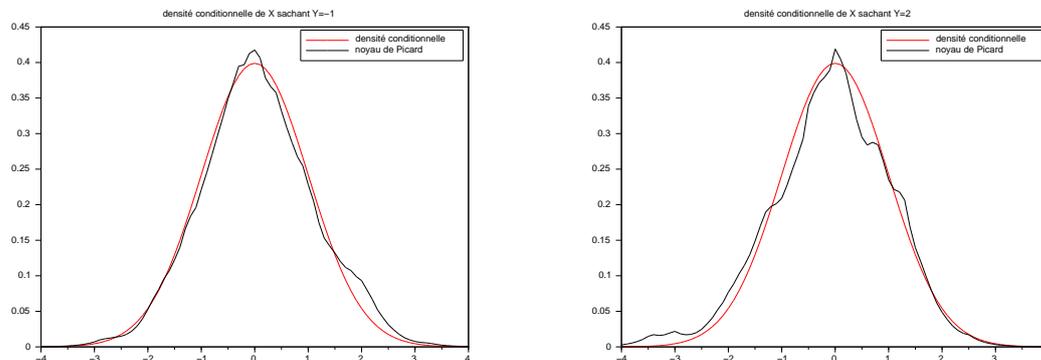


FIGURE 6.17 – Estimations par noyau de Picard de la densité conditionnelle

6.3.7 Algorithme

Algo filtre par noyau

Données : p_{X_0} densité de la loi initiale , h_t et f_t fonctions du modèle,

K_h noyau de convolution p_V densité du bruit d'observation.

Résultats : estimation de la densité du paramètre θ , densité du filtre optimal

DEBUT

Initialisation des N 1 particules : $\xi_0^i = (\tilde{x}_0^i, \tilde{\theta}_0^i) \sim p_{X_0}(x, \theta)$,

Pour $t = 1$ à N

Génération des particules : $(\bar{x}_{t-1}^i, \bar{\theta}_{t-1}^i) \sim \hat{p}_{t-1}^n(x, \theta | y_{1:t-1})$

Evolution des particules : $\tilde{\omega}_t^i \sim p_W(v)$, $\tilde{\eta}_t^i \sim p_V(v)$, $\tilde{\theta}_t^i = \bar{\theta}_{t-1}^i$.

$\tilde{x}_t^i = f_t(\bar{x}_{t-1}^i, \bar{\theta}_{t-1}^i, \tilde{\omega}_{t-1}^i)$

$\tilde{y}_t^i = h_t(\tilde{x}_t^i, \tilde{\theta}_t^i, \tilde{\eta}_t^i)$

Estimation des densités : $\hat{p}_t^n(x, \theta | y_{1:t}) = \frac{\sum_{i=1}^n K_h(\tilde{y}_t^i - y_t) K_h(\tilde{\theta}_t^i - \theta) K_h(\tilde{x}_t^i - x)}{\sum_{i=1}^n K_h(\tilde{y}_t^i - y_t)}$

$$\hat{p}_t^n(\theta | y_{1:t}) = \frac{\sum_{i=1}^n K_h(\tilde{y}_t^i - y_t) K_h(\tilde{\theta}_t^i - \theta)}{\sum_{i=1}^n K_h(\tilde{y}_t^i - y_t)}$$

$$\hat{p}_t^n(x | y_{1:t}) = \frac{\sum_{i=1}^n K_h(\tilde{y}_t^i - y_t) K_h(\tilde{x}_t^i - x)}{\sum_{i=1}^n K_h(\tilde{y}_t^i - y_t)}$$

Fin pour t .

FIN

TABLE 6.5 – Algorithme filtrage par noyau de convolution

Applications numériques

Particules initiales :

$$\begin{aligned}\theta &= (a, \sigma^2) \quad a_0 \sim \mathcal{U}[a_0 - h_1; \bar{a}_0 + h_1], \\ \sigma_0^2 &\sim \mathcal{U}[a_0 - h_1; \bar{\sigma}_0^2 + h_1], \\ X_0 &\sim \mathcal{N}(0, \bar{\sigma}_0)\end{aligned}$$

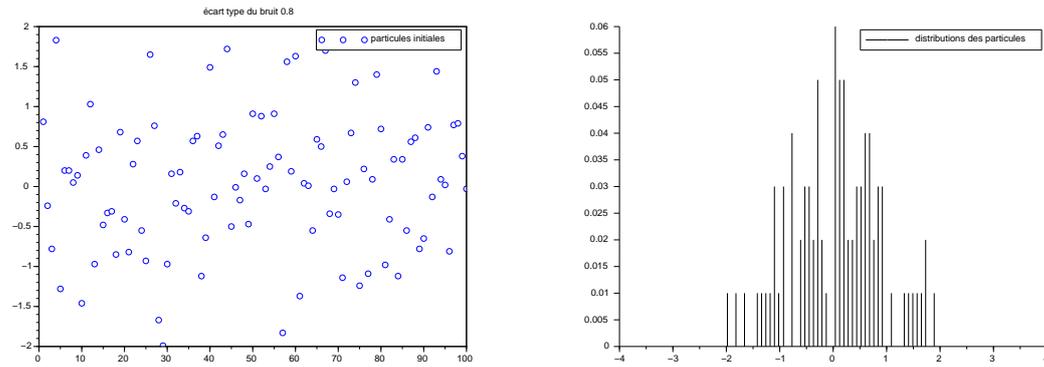


FIGURE 6.18 – Distribution des particules initiales

Evolution des particules

$$\begin{aligned}w_{t-1}^i &\sim \mathcal{N}(0, \sigma_W^i) \\ \bar{x}_t^1 &= a_t^i \bar{x}_{t-1}^i + w_{t-1}^i \\ v_{t-1}^i &\sim \mathcal{N}(0, \sigma_v^i) \\ \bar{y}_t^1 &= \sqrt{1 + (\bar{x}_t^i)^2} + v_t^i\end{aligned}$$

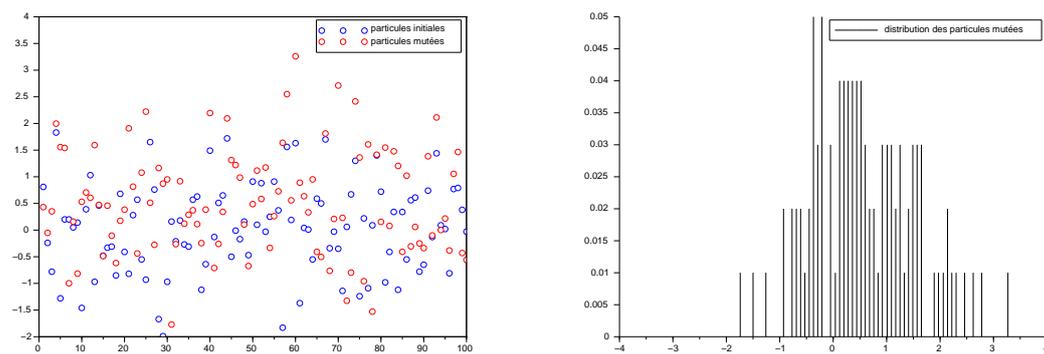


FIGURE 6.19 – Distribution des particules en mutation

Régularisation des filtres

K_h noyau d'Epanechnikov

$$\hat{p}_t^n(x|y_{1:t}) = \frac{\sum_{i=1}^n K_h(\tilde{y}_t^i - y_t) K_h(\tilde{x}_t^i - x)}{\sum_{i=1}^n K_h(\tilde{y}_t^i - y_t)}$$

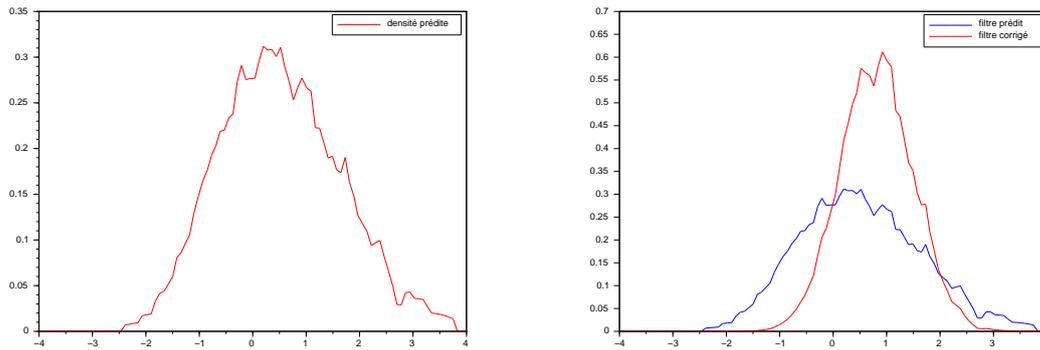


FIGURE 6.20 – Densités des filtres prédit et corrigé

Génération des particules

$$\bar{x}_t^i \sim \hat{p}_t^n(x|y_{1:t})$$

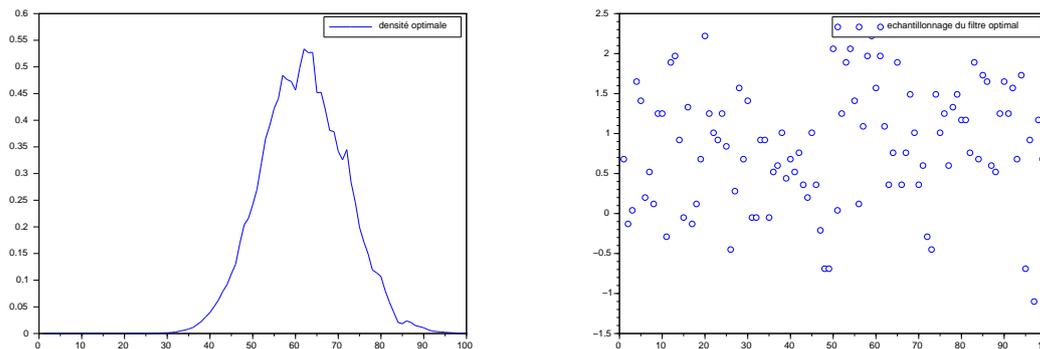


FIGURE 6.21 – Echantillonnage du filtre optimal

6.3.8 Oubli de la condition initiale

Considérons le système
$$\begin{cases} X_n = a \cdot X_{n-1} + \sigma_n \cdot W_0 \\ Y_n = \sqrt{1 + X_n^2} + V_n \end{cases} \quad \text{où } W_0 \sim \mathcal{N}(0, 1).$$

V_n est le bruit blanc d'observation non nécessairement gaussien.

Il s'agit d'estimer le filtre optimal et les paramètres a et σ_n .

Les paramètres peuvent être estimés par

$$\hat{a}_n = \frac{\sum_{k=1}^n \hat{X}_{k-1} \cdot \hat{X}_k}{\sum_{k=1}^n \hat{X}_{k-1}^2}$$

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n (\hat{X}_k - \hat{a}_n \cdot \hat{X}_{k-1})^2$$

où \hat{X}_k est l'estimation de X à la k -ième itération. Ces formules sont mises en défaut dans le cas où la première valeur de X est nulle.

X_0 suit une loi gaussienne centrée.

a_0 et σ_W^0 suivent des lois uniformes.

Dans la figure suivante, avec $a_0 \sim \mathcal{U}[0.7; 1.1]$ et $\sigma_W^0 \sim \mathcal{U}[1.7; 2.1]$, on obtient les estimations par noyau de convolution d'Epanechnikov avec $h = 1$ des densités :

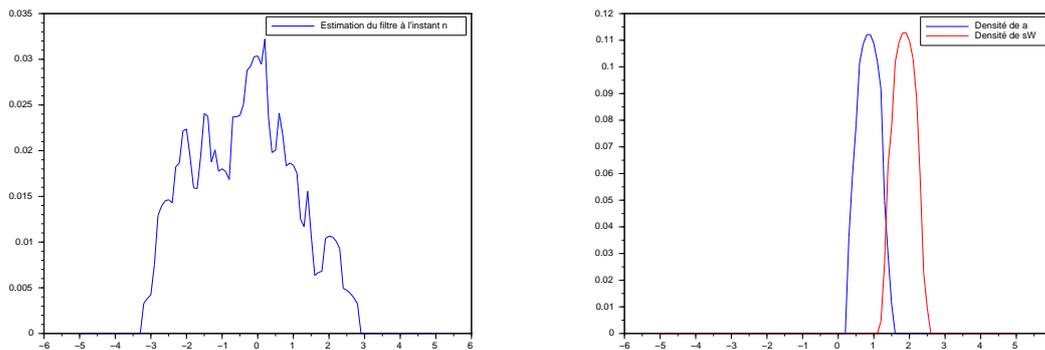


FIGURE 6.22 – Densités du filtre et des paramètres

Considérons les cas suivants :

Y est la valeur observée,

\hat{X} est la moyenne estimée du filtre,

σ_X est l'écart-type du filtre

avec 9 itérations, on a :

Conditions initiales : $X_0 = 0$, $S_{01} = 0.8$, $a \sim \mathcal{U}[0.7; 1.1]$ et $\sigma_W \sim \mathcal{U}[1.7; 2.1]$

n	Y	\hat{X}	\hat{s}_X	\hat{a}	\hat{s}_W
1	1.7966707	- 0.4062612	1.6542424	0.9	1.9
2	0.7437838	0.2665080	2.1920109	0.85	1.85
3	1.5760042	0.1714222	1.7913355	0.9	1.9
4	2.6075117	- 1.0083074	1.8484788	0.9	1.85
5	2.7045816	0.5833369	1.8057736	0.85	1.9
6	0.3063468	- 0.0842307	1.1799115	0.9	1.9
7	2.1206878	0.4520169	1.4853735	0.85	1.9
8	0.0927576	0.7719133	0.7602809	0.85	1.85
9	0.5409059	0.6236139	1.5701402	0.9	1.9

Conditions initiales : $X_0 = 0$, $S_{01} = 1.2$, $a \sim \mathcal{U}[0.7; 1.1]$ et $\sigma_W \sim \mathcal{U}[1.7; 2.1]$

n	Y	\hat{X}	\hat{s}_X	\hat{a}	\hat{s}_W
1	1.7966707	- 0.0033626	1.8744165	0.9	1.9
2	0.7437838	- 0.2227299	1.5361767	0.85	1.9
3	1.5760042	0.6196535	2.0131198	0.85	1.9
4	2.6075117	- 0.8126852	1.196278	0.85	1.85
5	2.7045816	0.2673170	1.993008	0.85	1.9
6	0.3063468	- 0.5451384	1.631883	0.85	1.8
7	2.1206878	- 0.6748219	1.9034892	0.85	1.85
8	0.0927576	- 0.2202775	1.4767549	0.9	1.85
9	0.5409059	- 0.1002396	1.5885445	0.85	1.9

Conditions initiales : $X_0 = 0$, $S_{01} = 2$, $a \sim \mathcal{U}[0.7; 1.1]$ et $\sigma_W \sim \mathcal{U}[1.7; 2.1]$

n	Y	\hat{X}	\hat{s}_X	\hat{a}	\hat{s}_W
1	1.7966707	0.5860447	2.2229572	0.9	1.9
2	0.7437838	- 0.1303366	2.3378443	0.9	1.95
3	1.5760042	0.0212632	1.8241213	0.9	1.9
4	2.6075117	0.8042460	2.4733732	0.9	1.9
5	2.7045816	0.0391342	1.6620968	0.9	1.9
6	0.3063468	0.2627481	1.5421818	0.9	1.9
7	2.1206878	- 0.3670376	2.0092802	0.9	1.9
8	0.0927576	- 0.1609236	1.881182	0.9	1.9
9	0.5409059	0.3967783	1.2610719	0.9	1.9

TABLE 6.6 – Estimations des paramètres avec $\sigma_0 = 0.8, \sigma_0 = 1.2, \sigma_0 = 2 \sigma_W = 0.2$ et $\sigma_a = 0.2$

La condition initiale varie avec l'écart type de X_0 avec 0.8, 1.2 et 2. On constate que les valeurs de les moyennes du filtre sont très proches de 0 et d'écart type proches de 2. Les valeurs du paramètre a sont pratiquement égales à 0.9 et les valeurs des l'écart type du bruit sont pratiquement égales à 1.9.

Conditions initiales : $X_0 = 0$, $S_{01} = 0.8$, $a \sim \mathcal{U}[0.8; 1]$ et $\sigma_W \sim \mathcal{U}[1.8; 2]$

n	Y	\hat{X}	\hat{s}_X	\hat{a}	\hat{s}_W
1	1.7966707	- 0.4559336	1.6590117	0.9	1.85
2	0.7437838	0.2335341	2.2282359	0.85	1.85
3	1.5760042	0.1686687	1.8034567	0.9	1.85
4	2.6075117	- 1.0949632	1.7664042	0.9	1.85
5	2.7045816	0.4674864	1.8671267	0.85	1.85
6	0.3063468	0.2715537	1.0634355	0.9	1.9
7	2.1206878	0.3692824	1.4770553	0.85	1.9
8	0.0927576	0.8731681	0.7690005	0.85	1.85
9	0.5409059	0.5682117	1.6143253	0.9	1.85

Conditions initiales : $X_0 = 0$, $S_{01} = 1.2$, $a \sim \mathcal{U}[0.8; 1]$ et $\sigma_W \sim \mathcal{U}[1.8; 2]$

n	Y	\hat{X}	\hat{s}_X	\hat{a}	\hat{s}_W
1	1.7966707	0.1140678	1.8632268	0.9	1.9
2	0.7437838	- 0.2860061	1.5806527	0.85	1.9
3	1.5760042	0.7049347	1.9188016	0.85	1.9
4	2.6075117	- 0.8702603	1.3179739	0.85	1.9
5	2.7045816	0.2028505	1.9888809	0.85	1.9
6	0.3063468	- 0.3514259	1.6886749	0.85	1.85
7	2.1206878	- 0.6989580	1.9563635	0.85	1.85
8	0.0927576	- 0.1187697	1.5131002	0.85	1.85
9	0.5409059	- 0.0922296	1.6212188	0.85	1.9

Conditions initiales : $X_0 = 0$, $S_{01} = 2$, $a \sim \mathcal{U}[0.8; 1]$ et $\sigma_W \sim \mathcal{U}[1.8; 2]$

n	Y	\hat{X}	\hat{s}_X	\hat{a}	\hat{s}_W
1	1.7966707	0.5393769	2.250504	0.85	1.85
2	0.7437838	0.0949273	2.3240905	0.9	1.9
3	1.5760042	0.1125883	1.8134498	0.9	1.85
4	2.6075117	0.8390705	2.422157	0.85	1.85
5	2.7045816	0.0655019	1.7053204	0.85	1.9
6	0.3063468	0.1139649	1.6195687	0.9	1.9
7	2.1206878	- 0.4497896	1.9548814	0.85	1.85
8	0.0927576	0.2492113	2.0065116	0.85	1.9
9	0.5409059	0.3308666	1.2457184	0.85	1.85

TABLE 6.7 – Estimations des paramètres avec $\sigma_0 = 0.8, \sigma_0 = 1.2, \sigma_0 = 2$ $\sigma_W = 0.1$ et $\sigma_a = 0.1$

On a les mêmes constatations avec les variations de l'écart type de X_0 même en réduisant les amplitudes des lois uniformes des paramètres à l'instant initial. C'est-à-dire que a est pratiquement égal à 0.9 et l'écart type du bruit s_W est pratiquement égal à 1.9.

Les figures suivantes montrent les différences entre les trajectoires du filtre où $\sigma_0 = 0.8$ et $\sigma_0 = 1.2$ dans la figure 6.23, $\sigma_0 = 0.8$ et $\sigma_0 = 2$ dans la figure 6.24, $\sigma_0 = 1.2$ et $\sigma_0 = 2$ dans la figure 6.25.

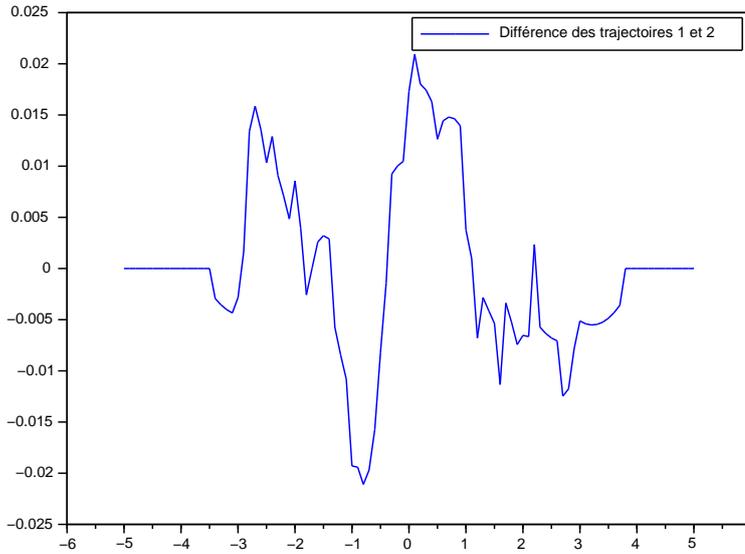


FIGURE 6.23 – Différences des trajectoires 1 et 2

L'écart entre les deux trajectoires est pratiquement nul avec une précision entre -0.02 et 0.02. Cet écart est mesuré par $y_1 = \int_{-5}^5 |f_1(x) - f_2(x)| dx$ ou $y_2 = \int_{-5}^5 |f_1(x) - f_3(x)| dx$ ou encore $y_3 = \int_{-5}^5 |f_2(x) - f_3(x)| dx$ où f_1 est l'estimation du filtre par noyau d'Epanechnikov avec $\sigma_0 = 0.8$, f_2 est l'estimation du filtre par noyau d'Epanechnikov avec $\sigma_0 = 1.2$ et f_3 est l'estimation du filtre par noyau d'Epanechnikov avec $\sigma_0 = 2$.

n	Y	y_1	y_2	y_3
1	1.7966707	0.0614542	0.0886433	0.0721347
2	0.7437838	0.1163110	0.0815137	0.1237408
3	1.5760042	0.0553561	0.0516387	0.0748611
4	2.6075117	0.0706533	0.0996173	0.1064049
5	2.7045816	0.0629586	0.0552693	0.0490538
6	0.3063468	0.0810051	0.0789582	0.1016946
7	2.1206878	0.0754551	0.0785948	0.0659318
8	0.0927576	0.1030425	0.1176607	0.0670833
9	0.5409059	0.0499907	0.0488128	0.0649433

TABLE 6.8 – Evaluation de l'écart avec $s_a = 0.2$, $s_W = 0.2$

À l'instant initiales, les particules de a sont tirées de la loi uniforme $\mathcal{U}[0.7; 1.1]$ et les particules de s_W de la loi uniformes $\mathcal{U}[1.7; 2.1]$.

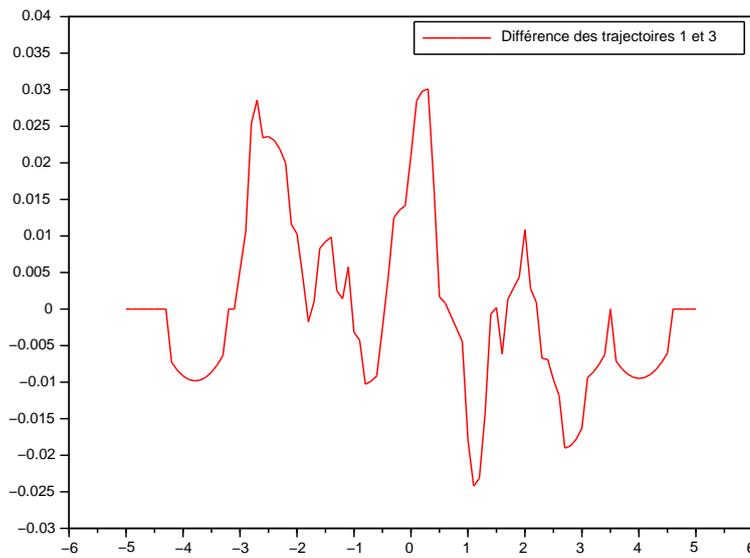


FIGURE 6.24 – Différences des trajectoires 1 et 3

Les écarts sont pratiquement nuls. L'oubli de la condition initiale est bien vérifié dans la table 6.8 et dans la table 6.9. Cette propriété justifie l'invariance des paramètres au cours du temps.

n	Y	y_1	y_2	y_3
1	1.7966707	0.0609338	0.0843802	0.0792150
2	0.7437838	0.1133568	0.0959764	0.1169030
3	1.5760042	0.0515159	0.0594181	0.0794888
4	2.6075117	0.0667960	0.1026314	0.1072182
5	2.7045816	0.0582282	0.0612660	0.0583214
6	0.3063468	0.0900029	0.0787239	0.0863065
7	2.1206878	0.0802948	0.0705191	0.0627574
8	0.0927576	0.1015698	0.1051982	0.0628485
9	0.5409059	0.0451941	0.0625455	0.0656147

TABLE 6.9 – Evaluation de l'écart avec $s_a = 0.1$, $s_W = 0.1$

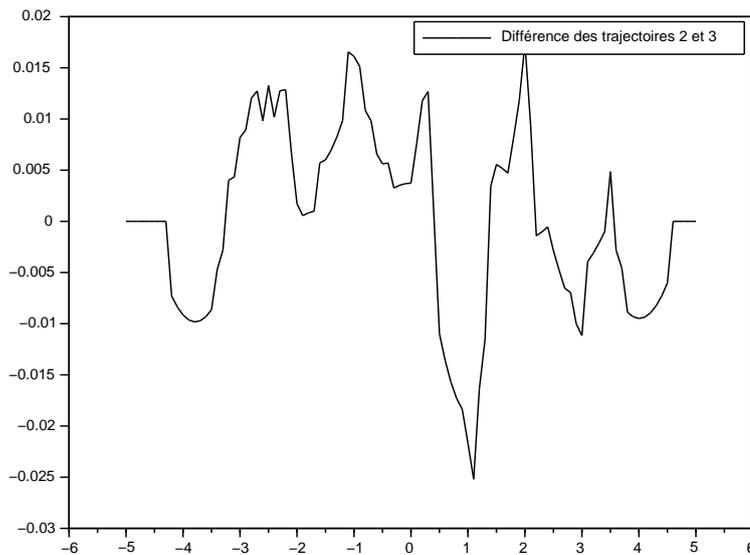


FIGURE 6.25 – Différences des trajectoires 2 et 3

6.4 Conclusion

Dans le problème de système commandé, le filtrage particulaire est la méthode adoptée. Le système est non linéaire et éventuellement, le bruit d'observation uniforme. Selon les différentes conditions initiales, les particules du filtre à chaque instant sont distribuées autour de l'estimé attendu. Il y a bien oublié de la condition initiale pour le filtrage particulaire.

Dans le problème de système paramétré, le filtrage particulaire par noyau de convolution est le mieux adapté. La régularisation des filtres permet de vérifier plus facilement la propriété d'oubli de la condition initiale.

L'inconvénient avec les filtres particuliers est dans leurs applications dans un système de dimension plus grande. En effet, les particules s'évaluent dans un espace virtuel ce qui entraîne une très grande complexité de l'algorithme.

Conclusion de la troisième partie

Dans cette partie, la poursuite de trajectoire et la commande optimale sont des applications intéressantes dans le sens où nous avons pu vérifier les théories sur le filtrage non linéaire en temps discret.

La théorie développée dans le chapitre 3 à propos de l'existence de mesure avec laquelle on a construit le filtre est justifiée. La méthode de Monte Carlo appliquée dans un espace en dimension 2 est la base de simulation.

Le filtrage de Kalman étendu et le filtrage de Uhlmann et Julier sont des méthodes adaptées à un système non linéaire gaussien. Le premier a son avantage dans l'approximation d'une fonction linéaire en fonction linéaire. Le second a son avantage dans l'approximation des états par des σ -points. Les deux filtres ont des points communs en tant que variantes du filtrage de Kalman.

Les résultats obtenus sont pratiquement identiques et l'oubli de la condition initiale est aisément vérifié.

Dans le problème de commande optimale et d'estimation paramétrique, la mise en œuvre du filtrage particulaire et du filtrage par noyau de convolution rencontre des difficultés. En effet, l'approximation particulaire est plus vraisemblable avec un nombre de particules plus important. Ce nombre de particules considérable ralentit l'exécution de l'algorithme et pourrait nuire l'idée de traiter le filtrage en temps réel d'observation. Néanmoins, ces filtres ont l'avantage de traiter des systèmes non linéaires et non gaussiens.

L'oubli de la condition initiale est aisément vérifié dans le filtrage particulaire. Comme le filtre par noyau de convolution n'est que la régularisation du filtre particulaire, l'oubli de la condition initiale s'en suit également.

On peut très bien appliquer les filtres particuliers en traitement d'images. Comme dans les filtres de Kalman, les filtres particuliers sont appliqués en fonction de la modélisation du problème à étudier. Le choix du filtrage à utiliser dépend de cette modélisation surtout dans les formes des bruits d'état et d'observation.

Conclusion générale

Sommaire

Résultats	161
Discussions et perspectives	162

Résultats

En traitement de signal, le filtrage est une méthode incontournable dans l'étude du signal non observé. Cette méthode rencontre des difficultés dans son application. Elle nécessite des hypothèses assez lourdes qui sont difficilement justifiées dans des cas plus réels. La mise en œuvre dans les applications est assez complexe.

Théoriquement, le signal à étudier est régi par une équation différentielle ou une équation aux dérivées. Dans ce mémoire, les observations sont partielles. Ceci est dû à l'adaptation pratique des mesures. Un grand nombre de mesures d'observations est nécessaire mais ralentit considérablement la vitesse des calculs.

Dans la construction du modèle, il est utile de faire une hypothèse sur la loi a priori de la condition initiale. L'étude de stabilité justifie la vraisemblance des filtres obtenus dans l'application des différentes méthodes.

Un cas particulier des bruits d'observation uniforme fait ressortir une des qualités intéressantes du filtrage particulière.

L'existence de la solution de filtrage dépend de la structure markovienne du signal étudié. Cette structure, qui est nécessaire dans la décomposition classique prédiction-corréction, permet aussi d'introduire le théorème de Girsanov assurant l'existence et l'unicité presque sûrement.

Une solution théorique avec le filtrage de Kalman permet de justifier les théories pour les modèles linéaires et non linéaires mais avec des hypothèses assez fortes. Dans le cas de forte non linéarité du modèle, cette méthode n'est plus applicable. Le filtrage de Kalman inodore peut résoudre le problème de non régularité des fonctions dans le modèle mais un modèle dans un espace de grande dimension est assez compliqué à traiter numériquement.

Par contre, le filtrage particulière est une des méthodes assez simple dans l'approximation numérique avec les théorèmes classiques de convergence comme les lois de grands nombres et le théorème de la limite centrale. Mais dans la vérification des différentes propriétés comme la stabilité, cette méthode de filtrage manque de précision.

L'introduction du filtrage par noyau de convolution dans le filtrage particulaire est une solution intéressante dans le sens où l'échantillonnage des particules se fait avec des densités régularisées et finalement les approximations sont plus précises.

Cette thèse permet de montrer quelques aspects de filtrages non linéaires. Elle a permis de voir les difficultés dans l'implémentation des programmes non seulement dans la simulation mais surtout dans la justification des théories. Dans la poursuite de trajectoire ou dans l'estimation paramétrisque ou encore dans l'estimation des commandes, les théories ont pu être vérifiées et cela peut être étendu dans d'autres phénomènes beaucoup plus complexes.

Des méthodes utilisant les éléments finis ou les réseaux de neurones ont été déjà introduites dans [CAMPILLO,1986] et [ROSSI et VILLA, 2004]. On pourrait très bien approfondir ces recherches.

Discussions et perspectives

La complexité dans les différents programmes n'est pas étudié. La méthode de "diviser pour regner" pourrait être appliquée pour améliorer cette complexité. Dans les applications, le nombre de particules n'est pas assez important, de l'ordre d'une centaine. L'augmentation du nombre de particules entraîne naturellement un ralentissement dans l'exécution d'un côté mais de l'autre côté, on améliore la précision et la convergence des méthodes.

Le filtrage non linéaire est une méthode assez riche dans la résolution de différents problèmes classiques comme les estimations paramétriques et les estimations des commandes. On pourrait traiter les problèmes de commandes optimales dans un horizon fini en considérant une statistique adéquate. Le suivi en temps réel avec la décomposition prédiction-réduction n'est plus nécessaire mais la prévision en un instant ultérieur gagne son intérêt.

La méthode particulaire pourrait être appliquée pour la résolution des équations différentielles et équations aux dérivées partielles. La méthode des éléments finis appliquées aux particules aurait une certaine place intéressante dans ces résolutions.

Les bruits blancs intervenant dans les modèles sont soit uniformes, soit gaussiens. Le choix de ces bruits n'est pas totalement justifié. Comme les signaux ne sont que partiellement observés, des bruits définis par des lois discrètes pourrait être introduits dans les modèles. Le bruit blanc poissonnien serait un bruit perturbateur intéressant. Ce bruit intervient surtout dans des processus de comptage.

Naturellement, Les bruits de perturbations dans les modèles sont blancs. Cet état joue un grand rôle dans les applications . L'introduction des bruits colorés pourrait être un obstacle dans la justification des théories.

Quatrième partie

Annexes

Annexe A

Méthode de Monte-Carlo de chaînes de Markov

A.1 Bayes et Simulation

Pour estimer un paramètre inconnu, considérons un échantillon $x = (x_1, \dots, x_n)$ obtenu par une suite de variables aléatoires $(X_i, 1 \leq i \leq n)$ i.i.d de densité commune $f(x|\theta)$. La vraisemblance¹ est telle que

$$L(x|\theta) = \prod_{i=1}^n f(x_i|\theta)$$

et la log-vraisemblance est telle que :

$$\ln L(x|\theta) = \sum_{i=1}^n \ln f(x_i|\theta)$$

L'estimateur du maximum de vraisemblance de θ est défini par :

$$\frac{\partial}{\partial \theta} \ln L(x|\theta) = 0$$

Une estimation de θ est telle que :

$$\hat{\theta} = \arg \max L(x|\theta)$$

A.1.1 Exemple : loi Gamma

La loi Gamma est de densité :

$$f(x|\alpha, \beta) = \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha) \beta^\alpha} \mathbb{I}_{\mathbb{R}^+}(x)$$

Sa log-vraisemblance est donc :

$$\ln L(x|\alpha, \beta) = -n \ln \Gamma(\alpha) - n\alpha \ln \beta + (\alpha - 1) \sum_{i=1}^n \ln x_i - \frac{1}{\beta} \sum_{i=1}^n x_i$$

1. densité de la loi de l'échantillon

L'estimateur du maximum de vraisemblance de β est si α est connu :

$$\hat{\beta} = \frac{1}{n\alpha} \sum_{i=1}^n x_i$$

Si α est inconnu, on calcule cet estimateur par :

$$\frac{\partial}{\partial \alpha} \ln L(x|\alpha, \beta) = 0$$

$$\frac{\partial}{\partial \beta} \ln L(x|\alpha, \beta) = 0$$

A.1.2 Modèles de mélange

Si X_i suit la loi de densité $f_j(x_i)$ avec la probabilité p_j , considérons la densité

$$f(x_i) = p_1 f_1(x_i) + \dots + p_k f_k(x_i)$$

Pour un échantillon (x_1, \dots, x_n) la vraisemblance est :

$$L(x|\theta, p) = \prod_{i=1}^n [p_1 f_1(x_i) + \dots + p_k f_k(x_i)]$$

La fonction est multimodale et l'estimateur du maximum de vraisemblance n'est plus adapté. L'estimateur de θ est obtenu par la méthode bayésienne.

A.1.3 Méthodes bayésiennes

La vraisemblance est

$$f(x|\theta) = \prod_{i=1}^n f(x_i|\theta)$$

Notons la loi à priori de θ par $\pi(\theta)$ et on construit la loi à posteriori de θ telle que

$$\text{avec la formule de Bayes, } f(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\mathbb{R}^d} f(x|\theta)\pi(\theta)d\theta}$$

$$\text{où } \int_{\mathbb{R}^d} f(x|\theta)\pi(\theta)d\theta \text{ est la loi marginale de } x.$$

Loi normale

La vraisemblance est :

$$f(x|\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x_i - \mu)^2}{2\sigma^2} \right]$$

La loi à priori est gaussienne : $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$ avec

$$\pi(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp \left[-\frac{(\mu - \mu_0)^2}{2\sigma_0^2} \right]$$

La loi à posteriori est gaussienne : $\mu|x \sim \mathcal{N}(\mu_n, \sigma_n^2)$ avec

$$\mu_n = \left[\frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} \right] \frac{1}{n} \sum_{i=1}^n x_i + \left[\frac{\sigma^2}{\sigma^2 + n\sigma_0^2} \right] \mu_0$$

Par définition, une loi à priori $\pi(\theta)$ est conjuguée si $f(x|\theta)$ et $\pi(\theta)$ appartiennent à la même famille de lois.

Pour le cas gaussien,

$$f(x|m, \sigma^2) \propto \frac{1}{(\sigma^2)^{n/2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2 \right]$$

La loi conjuguée pour m est la loi normale $m \sim \mathcal{N}(\mu, \beta_0^2)$.

La loi conjuguée pour σ^2 est la loi inverse gamma,

$$\pi(\sigma^2|\kappa, \gamma) \propto \frac{1}{(\sigma^2)^{\kappa+1}} \exp \left(-\frac{\gamma}{\sigma^2} \right)$$

A.1.4 Algorithme d'acceptation-rejet

Soit une loi d'intérêt de densité f et une loi de proposition de densité g telle que :

$$f(x) \leq Mg(x)$$

sur le support de f . Alors, on peut simuler suivant f avec l'algorithme suivant :

1. Générer $X \sim g$ et $U \sim \mathcal{U}([0, 1])$
2. Accepter $Y = X$ si

$$U \leq \frac{f(X)}{Mg(X)}$$

3. Retourner en 1 si rejet

Exemple : Cauchy-Normale

La loi cible est

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

La loi de proposition est

$$g(x) = \frac{1}{\pi} \frac{1}{1+x^2}$$

M est choisi tel que

$$\frac{f(x)}{g(x)} = \sqrt{\frac{\pi}{2}} (1+x^2) \exp(-x^2/2) \leq \sqrt{\frac{2\pi}{e}} = 1.52$$

Valeur atteinte en ± 1 . Probabilité d'acceptation $1/M \simeq 0.66$.

A.1.5 Echantillonnage d'importance

h est une fonction mesurable bornée et f la densité de Θ

$$\mathbb{E}[h(\Theta)] = \int_{\Omega} \left[h(\theta) \frac{f(\theta)}{g(\theta)} \right] g(\theta) d\theta$$

g est facilement simulable.

On génère un échantillon $(\theta_1, \dots, \theta_n)$ distribué suivant g pour approcher l'intégrale :

$$\mathbb{E}[h(\Theta)] \simeq \frac{1}{m} \sum_{j=1}^m \frac{f(\theta_j)}{g(\theta_j)} h(\theta_j)$$

La loi de g doit être simple à simuler. Si le support de g contient celui de f , l'estimateur converge vers

$$\int_{\Omega} h(\theta) f(\theta) d\theta$$

La variance de l'estimateur est finie si

$$\mathbb{E} \left[h^2(\Theta) \frac{f(\Theta)}{g(\Theta)} \right] < \infty$$

A.2 Metropolis-Hastings

Pour approcher l'intégrale

$$\int_{\Omega} h(\theta) f(\theta) d\theta,$$

il n'est pas nécessaire de simuler suivant f . Le principe des méthodes MCMC est de construire une chaîne de Markov ergodique dont la loi stationnaire est f .

- On part d'une valeur $\theta^{(0)}$ et on construit $\theta^{(t)}$ à l'aide d'un noyau de transition tel que la loi cible est f .
- Pour t_0 assez grand, $\theta^{(t_0)}$ est distribué suivant f .
- Les valeurs générées $\theta^{(t_0)}, \theta^{(t_0+1)}, \dots$ sont dépendantes car $\theta^{(t)}$ est une chaîne de Markov.

Hypothèses

On connaît la loi cible f à une constante multiplicative près.

On définit une loi de proposition, ou loi instrumentale, $q(y|\theta)$

Algorithme

- Initialisation : choix de $\theta^{(0)}$
- À partir de $\theta^{(t)}$, on génère $y^{(t)}$ à l'aide de la loi de proposition et on accepte ou rejette cette valeur de $y^{(t)}$ à l'aide d'une procédure d'acceptation-rejet. La valeur retenue est notée $\theta^{(t+1)}$
- Les premières valeurs générées par l'algorithme ne seront pas utilisées pour l'inférence (burn-in)

Metropolis-Hastings

Étant donné $\theta^{(t)}$,

-Générer $y_t \sim q(y|\theta^{(t)})$.

-Accéptation-rejet $\theta^{(t+1)} = \begin{cases} y_t & \text{avec proba. } \rho(\theta^{(t)}, y_t), \\ \theta^{(t)} & \text{avec proba. } 1 - \rho(\theta^{(t)}, y_t), \end{cases}$

$$\rho(\theta, y) = \min \left\{ \frac{f(y) q(\theta|y)}{f(\theta) q(y|\theta)}, 1 \right\}$$

Pour le cas symétrique

$$\rho(\theta, y) = \min \left\{ \frac{f(y)}{f(\theta)}, 1 \right\}$$

On accepte toujours les valeurs de y_t augmentant la densité cible.

La loi cible f peut être connue à une constante multiplicative près.

La chaîne $(\theta^{(t)})_t$ peut prendre plusieurs fois la même valeur

Convergence

Probabilité d'acceptation

$$\mathbb{P} \left[\frac{f(y_t) q(\theta^{(t)}|y_t)}{f(\theta^{(t)}) q(y_t|\theta^{(t)})} \geq 1 \right] < 1.$$

C'est à dire que l'événement $\{\theta^{(t+1)} = \theta^{(t)}\}$ est possible.

Loi de proposition

$q(y|\theta) > 0$ pour tout (θ, y) .

En particulier, le support de la loi de proposition doit inclure le support de la loi cible.

Ergodicité

Pour h tel que $\mathbb{E}_f[|h(\Theta)|] < \infty$,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T h(\theta^{(t)}) = \int h(\theta) \cdot f(\theta) d\theta$$

Convergence en variation totale

$$\lim_{T \rightarrow \infty} \left\| \int K^n(\theta, \cdot) \mu(d\theta) - f \right\|_{TV} = 0$$

pour toute loi initiale μ , $K^n(\theta, \cdot)$ est le noyau de la chaîne après n transitions.

En particulier

$$\lim_{T \rightarrow \infty} \mathbb{P}[\theta^{(t)} \in A] = \int_A f(\theta) d\theta$$

Cas indépendant

La loi de proposition $q(y|\theta^{(t)})$ est indépendante de $\theta^{(t)}$.
Étant donnée $\theta^{(t)}$,

Générer $y_t \sim q(y)$

$$\text{Accéptation-rejet : } \theta^{(t+1)} = \begin{cases} y_t & \text{avec proba. } \min \left\{ \frac{f(y_t)}{f(\theta^{(t)})} \frac{q(\theta^{(t)})}{q(y_t)}, 1 \right\}, \\ \theta^{(t)} & \text{sinon} \end{cases},$$

Marche aléatoire

La loi de proposition q est telle que

$$y_t = \theta^{(t)} + \epsilon_t$$

où ϵ_t indépendant de $\theta^{(t)}$, c'est-à-dire que $q(y|\theta) = q(y-\theta)$; Si q est symétrique, on obtient l'algorithme suivant :

Étant donnée $\theta^{(t)}$,

Générer $y_t \sim q(y - \theta^{(t)})$

$$\text{Accéptation-rejet : } \theta^{(t+1)} = \begin{cases} y_t & \text{avec proba. } \min \left\{ \frac{f(y_t)}{f(\theta^{(t)})}, 1 \right\}, \\ \theta^{(t)} & \text{sinon} \end{cases},$$

Exemple loi normale

Simulation des données suivant la loi normale $\mathcal{N}(0, 1)$.

Métropolis-Hastings-Indépendant avec $q(y) \sim \mathcal{U}[-3, +3]$.

Algo de Metropolis6Hastings-Marche aléatoire avec

$$q(\epsilon_t) \sim \mathcal{U}[-\delta, +\delta]$$

Probabilité d'accéptation

$$\min\{\exp\{(\theta^{(t)})^2 - y_t^2\}, 1\}$$

L'échantillonneur de Gibbs

Pour simuler suivant une loi $f(\theta)$ avec $\theta = (\theta_1, \dots, \theta_p)$, on peut utiliser l'algorithme suivant :

- Initialisation : générer un vecteur $\theta = (\theta_1, \dots, \theta_p)$ suivant une loi de proposition π_0

- Simuler suivant les lois conditionnelles pour $i = 1, 2, \dots, p$

$$\theta_i | \theta_1, \theta_2, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p \sim f_i(\theta_i | \theta_1, \theta_2, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p)$$

Plus précisément,

Étant donné $\theta^{(t)} = (\theta_1^{(t)}, \dots, \theta_p^{(t)})$,

- Générer $\theta_1^{(t+1)} \sim f_1(\theta_1 | \theta_2^{(t)}, \dots, \theta_p^{(t)})$,

- Générer $\theta_2^{(t+1)} \sim f_2(\theta_2 | \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_p^{(t)})$,

.....

- Générer $\theta_p^{(t+1)} \sim f_p(\theta_p | \theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{p-1}^{(t+1)})$,

Seules les lois conditionnelles f_1, \dots, f_p sont utilisées pour la simulation. Donc, même pour un problème de grande dimension, toutes les simulations sont univariées.

Cas gaussien

$$f(x|m, \sigma^2) \propto (\sigma^2)^{-n/2} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2 \right]$$

Lois à priori

Moyenne

$$m \sim \mathcal{N}(m_0, \sigma_0^2)$$

Variance

$$\sigma^2 \sim \mathcal{IG}(\alpha, \beta)$$

Lois conditionnelles

Moyenne

$$m | \sigma^2, x \sim \mathcal{N}(M, \Sigma^2)$$

avec

$$M = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} \left(\frac{1}{n} \sum_{i=1}^n x_i \right) + \left(\frac{\sigma^2}{\sigma^2 + n\sigma_0^2} \right) m_0$$

et

$$\Sigma^2 = \frac{\sigma^2 \sigma_0^2}{n\sigma_0^2 + \sigma^2}$$

Variance

$$\sigma^2 | m, x \sim \mathcal{IG} \left(\frac{n}{2} + \alpha, \frac{1}{2} \sum_{i=1}^n (x_i - m)^2 + \beta \right)$$

Completion

La densité g est une complétion de f si

$$\int g(\theta, \eta) d\eta = f(\theta),$$

c'est-à-dire si f est une loi marginale de g .

Les lois conditionnelles de g sont parfois plus simples à simuler que celles de f .

Annexe B

Chaînes de Markov cachées

B.1 Introduction

X_n est une chaîne de Markov de loi initiale ρ et de matrice de transition Q . X_n est une variable latente. Y_n est la variable d'observation de loi d'émission ψ . Conditionnellement aux états X_n , les observations Y_n sont mutuellement indépendantes et que chaque observation Y_n ne dépend que de X_n .

$$\mathbb{P}(X_0 = x) = \rho(x) \quad (\text{B.1})$$

$$\mathbb{P}(X_{n+1} = x' | X_n = x) = Q(x, x') \quad (\text{B.2})$$

$$\mathbb{P}(Y_n = y | X_n = x) = \psi(x, y) \quad (\text{B.3})$$

Étant donné une chaîne de Markov X_n non observable, on dispose d'une suite d'observations Y_n fonction de X_n telle que :

$$Y_n = h_n(X_n) + V_n$$

V_n est un bruit blanc additif.

Pour assurer l'existence d'une solution unique, on utilise des informations supplémentaires [Le GLAND, 2011] sur la suite cachée. On cherche à minimiser le critère :

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

où le terme d'attache aux données est :

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

ou

$$d_k(x) = \frac{1}{2} (Y_k - h_k(x_k))^t I_k (Y_k - h_k(x_k))$$

Et l'information à priori est telle que :

$$c_0(x) = \frac{1}{2} |x - \mu|^2$$

ou

$$c_0(x) = \frac{1}{2} (x - \mu)^t \Sigma_0^{-1} (x - \mu)$$

et

$$c_k(x, x') = \frac{1}{2}|x' - f_k(x)|^2$$

ou

$$c_k(x, x') = \frac{1}{2}(x' - f_k(x))^t Q_k^{-1}(x' - f_k(x))$$

L'état x_k doit être proche de $f_k(x_{k-1})$ et x_0 est proche de μ .

Les fonctions coûts sont, à une constante additive près, de la forme :

$$c_0(x) = -\log p_0(x)$$

et

$$c_k(x, x') = -\log p_k(x'|x)$$

Et finalement, le critère à minimiser est :

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

ce qui revient à minimiser,

$$\exp[J_{x_{0:n}}] = \underbrace{p_0(x_0) \prod_{k=1}^n p_k(x_{k-1}, x_k)}_{p_{0:n}(x_{0:n})} \exp \left[\sum_{k=0}^n d_k(x_k) \right].$$

ou encore,

$$\exp[J_{x_{0:n}}] = p_0(x_0) \prod_{k=1}^n p_k(x_{k-1}, x_k) \exp \left[\sum_{k=0}^n \frac{1}{2} |Y_k - h_k(x_k)|^2 \right].$$

B.2 Algorithmes

$$\begin{aligned} \mathbb{P}(X_{0:n} = x_{0:n}, Y_{0:n} = y_{0:n}) &= \mathbb{P}(X_0 = x_0, \dots, X_n = x_n, Y_0 = y_0, \dots, Y_n = y_n) \\ &= \mathbb{P}(X_0 = x_0, \dots, X_n = x_n) \mathbb{P}(Y_0 = y_0, \dots, Y_n = y_n | X_0 = x_0, \dots, X_n = x_n) \\ &= \rho(x_0) \cdot \prod_{k=1}^n Q(x_{k-1}, x_k) \cdot \mathbb{P}(Y_0 = y_0 | X_0 = x_0) \dots \mathbb{P}(Y_n = y_n | X_n = x_n) \\ &= \rho(x_0) \cdot \psi(x_0, y_0) \cdot \prod_{k=1}^n Q(x_{k-1}, x_k) \cdot \psi(x_{k-1}, x_k) \end{aligned}$$

B.2.1 Algorithme forward de Baum

L'algorithme forward de Baum [CAMPILLO,2006] permet de calculer la vraisemblance de la suite d'observations. Il permet aussi de déterminer l'estimation de l'état courant X_n sachant les observation $Y_n, k \leq k \leq n$: c'est le problème de filtrage.

Notons la loi jointe de (Y_0, \dots, Y_n, X_n) par :

$$\alpha_x^n(y_{0:n}) = \mathbb{P}(Y_{0:n} = y_{0:n}, X_n = x)$$

$$\alpha_x^n = \mathbb{P}(Y_{0:n}, X_n = x)$$

Initialisation

$$\alpha_{x_0}^0(y_0) = \mathbb{P}(Y_0 = y_0, X_0 = x_0) = \rho(x_0) \cdot \psi(x_0, y_0)$$

pour $i = 1 : n$

$$\begin{aligned} \alpha_{x_{k+1}}^{k+1}(y_0, \dots, y_{k+1}) &= \mathbb{P}(Y_0 = y_0, \dots, Y_{k+1} = y_{k+1}, X_{k+1} = x_{k+1}) \\ &= \mathbb{P}(X_{k+1} = x_{k+1}) \cdot \mathbb{P}(Y_0 = y_0, \dots, Y_{k+1} = y_{k+1} | X_{k+1} = x_{k+1}) \\ &= \sum_{x_k \in E} \mathbb{P}(X_k = x_k, X_{k+1} = x) \cdot \mathbb{P}(Y_0 = y_0, \dots, Y_{k+1} = y_{k+1} | X_{k+1} = x_{k+1}) \\ &= \sum_{x_k \in E} Q(x_k, x_{k+1}) \cdot \mathbb{P}(X_k = x_k) \cdot \mathbb{P}(Y_0 = y_0, \dots, Y_k = y_k) \mathbb{P}(Y_{k+1} = y_{k+1} | X_{k+1} = x_{k+1}) \\ &= \psi(x_{k+1}, y_{k+1}) \sum_{x_k \in E} Q(x_k, x_{k+1}) \cdot \mathbb{P}(Y_0 = y_0, \dots, Y_k = y_k, X_k = x_k) \\ \alpha_x^{k+1}(y_0, \dots, y_{k+1}) &= \psi(x_{k+1}, y_{k+1}) \sum_{x_k \in E} Q(x_k, x_{k+1}) \cdot \alpha_{x_k}^k(y_0, \dots, y_k) \end{aligned}$$

D'où la vraisemblance

$$\mathbb{P}(Y_0 = y_0, \dots, Y_n = y_n) = \sum_{x \in E} \alpha_x^n(y_0, \dots, y_n)$$

et le filtrage

$$\mathbb{P}(X_n = x | Y_0 = y_0, \dots, Y_n = y_n) = \frac{\alpha_x^n(y_0, \dots, y_n)}{\sum_{x \in E} \alpha_x^n(y_0, \dots, y_n)}$$

B.2.2 Algorithme backward de Baum

Avec l'algorithme forward de Baum, l'algorithme backward de Baum [CAMPILLO,2006] permet d'estimer un état intermédiaire X_k , $k \leq n - 1$: c'est le problème de lissage.

Notons la loi conditionnelle de (Y_{k+1}, \dots, Y_n) sachant $X_k = x$ par :

$$\mathbb{P}(Y_{k+1} = y_{k+1}, \dots, Y_n = y_n | X_k = x) = \beta_x^k(y_{k+1}, \dots, y_n)$$

Initialisation

$$\beta_x^n = 1$$

pour $i = n : k$

B.2.3 Algorithme de Viterbi

L'algorithme de Viterbi [CAMPILLO,2006] permet d'estimer la trajectoire complète de $X_{0:n}$.

B.2.4 Algorithme de Baum-Welch

L'algorithme de Baum-Welch [CAMPILLO,2006] permet de déterminer les paramètres du modèle.

Annexe C

Filtrage Statistique en Traitement de Signal

C.1 Processus aléatoire

Un **processus aléatoire du second ordre**, à temps discret est une suite $\{X_t, t \in \mathbb{Z}\}$ définie sur un espace de probabilité $(\Omega, \mathbf{A}, \mathbb{P})$ à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ telle que $\mathbb{E}(|X_t|^2) < \infty, \forall t \in \mathbb{Z}$.

La **fonction moyenne** est : $\mu(t) = \mathbb{E}X_t \in \mathbb{R}^d$.

La **fonction covariance** est : $K(s, t) = \mathbb{E}[(X_s - \mu(s))(X_t - \mu(t))^t]$ une matrice symétrique $d \times d$.

Avec le produit scalaire défini sur $L^2(\Omega, \mathbf{A}, \mathbb{P})$,

$$\mu(t) = \langle X_t, 1 \rangle \quad (\text{C.1})$$

$$K(s, t) = \langle X_s, X_t \rangle - \mu(s) \cdot \mu(t) \quad (\text{C.2})$$

C.1.1 Processus stationnaire

Un processus $X = (X_t)_{t \in \mathbb{Z}}$ est strictement stationnaire ou fortement stationnaire si $\forall t_1 < \dots < t_n$, et pour tout $s \in \mathbb{Z}$, $(X_{t_1+s}, \dots, X_{t_n+s})$ a même loi de probabilité que $(X_{t_1}, \dots, X_{t_n})$.

Un processus $X = (X_t)_{t \in \mathbb{Z}}$ est donc **fortement stationnaire** si les lois jointes $(X_{t_1}, \dots, X_{t_k})$ sont invariantes par translation dans le temps.

Un processus $X = (X_t)_{t \in \mathbb{Z}}$ est stationnaire au second ordre ou stationnaire au sens faible si :

- $\mathbb{E}X_t^2 < \infty$

- $\mathbb{E}X_t = m$ est indépendante de t .

- $K(t, t+s) = \mathbb{E}[(X_{t+s} - m)(X_t - m)^t] = \mathbb{E}[(X_t - s)(X_0 - m)^t] = K(0, n)$.

Un processus $X = (X_t)_{t \in \mathbb{Z}}$ est donc **faiblement stationnaire** si sa la fonction moyenne et la fonction covariance sont invariantes par translation dans le temps.

Dans ce cas, la fonction moyenne μ est constante égale à m , la fonction covariance $K(t, s)$ ne depend que de $s - t$ et il existe une fonction γ telle que $K(t, s) = \gamma(s - t)$.

La variance de X_t est telle que :

$Var X_t = \gamma(0)$ et notons la **fonction de corrélation** par :

$$\rho(t) = \frac{\gamma(t)}{\gamma(0)}$$

C.1.2 Bruit blanc

Un processus $X = (X_t)_{t \in \mathbb{Z}}$ est à accroissements indépendants si pour tous, $t_1 < \dots < t_n$, les variables aléatoires $X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ sont indépendantes.

Un bruit blanc est un processus stationnaire à accroissements indépendants, un processus i.i.d, variables indépendantes et identiquement distribuées.

Un processus $\varepsilon = (\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc s'il satisfait les deux conditions suivantes :

- $\mathbb{E} \varepsilon_t = 0$
- $\gamma(t-s) = \mathbb{E} \varepsilon_t \varepsilon_s = \sigma^2 \cdot \delta_s^t$ avec $\gamma(0) = \sigma^2$.

Les bruits blancs sont des processus stationnaires particuliers sans mémoire. Le terme bruit blanc provient de l'analogie dans le domaine des fréquences entre la densité spectrale d'une variable i.i.d et le spectre de la lumière blanche dans le spectre des couleurs.

C.1.3 Processus stationnaire d'ordre 2

Théorème de décomposition de Wold

Tout processus $X = (X_t)_{t \in \mathbb{Z}}$ d'ordre 2 peut être représenté sous la forme

$$X_t = \sum_{s=0}^{\infty} \psi_s \cdot \varepsilon_{t-s} + \kappa_t.$$

Les paramètres $\psi_s \in \mathbb{R}$ satisfont $\psi_0 = 1$ et $\sum_{s=0}^{\infty} \psi_s^2 < \infty$ et ε_t est un bruit blanc i.i.d $\mathcal{N}(0, \sigma_\varepsilon^2)$.

Le terme κ_t désigne la composante linéaire déterministe telle que $cov(\kappa_t, \varepsilon_{t-s}) = 0$.

L'opérateur retard L est défini tel que :

$$LX_t = X_{t-1}, \quad \forall t \in \mathbb{Z}$$

On a $L^j X_t = X_{t-j}$, $\forall j \in \mathbb{Z}$, $L^0 X_t = X_t$.

Tout processus stationnaire d'ordre 2 peut être donc représenté sous la forme ;

$$X_t = \Psi(L)\varepsilon_t + \kappa_t$$

où $\Psi(L) = \psi_0 L^0 + \psi_1 L^1 + \dots + \psi_q L^q$.

C.1.4 Processus à moyenne mobile

Le processus $(X_t, t \in \mathbb{Z})$ satisfait une représentation *MA* (Moving Average) d'ordre q , notée $MA(q)$, si et seulement si :

$$X_t = m + \sum_{j=0}^q \theta_j L^j \varepsilon_t$$

Exemples

Le processus $X_t = \varepsilon_t + 4\varepsilon_{t-3}$ est un processus $MA(3)$ de moyenne nulle.

Le processus $X_t = 2 + \varepsilon_t + 0.4 \varepsilon_{t-1}$ est un processus $MA(1)$ de moyenne 2.

Algorithme

```
function MA(a,m,N)
    k=length(a)
    e=grand(N,1,'nor',0,1)
    for j=k+1:N
        s=0
        for i=1:k
            s=s+a(i)*e(j-i)
        end
        x(j)=m+s
    end
    plot(x)
endfunction}
```

C.1.5 Processus autoregressif

Le processus $(X_t, t \in \mathbb{Z})$ satisfait une représentation *AR* (AutoRegressive) d'ordre p , notée $AR(p)$, si et seulement si :

$$\sum_{j=0}^p \phi_j L^j X_t = c + \varepsilon_t$$

On a

$$\mathbb{E}(X_t) = m = \frac{c}{1 + \phi_1 + \dots + \phi_p}$$

Exemples

$Y_t = -0.5Y_{t-1} + 0.2Y_{t-2} + \varepsilon_t$ est un processus $AR(3)$ de moyenne nulle.

$Y_t = 0.8 + 0.7Y_{t-1} + 0.25Y_{t-2} + \varepsilon_t$ est un processus $AR(2)$ de moyenne $0.8/(1-0.7-0.25)=16$.

Algorithme

C.1.6 Processus ARMA

Le processus $(X_t, t \in \mathbb{Z})$ satisfait une représentation *ARMA* d'ordre p et q , notée *ARMA*(p, q), si et seulement si :

$$\sum_{j=0}^q \theta_j L^j X_t = c + \sum_{j=0}^p \psi_j L^j \varepsilon_t$$

Exemples

$Z_t = 2 - 0.5Z_{t-1} + 0.2Z_{t-2} + \varepsilon_t - 0.5\varepsilon_{t-1}$ est un processus *ARMA*(3, 1) de moyenne 1.5385.

$Z_t - 0.2Z_{t-1} - 1.5Z_{t-2} = \varepsilon_t + 0.5\varepsilon_{t-1} + 0.6\varepsilon_{t-4}$ est un *ARMA*(2, 4) de moyenne nulle.

Algorithme

C.1.7 Filtre de moyenne mobile

L'opération de convolution

$$X_n = \sum_{k=0}^N h_k \cdot Y_{N-k} \quad (\text{C.3})$$

effectuée par le filtre de moyenne mobile $(h_k)_{0 \leq k \leq N}$ sur le processus Y au cours du temps se représente commodément à l'aide de l'opérateur de translation T tel que :

$$X_n = \sum_{k=0}^N h_k \cdot T^k \cdot Y_n = H(T)Y_n \quad (\text{C.4})$$

$H(T) = \sum_{k=0}^N h_k \cdot T^k$ un polynôme.

Notons $U = [-\pi, +\pi]$.

La **fonction réponse** du filtre de moyenne mobile est une fonction a sur U définie par :

$$a(\lambda) = \sum_{k=0}^N h_k \cdot \exp(-ik\lambda). \quad (\text{C.5})$$

La mesure spectrale du processus Y est l'unique mesure positive bornée μ sur U telle que

$$\gamma(k) = \sum_U \exp(ik\lambda) \mu(d\lambda) \quad (\text{C.6})$$

Si μ possède une densité f par rapport à la mesure de Lebesgue sur U , on dit que Y est à **densité spectrale** f .

$$f(\lambda) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma(k) \cdot \exp(-ik\lambda)$$

Pour le bruit blanc $\varepsilon = (\varepsilon_n)_{n \in \mathbb{Z}}$, la densité spectrale est

$$f(\lambda) = \frac{\sigma^2}{2\pi}$$

Proposition

Y est un processus stationnaire au second ordre, $(h_k)_{0 \leq k \leq N}$ un filtre de moyenne mobile et X le processus défini par :

$$X_n = \sum_{k=0}^N h_k X_{N-k}$$

Alors la mesure spectrale de X est

$$\mu_X(d\lambda) = |a(\lambda)|^2 \mu_Y(d\lambda)$$

où $a(\lambda)$ est la fonction réponse du filtre $(h_k)_{0 \leq k \leq N}$ et μ_Y est la mesure spectrale de Y .

C.2 Filtrage optimal

Un signal Y est observé à intervalle de temps régulier. Le traitement du signal consiste à extraire le signal utile X au vu de cette suite d'observations. On filtre le signal Y à partir d'une fonction suivant les informations a priori de Y et de X . Les données disponibles sont entachées de bruits aléatoires indésirables. Le filtre est conçu suivant certains critères. Comme la transmission de l'information n'est jamais parfaite, on exige au filtre que la variance de l'erreur d'estimation soit minimale.

Ayant observé les échantillons à l'instant 1 à n du signal Y , on cherche un traitement ou filtre h qui, appliqué à Y donne la meilleure estimation possible d'une grandeur aléatoire inconnue X_n .

On exige que le filtre soit optimal dans le sens où la variance de l'erreur d'estimation soit minimale.

Considéons un filtre h_n tel que $h_n(Y_1, \dots, Y_n)$ soit une estimation de X_n et notons

$$J_n = \mathbb{E} [(X_n - h_n(Y_1, \dots, Y_n))^2]$$

Le filtre optimal est tel que

$$\tilde{h}_n = \arg \min_{h_n \in H} J_n = [\tilde{h}_1, \dots, \tilde{h}_n] \quad (\text{C.7})$$

Où H est l'ensemble de tous les filtres.

\tilde{h}_n est le vecteur qui annule le gradient de J_n c'est-à-dire $\vec{\text{grad}}_{h_n} J_n = 0$.

$$\vec{\text{grad}}_{h_n} J_n = 0 \iff \frac{\partial}{\partial \tilde{h}_{ti}} J_n = 0$$

L'estimé optimal est donné par la valeur de l'espérance conditionnelle,

$$\hat{X}_n = \tilde{h}(Y_1, \dots, Y_n) = \mathbb{E}(X_n | Y_1, \dots, Y_n)$$

C'est à dire que l'estimation optimale de X_n à partir des observations Y_1, \dots, Y_n revient à projeter X_n sur le sous espace engendré par Y_1, \dots, Y_n .

Lemme

Si (Y_1, \dots, Y_n, X_n) un vecteur gaussien centré. Alors $\mathbb{E}(X_n|Y_1, \dots, Y_n)$ est la projection orthogonale de X_n sur $\mathcal{B}(Y_1, \dots, Y_n)$ au sens de l'Hilbert $L^2(\Omega, \mathcal{A}, \mathbb{P})$ et il existe des réels $\alpha_1, \dots, \alpha_n$ tels que

$$\mathbb{E}(X_n|Y_1, \dots, Y_n) = \sum_{i=1}^n \alpha_i \cdot Y_i.$$

Donc si X et Y sont conjointement gaussiens, \tilde{h}_n est un filtre linéaire et l'estimation de X_n est donnée par :

$$\hat{X}_n = \tilde{h}_n'(Y_n, \dots, Y_0)' = \sum_{k=0}^n \tilde{h}_k \cdot Y_{n-k}$$

Équation de Wiener-Hopf

$$\mathbb{E}(\underline{Y}_n \cdot \underline{Y}'_n) \cdot \tilde{h}_n = \mathbb{E}(X_n \cdot \underline{Y}_n) \quad (\text{C.8})$$

Le signal $\varepsilon_n = X_n - \hat{X}_n$ représentant l'erreur d'estimation minimale est appelé processus d'innovation. C'est un bruit blanc de variance

$$\mathbb{E}(\varepsilon_n^2) = \mathbb{E}X_n^2 - [\mathbb{E}(X_n \cdot \underline{Y}_n)]' \cdot [\mathbb{E}(\underline{Y}_n \cdot \underline{Y}'_n)]^{-1} [\mathbb{E}(X_n \cdot \underline{Y}_n)]$$

Signal noyé dans un bruit

Considérons W_n un bruit blanc perturbateur indépendant du signal X et de variance s_w .

$$Y_k = X_k + W_k \quad 0 \leq k \leq n \quad (\text{C.9})$$

$$\mathbb{E}(W_{n_1} \cdot W_{n_2}) = s_w \cdot \delta(n_1 - n_2)$$

$$\mathbb{E}(X_{n_1} \cdot W_{n_2}) = 0, \text{ pour tout } n_1, n_2$$

La matrice de covariance de Y est : $Q_n^y = \mathbb{E}[\mathbb{E}(\underline{Y}_n \cdot \underline{Y}'_n)] = \mathbb{E}[\mathbb{E}(X_n \cdot \underline{X}'_n)] + s_w I_n$

$$Q_n^y = Q_n^x + s_w \cdot I_n$$

Q_X est la matrice de covariance de X .

$$\mathbb{E}(X_n \cdot \underline{Y}_n) = \mathbb{E}(X_n \cdot \underline{X}_n)$$

Le filtre de Wiener s'écrit :

$$\tilde{h}_n = (Q_n^x + s_w \cdot I_n)^{-1} \cdot Q_n^y$$

Q_n^x est une matrice symétrique.

Si X est stationnaire, Q_n^x vérifie la structure de Toeplitz symétrique.

$$Q_n^x = \begin{bmatrix} r_0 & r_1 & \dots & r_n \\ r_1 & r_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_1 \\ r_n & \dots & r_1 & r_0 \end{bmatrix}$$

Signal auto régressif d'ordre p

$$X_n = \alpha_1 X_{n-1} + \dots + \alpha_p X_{n-p} + u_n$$

u_n est un bruit blanc.

La fonction de corrélation de X est telle que :

$r_k = \mathbb{E}(X_n \cdot X_{n-k})$ et on a

$$r_k = \alpha_1 \cdot r_{k-1} + \dots + \alpha_p \cdot r_{k-p}$$

Le vecteur filtre optimal est alors :

$$\tilde{h}_t = \Gamma_t^{-1} r_t = \begin{bmatrix} r_0 + s_w & r_1 & \dots & r_n \\ r_1 & r_0 + s_w & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_1 \\ r_n & \dots & r_1 & r_0 + s_w \end{bmatrix}^{-1} \cdot \begin{bmatrix} r_0 \\ r_1 \\ \vdots \\ r_n \end{bmatrix}$$

Exemple d'un processus du premier ordre

Considérons un signal inconnu autoregressif d'ordre 1

$$X_t = \alpha X_{t-1} + u_t$$

u_t est un bruit blanc de variance s_u .

On en déduit $r_k = \alpha r_{k-1} = \alpha^k r_0$.

Annexe D

Articles publiés

D.1 Oubli de la condition initiale dans un filtrage non linéaire gaussien et filtre de Kalman

Article publié dans LESA ESPA MADA-ENELSA vol1,2013 p.1-9

Résumé

Dans cet article, on étudie les filtrages non linéaires gaussiens. Le filtre optimal est déterminé à partir d'une théorie probabiliste. Comme on est dans le cas gaussien, on peut aussi appliquer le filtrage de Kalman pour avoir des estimations des paramètres du filtre. Sous certaines hypothèses, cette solution théorique permet d'affirmer qu'on peut oublier la condition initiale, c'est à dire que quelque soit la loi initiale, on obtient toujours la même loi du filtre optimal. La méthode d'approximation de Monte Carlo des intégrales sert de base de simulation pour les calculs numériques en application.

Mots-clefs Filtrage non linéaire, Filtre de Kalman, théorème de Girsanov, Approximation de Monte Carlo, Chaîne de Markov.

Abstract

In this paper, we study the filtering nonlinear Gaussian. The optimal filter is determined from a theory probabilistic. As is the case in Gaussian, one can also apply filtering Kalman estimates for of its parameters of the filter. Under certain assumptions, the theoretical solution lets say we can forget initial condition, i.e that some is the original legislation, we always get the same law of optimal filter. Method approximation Monte Carlo integral is the basis for simulation numerical calculations in applications.

Keywords Nonlinear filter, Kalman filter, Girsanov theorem, Monte Carlo approximation, Markov chain.

Introduction

Le filtrage est une opération qui consiste à estimer l'état d'un système dynamique à partir d'observations partielles et bruitées : on cherche à chaque instant n à estimer la vraie valeur de X_n , à partir des observations Y_1, \dots, Y_n en calculant le filtre optimal, la loi conditionnelle $\mu_n(dx) = \mathbb{P}_{X_n}(dx|\mathcal{Y}_n)$ de X_n sachant \mathcal{Y}_n où $\mathcal{Y}_n = \sigma(\mathcal{Y}_n)$ ¹.

Intuitivement, la solution du problème de filtrage dépend de l'état initial X_0 . La condition initiale, définie par la loi initiale de X_0 , est peut être oubliée dans le sens où en considérant une autre loi initiale $\bar{\mu}_0$, différente de la véritable loi initiale μ_0 , les filtres optimaux déduits sont asymptotiquement presque sûrement, et même dans L^1 , confondus.

D'abord, on considère une chaîne de Markov [1] $X = (X_n)_{n \in \mathbb{N}}$ non observée dans \mathbb{R}^d de loi initiale μ_0 et de matrice de transition π . La chaîne n'est pas directement observée, elle est dite cachée. On dispose d'observations $Y = (Y_n)_{n \in \mathbb{N}}$ à valeurs dans \mathbb{R}^k perturbée par une suite $V = (V_n)_{n \in \mathbb{N}}$ de bruits d'observations.

$$\begin{cases} Y_n &= g(X_n) + V_n \\ X_{n+1} &= h(X_n) + W_n. \end{cases}$$

On établit alors une relation de récurrence pour $\mu_n(dx)$ avec un algorithme récursif qui permet le traitement dans l'immédiat, la méthode classique de décomposition en deux étapes, en loi prédite $\mu_n^-(dx) = \mathbb{P}_{X_n}(dx|\mathcal{Y}_{n-1})$ puis en loi corrigée $\mu_n(dx) = \mathbb{P}_{X_n}(dx|\mathcal{Y}_n)$.

Le cas de filtre non linéaire gaussien est alors étudié dans lequel une théorie classique de calcul de probabilités permet de donner une solution théorique en établissant des relations de récurrence.

On aborde ensuite le cas de filtre de Kalman étendu, une approximation du filtre de Kalman des cas linéaires gaussiens, un algorithme permettant la détermination du filtre optimal par ses paramètres.

La partie application permet de faire une simulation pour le filtre optimal [8] et son filtre prédit et de vérifier l'oubli de la condition initiale avec différentes lois initiales gaussiennes centrées de différentes variances.

1. Filtrage non linéaire gaussien

On considère le modèle de chaînes de Markov cachées homogènes[3] tel que l'état du système non observé $X = \{X_n; n \geq 0\}$, une chaîne de Markov homogène à valeurs dans l'espace mesurable $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est de loi initiale μ_0 et de probabilité de transition π où pour toute fonction mesurable $f : \mathbb{R}^d \rightarrow \mathbb{R}^+$

$$\mathbb{E}[f(X_{n+1})|X_n = y] = \int_{\mathbb{R}^d} f(x)\pi(y; dx)$$

$$\mathbb{E}[f(X_0)] = \int_{\mathbb{R}^d} f(x)\mu_0(dx)$$

1. $\sigma(\mathcal{Y}_n)$ est la tribu engendrée des $Y_k, 0 \leq k \leq n$

et l'observation est un processus $Y = \{Y_n; n \geq 0\}$ donné par : $Y_n = h(X_n) + V_n$, avec $h : \mathbb{R}^d \rightarrow \mathbb{R}^k$ une fonction mesurable.

On montre qu'il existe une mesure de probabilité σ_n telle que sous cette mesure, X_n est une chaîne de Markov et de probabilité de transition indépendante de $\sigma(\mathcal{Y}_n)$ et que

$$\mu_n(dx) = \left(\int_{\mathbb{R}^d} \sigma_n(dy) \right)^{-1} \sigma_n(dx)$$

Cette mesure permet d'établir les relations de récurrence dans la prédiction-corréction.

$$\text{Comme } \mathbb{E}(Y_n | X_n = x) = \mathbb{E}(h(x) + V_n) = \int_{\mathbb{R}^d} (h(x) + v) f_V(v) dv,$$

la densité de Y_n sachant $X_n = x$ est $g(y) = f_V(y - h(x))$.

Notons

$$Z(x, y) := \exp \left[\langle Q^{-1}h(x), y \rangle - \frac{1}{2} |Q^{-1/2}h(x)|^2 \right] \quad (\text{D.1})$$

Pour un échantillon de taille $n : (X_1, Y_1), \dots, (X_n, Y_n)$, considérons alors la vraisemblance $L_n = Z(X_1, Y_1) \times \dots \times Z(X_n, Y_n)$. Pour tout $n \geq 1$, on montre aisément que $\mathbb{E}(L_n^{-1}) = 1$, et d'après le théorème de Radon-Nikodym [4], on définit alors la probabilité $\bar{\mathbb{P}}_n$ sur (Ω, \mathcal{F}) par

$$\frac{d\bar{\mathbb{P}}_n}{dP} = L_n^{-1}, \text{ et on construit la mesure aléatoire}$$

$$\sigma_n \text{ telle que : } (\forall f \in \mathcal{C}_b(E)), \int_{\mathbb{R}^d} f(x) \sigma_n(dx) = \bar{\mathbb{E}} [f(X_n) L_n | \mathcal{Y}_n]$$

ou encore :

$$\mathbb{E}[f(X_n) | \mathcal{Y}_n] \times \bar{\mathbb{E}}_n[L_n | \mathcal{Y}_n] = \bar{\mathbb{E}}_n[f(X_n) L_n | \mathcal{Y}_n]$$

et on a la **formule de Kallianpur-Striebel** ²

$$\mathbb{E}[f(X_n) | \mathcal{Y}_n] = \frac{\bar{\mathbb{E}} [f(X_n) L_n | \mathcal{Y}_n]}{\bar{\mathbb{E}} [L_n | \mathcal{Y}_n]}$$

$$(\forall f \in \mathcal{C}_b(E))$$

En d'autres termes,

$$\mu_n(dx) = \left(\int_{\mathbb{R}^d} \sigma_n(dy) \right)^{-1} \sigma_n(dx).$$

En appliquant le **théorème de Girsanov** [2], [3] :

Sous $\bar{\mathbb{P}}_n$, (X_0, \dots, X_n) est une chaîne de Markov de loi initiale μ_0 et de probabilité de transition π indépendante de la suite (Y_1, \dots, Y_n) .

Sous la loi $\bar{\mathbb{P}}_n$, X_1, \dots, X_n et Y_1, \dots, Y_n sont donc indépendantes et on peut alors définir une loi $\bar{\mathbb{P}}$, d'espérance $\bar{\mathbb{E}}$ de la façon suivante

$$\bar{\mathbb{E}}_n[\phi(X_1, \dots, X_n, Y_1, \dots, Y_n) | Y_1, \dots, Y_n] = \bar{\mathbb{E}}[\phi(X_1, \dots, X_n, Y_1, \dots, Y_n)]$$

2. propriété bayésienne

et on obtient les relations de récurrence sur les mesures σ_n et μ_n :

$$\sigma_n(dx) = Z(x, Y_n) \int_{\mathbb{R}^d} \pi(x'; dx) \sigma_{n-1}(dx')$$

avec la loi de X_0 , $\sigma_0(dx)$ ou $\mu_0(dx)$.

$$\mu_n(dx) = C_n Z(x, Y_n) \int_{\mathbb{R}^d} \pi(x'; dx) \mu_{n-1}(dx')$$

où C_n est une constante de normalisation telle que :

$$C_n^{-1} := \int_{\mathbb{R}^d} Z(x, Y_n) \int_{\mathbb{R}^d} \pi(x'; dx) \mu_{n-1}(dx')$$

permettant d'établir des équations de récurrence pour le filtre optimal. Ce qui entraîne la décomposition en deux étapes classiques :

Étape de prédiction :

$$\mu_n^-(dx) = \int_{\mathbb{R}^d} \pi(x'; dx) \mu_{n-1}(dx')$$

Étape de correction :

$$\mu_n(dx) = C_n Z(x, Y_n) \mu_n^-(dx) \text{ avec } Z(x, Y_n) = \exp \left[\langle Q^{-1} h(x), Y_n \rangle - \frac{1}{2} |Q^{-1/2} h(x)|^2 \right]$$

Pour cette solution théorique, le bruit d'observation est gaussien. Si de plus le bruit de mesure est aussi gaussien, il est préférable d'appliquer le filtrage de Kalman pour déterminer le filtre optimal. Seules l'estimation de sa moyenne \hat{X}_n et l'estimation de sa matrice de covariance R_n sont nécessaires à la détermination de cette loi.

2. Filtre de Kalman étendu

Considérons le système non linéaire gaussien [7] tel que :

$$\begin{cases} Y_n & = h_n(X_n) + V_n & (1) \\ X_{n+1} & = f_n(X_n) + g_n(X_n)W_n & (2) \end{cases}$$

V_n représente le bruit de mesure, bruit blanc gaussien de matrice de covariance Q_n^V .

L'état X_n du système dynamique (2) n'est pas observé et son équation d'évolution est perturbée par un bruit blanc gaussien de matrice de covariance Q_n^W et, de condition initiale X_0 gaussienne $\mathcal{N}(\bar{X}_0, Q_0)$.

Les bruits $\{W_n\}$, $\{V_n\}$ et la condition initiale X_0 sont mutuellement indépendants.

Le critère d'optimalité est de minimiser la variance de l'erreur d'estimation donc de déterminer la loi conditionnelle de X_n sachant $\sigma(\mathcal{Y}_n)$.

Le filtre de Kalman [5] est adapté pour les filtres linéaires gaussiens. Dans le cas non linéaire, dans l'extension du filtre de Kalman, on linéarise par une approximation fonctionnelle le système.

Soit une suite déterministe \bar{x}_n dans \mathbb{R}^d , solution approchée du système, appelée **trajectoire nominale**. Les fonctions f_n et g_n sont supposées dérivables. On linéarise f_n et g_n autour de \bar{x}_n autour des développements habituels et on approxime

$$\begin{aligned} f_n(x) &\simeq f_n(\bar{x}_n) + \vec{\nabla} f_n(\bar{x}_n)(x - \bar{x}_n), \\ g_n(x) &\simeq g_n(\bar{x}_n) \end{aligned}$$

Il en est de même pour h_n autour de \bar{x}_n , $h_n(x) \simeq h_n(\bar{x}_n) + \vec{\nabla} h_n(\bar{x}_n)(x - \bar{x}_n)$.

On obtient le système linéarisé suivant :

$$\begin{cases} Y_n &= H_n(X_n - \bar{x}_n) + \tilde{h}_n + V_n \\ X_{n+1} &= F_n(X_n - \bar{x}_n) + \tilde{f}_n + G_n W_n \end{cases}$$

où $f_n(\bar{x}_n) = \tilde{f}_n$, $F_n = \vec{\nabla} f_n(\bar{x}_n)$, $G_n = g_n(\bar{x}_n)$, $H_n = \vec{\nabla} h_n(\bar{x}_n)$ et $h_n(\bar{x}_n) = \tilde{h}_n$.

En notant, $\tilde{X}_n = X_n - \bar{x}_n$, $\tilde{f}_n = f_n - \bar{x}_{n+1}$ et $\tilde{W}_n = G_n W_n$, on a un nouveau système linéaire :

$$\begin{cases} Y_n &= H_n \tilde{X}_n + \tilde{h}_n + V_n & (3) \\ \tilde{X}_{n+1} &= F_n \tilde{X}_n + \tilde{f}_n + \tilde{W}_n & (4) \end{cases}$$

Les bruits $\{\tilde{W}_n\}$, $\{V_n\}$ et la condition initiale $\tilde{X}_0 = X_0 - \bar{x}_0$ sont mutuellement indépendants. La condition initiale est gaussienne $\mathcal{N}(\tilde{X}_0 - \bar{x}_0, Q_0)$.

\tilde{W}_n , la perturbation aléatoire est un bruit blanc gaussien de matrice de covariance $Q_n^{\tilde{W}}$.

Dans le cas gaussien, seules la moyenne $\hat{X}_n = \mathbb{E}[\tilde{X}_n | \mathcal{Y}_n]$ et la matrice de covariance $R_n = \mathbb{E}[(\tilde{X}_n - \hat{X}_n)(\tilde{X}_n - \hat{X}_n)^t]$ sont nécessaires à la définition de cette loi.

Pour la prévision,

$$\begin{cases} \hat{X}_{n-} &= \mathbb{E}[\tilde{X}_n | \mathcal{Y}_{n-1}] \\ R_{n-} &= \mathbb{E}[(\tilde{X}_n - \hat{X}_{n-})(\tilde{X}_n - \hat{X}_{n-})^t] \end{cases}$$

On construit le processus d'innovation défini par l'information apportée par Y_n par rapport aux observations passées \mathcal{Y}_{n-1} : $I_n = Y_n - \hat{Y}_{n-}$ où $\hat{Y}_{n-} = \mathbb{E}[Y_n | \mathcal{Y}_{n-1}]$.

L'innovation est donc : $I_n = Y_n - (H_n \hat{Y}_{n-} + \tilde{h}_n)$.

Le processus I_n est un processus gaussien à valeurs dans \mathbb{R}^k , en particulier I_n est un vecteur aléatoire gaussien de moyenne nulle et de matrice de covariance $Q_n^I = H_n R_{n-} H_n^t + Q_n^V$ indépendant de \mathcal{Y}_{n-1} . Et on a le théorème :

Théorème de Kalman-Bucy[9]

Si la matrice de covariance Q_n^V est inversible pour tout $n \in \mathbb{N}$, alors les processus \hat{X} et R_n sont définies par les équations suivantes :

Prédiction

$$\begin{cases} \hat{X}_n^- &= F_n \hat{X}_{n-1} + \tilde{f}_n \\ R_n^- &= F_n R_{n-1} F_n^t + Q_n^{\tilde{W}} \end{cases}$$

Corréction

$$\begin{cases} \hat{X}_n &= \hat{X}_n^- + K_n [Y_n - (H_n \hat{X}_n^- + \tilde{h}_n)] \\ R_n &= [I - K_n H_n] R_n^- \end{cases}$$

où le **gain de Kalman** est la matrice $K_n = R_n^- H_n^t [H_n R_n^- H_n^t + Q_n^V]^{-1} = R_n^- H_n^t [Q_n^I]^{-1}$ avec les initialisations :

$$\hat{X}_0^- = \bar{X}_0 - \bar{x}_0 = \mathbb{E}[X_0] - \bar{x}_0, \quad R_0^- = Q_0^X$$

Le gain matriciel de Kalman K_n est à déterminer de façon à fournir une estimation \hat{X}_n optimale au sens de l'erreur quadratique moyenne. On définit ainsi l'erreur à posteriori : $e_n = X_n - \hat{X}_n$ et la matrice de covariance $p_n = \mathbb{E}[e_n e_n^t]$. On peut alors définir l'erreur à priori $e_n^- = X_n - \hat{X}_n^-$, ainsi que la matrice de covariance de l'erreur à priori : $p_n^- = \mathbb{E}[e_n^- (e_n^-)^t]$.

La solution théorique et la solution par approximation par filtre de Kalman dépendent de la condition initiale. On peut montrer que, tout en restant dans le cas gaussien, qu'avec une condition initiale erronée, on peut avoir le même résultat qu'avec le véritable filtre. Il y a donc une stabilité du filtre par rapport à la mesure initiale.

3. Oubli de la condition initiale

On suppose donc que l'on utilise pour le filtre une densité $\bar{p}_0(x)$ alors que la véritable loi initiale est $p_0(x)$. Avec $\bar{p}_0(x)$, on met en œuvre un filtre $\bar{p}_n^-(x)$, la prédiction, et $\bar{p}_n(x)$, la corréction, alors que le vrai filtre est $p_n^-(x)$, la prédiction, et $p_n(x)$ la corréction. Le filtre oublie sa condition initiale [4] si :

$$\lim_{n \rightarrow \infty} \text{ps} \left| \int_{\mathbb{R}^d} f(x) p_n(x) dx - \int_{\mathbb{R}^d} f(x) \bar{p}_n(x) dx \right|$$

est égale à 0, $\forall f \in \mathcal{C}_b(\mathbb{R}^d)$ ou si :

$$\epsilon_n := \|p_n - \bar{p}_n\|_1 = \int_{\mathbb{R}^d} |p_n(x) - \bar{p}_n(x)| dx$$

avec $\lim_{n \rightarrow \infty} \epsilon_n = 0$

Appliquons le cas où $d = 1$. Le cas général se déduit assez facilement. Avec le système de filtrage :

$$\begin{cases} Y_n = g(X_n) + V_n & , W \sim \mathcal{N}(0, \sigma_W) \\ X_{n+1} = h(X_n) + W_n & , V \sim \mathcal{N}(0, \sigma_V). \end{cases} ,$$

on a les lois prédites et corrigées admettant les densités : $p_n^-(x)$ et $p_n(x)$ avec $\mu_n^-(dx) = p_n^-(x) dx$ et $\mu_n(dx) = p_n(x) dx$.

Et comme $\pi(x'; dx) = f_W(x - g(x')) dx$

$$\begin{cases} p_n^-(x) &= \int_{\mathbb{R}} f_W(x - g(x')) p_{n-1}(x') dx' \\ p_n(x) &= C_n Z(x, Y_n) p_n^-(x) \end{cases}$$

avec

$$C_n^{-1} = \int_{\mathbb{R}} \int_{\mathbb{R}} Z(y + g(x'), Y_n) f_W(y) p_{n-1}(x') dy dx'$$

La constante de normalisation est C_n .

$$\text{avec } Z(x, Y_n) = \exp \left[\frac{h(x)Y_n}{\sigma_V^2} - \frac{h(x)^2}{2\sigma_V^2} \right]$$

Analytiquement, $|p_n(x) - \bar{p}_n(x)| = |C_n Z(x, Y_n) p_n^-(x) - \bar{C}_n Z(x, Y_n) \bar{p}_n^-(x)|$.

La majoration de cette quantité n'est pas évidente avec les constantes de normalisation. Il est nécessaire de faire le calcul numériquement à partir d'une méthode de Monte Carlo[1] consistant à approximer numériquement les intégrales sous la forme

$$\int_{\mathbb{R}} \phi(x) p(x) dx \simeq \frac{1}{N} \sum_{i=1}^N \phi(X_i)$$

où X_i sont simulées suivant la loi de densité $p(x)$.

Cette approximation numérique est justifiée par la loi des grands nombres et le théorème central limite. Avec le principe d'échantillonnage d'importance, il est parfois utile de modifier l'intégrande :

$$\int_{\mathbb{R}} \phi(x) p(x) dx = \int_{\mathbb{R}} \phi(x) \frac{p(x)}{q(x)} q(x) dx$$

avec le support de $p(x)$ est contenu dans celui de $q(x)$. Les X_i seront simulées suivant la densité $q(x)$. Le choix de $q(x)$ est uniquement pour une raison pratique non pas pour une réduction de variance d'une meilleure convergence .

Avec cette approximation de Monte Carlo, avec N assez grand,

$$p_n^-(x) \simeq \frac{1}{N} \sum_{i=1}^N f_W(x - g(X_{i,n-1}))$$

$$Z(x, Y_n) = \exp \left[\frac{h(x)Y_n}{\sigma_V^2} - \frac{h(x)^2}{2\sigma_V^2} \right]$$

$$p_n(x) = C_n Z(x, Y_n) p_n^-(x)$$

$$C_n^{-1} \simeq \frac{1}{N} \sum_{i=1}^N Z(W_i + g(X_{i,n-1}), Y_n)$$

C_n est la constante de normalisation, W_i sont simulées suivant la loi de densité f_W et $X_{i,n-1}$ sont simulées suivant la loi de densité p_{n-1} .

Par approximation, on a donc :

$$\begin{aligned} |p_n(x) - \bar{p}_n(x)| &= |C_n Z(x, Y_n) p_n^-(x) - \bar{C}_n Z(x, Y_n) \bar{p}_n^-(x)| \\ &\simeq Z(x, Y_n) \left| \frac{C_n}{N} \sum_{i=1}^N f_W(x - g(X_{i,n-1})) - \frac{\bar{C}_n}{N} \sum_{i=1}^N f_W(x - g(\bar{X}_{i,n-1})) \right| \end{aligned}$$

et

$$\varepsilon_n \simeq \int_{\mathbb{R}} Z(x, Y_n) \left| \frac{C_n}{N} \sum_{i=1}^N f_W(x - g(X_{i,n-1})) - \frac{\bar{C}_n}{N} \sum_{i=1}^N f_W(x - g(\bar{X}_{i,n-1})) \right| dx$$

Avec l'approximation de Monte Carlo :

$$\varepsilon_n \simeq \sum_{j=1}^{N_2} Z(\hat{X}_j, Y_n) \left| \frac{C_n}{N} \sum_{i=1}^N f_W(\hat{X}_j - g(X_{i,n-1})) - \frac{\bar{C}_n}{N} \sum_{i=1}^N f_W(\hat{X}_j - g(\bar{X}_{i,n-1})) \right| \frac{1}{g_0(\hat{X}_j)}$$

$\hat{X}_j \sim \mathcal{N}(0, 1)$ et $g_0(x)$ la densité de de la loi normale centrée réduite.

Dans les simulations, les densités sont définies par des points espacés régulièrement. On définit aussi leurs fonctions de répartition par des points espacés régulièrement en utilisant la proposition suivante :

Si F est une fonction de répartition et $U \sim \mathcal{U}[0, 1]$, alors $X = F^{-1}(U)$ admet F comme fonction de répartition.

Simuler une valeur d'une variable aléatoire de fonction de répartition F , revient à simuler une valeur de la variable aléatoire U suivant la loi uniforme dans $[0, 1]$ et utiliser $X = \arg \max\{F^{-1}(U)\}$. On a les applications suivantes :

4. Applications

L'algorithme suivant donne les calculs des filtres prédit et corrigé.

Algo filtre h, g : fonctions; N, N_1 : entiers ;

f_v, f_W : densités; Y_n observations)

Résultats : $q(x)$ le filtre prédit et $p(x)$ le filtre corrigé

Debut

initialisation de $q(x)$ à $p_0(x)$

initialisation de $p(x)$ à $p_0(x)$

pour $n = 1$ à N faire

$$Z(x, Y_n) = \exp \left[\frac{h(x)Y_n}{\sigma_V^2} - \frac{h(x)^2}{2\sigma_V^2} \right]$$

pour $i = 1$ à N_1 faire

simuler W_i suivant p_W

simuler $X_{i,n-1}$ suivant p_{n-1}

$$C_n^{-1} = \sum_{i=1}^{N_1} \frac{1}{N_1} Z(W_i + g(X_{i,n-1}), Y_n)$$

$$p_n^-(x) = \sum_{i=1}^{N_1} \frac{1}{N_1} f_W(x - g(X_{i,n-1}))$$

fin pour i

$$q_n(x) = p_n^-(x) \text{ et } p_n(x) = C_n Z(x, Y_n) q_n(x)$$

fin pour n

$$q(x) = p_n^-(x) \text{ et } p(x) = p_n(x)$$

Fin.

Prenons pour simplifier $d = 1$. En considérant $g(x) = x$ et $h(x) = x^2$, on a le système :

$$\begin{cases} Y_n & = & X_n^2 + V_n \\ X_{n+1} & = & X_n + W_n \end{cases}$$

Sur Scilab, prenons $\sigma_0 = 1$, $\sigma_W = 2$, $\sigma_V = 1$, $N = 100$, $N_1 = 50$. C'est à dire que le bruit de mesure suit $\mathcal{N}(0, 2)$ et le bruit d'observation suit $\mathcal{N}(0, 1)$.

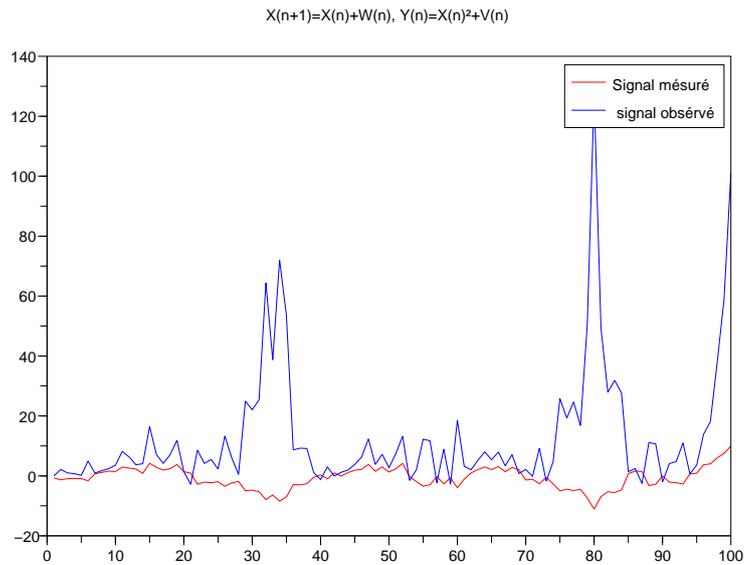
$$Z(x, Y_n) = \exp \left[xY_n - \frac{x^2}{2} \right]$$

$$p_n^-(x) = \int_{\mathbb{R}} f_W(x - x') p_{n-1}(x') dx' \text{ est approximée par } \frac{1}{N_1} \sum_{i=1}^{N_1} f_W(x - \tilde{X}_{n-1,i})$$

$$C_n^{-1} = \int_{\mathbb{R}} \int_{\mathbb{R}} Z(w + x, Y_n) f_W(w) dw \cdot p_n^-(x) dx \text{ approximée par : } \frac{1}{N_1} \sum_{i=1}^{N_1} Z(W_i + \tilde{X}_{n-1,i}, Y_n)$$

où $\tilde{X}_{n-1,i}$ suit la loi p_{n-1} et W_i suit la loi f_W .

Le filtre corrigé est $p_n(x) = C_n Z(x, Y_n) p_n^-(x)$.

Fig 1 : Trajectoires $N = 100$

La trajectoire du signal observé est proportionnelle à l'écart entre le signal mesuré et de la trajectoire de l'estimé, Figure D1.

La trajectoire du filtre prédit est pratiquement confondue à la trajectoire du filtre corrigé. Cette remarque est totalement différente pour un autre modèle tel que : Figure D2

$$\begin{cases} Y_n &= 50 \cdot \sin(X_n) + 50 \cdot \cos(X_n) + V_n \\ X_{n+1} &= X_n + W_n \end{cases}$$

Les trajectoires du filtre prédit et filtre corrigé ne diffèrent que de peu. Mais on ne peut pas déduire du signal observé le signal mesuré. De plus, la trajectoire du filtre corrigé est écartée de la trajectoire du signal mesuré. Le filtre de Kalman représente mieux le filtre.

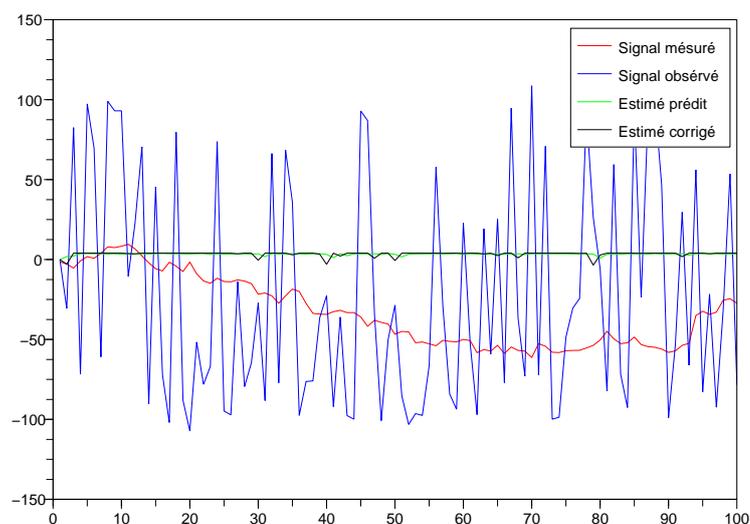


Fig 2 : Filtre sinusoïdal

L'algorithme suivant donne les calculs du filtrage de Kalman.

Algo Kalman : (Q_n^V et Q_n^W : matrices de covariance des bruits, f_n, h_n)

filtre prédit : \hat{X}_n^- , filtre corrigé : \hat{X}_n

Debut

initialisation $\bar{X}_0^- = \bar{X}_0 - \bar{x}_0$ et $R_0^- = Q_0^-$

pour $n = 1 : N$

$$\hat{X}_n^- = F_n \hat{X}_{n-1} + \tilde{f}_n$$

$$R_n^- = F_n R_{n-1} F_n^t + Q_n^W$$

$$Q_n^I = H_n R_n^- H_n^t + Q_n^V$$

$$K_n = R_n^- H_n^t Q_n^I$$

$$\hat{X}_n = \hat{X}_n^- + K_n [Y_n - (H_n \hat{X}_n^- + \tilde{h}_n)]$$

$$R_n = [I - K_n H_n] R_n^-$$

fin pour n

Fin

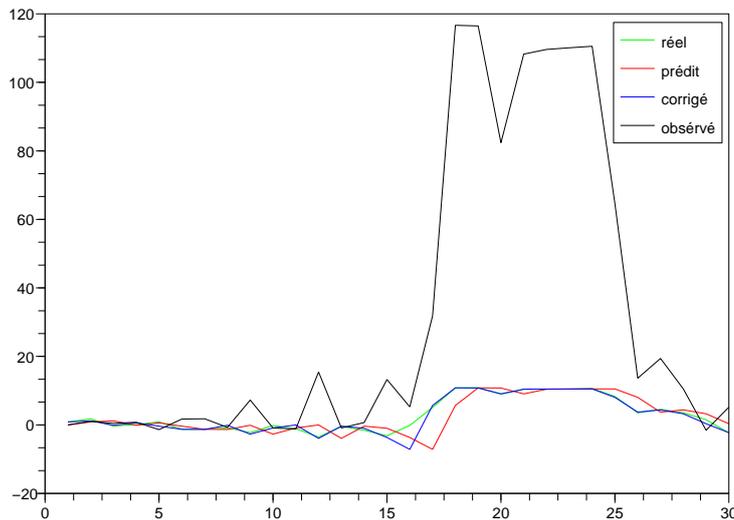


Fig 3 : Trajectoires par filtrage de Kalman

En pratique, l'algorithme du filtre de Kalman est plus adapté pour les filtres optimaux gaussiens.

Le développement de h_n se fait autour de \hat{X}_n^- et celui de f_n autour de \hat{X}_{n-1} .

En revenant sur la solution théorique, on a effectivement des densités gaussiennes. On est dans le cas où tout est gaussien. Le filtre prédit reste centré en 0. Par contre, avec l'innovation apportée par Y_n , la moyenne du filtre corrigé est variable, cf figure D4.

En considérant deux lois initiales de différentes variances telles que :

$$p_0(x) = \frac{1}{\sigma_0\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma_0^2}\right) \text{ et}$$

$$\bar{p}_0(x) = \frac{1}{\bar{\sigma}_0\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\bar{\sigma}_0^2}\right)$$

On a un algorithme pour le calcul de l'erreur :

$$\epsilon_n = \int_{\mathbb{R}^d} |p_n(x) - \bar{p}_n(x)| dx.$$

Algo erreur p_0 et \bar{p}_0 : lois initiales ;

N, N_2 : entiers ; g : densité, ϵ l'erreur

Debut

$$\epsilon(1) = 0$$

pour $n = 1$ à N faire

calculer $p_n(x)$ et $\bar{p}_n(x)$

$$\alpha_n(x) = |p_n(x) - \bar{p}_n(x)|$$

$$\epsilon(n) = \int_{\mathbb{R}} \alpha_n(x) dx$$

fin pour n

Fin

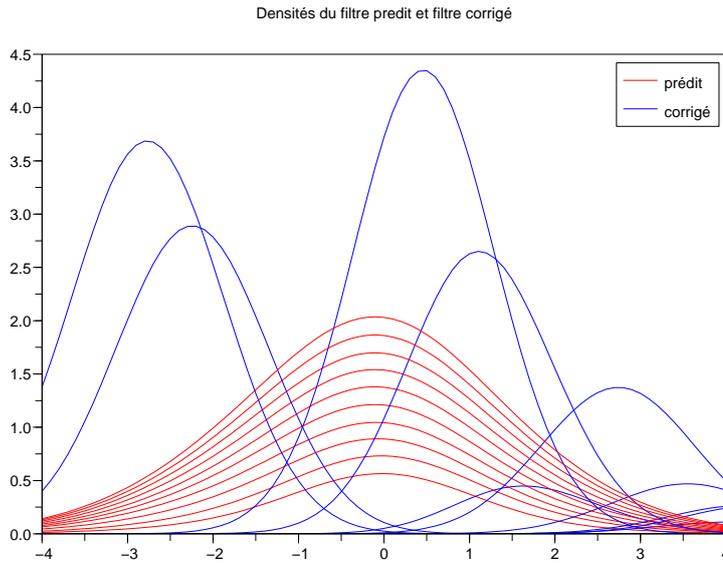


Fig 4 : Filtres prédits et corrigés

L'écart des courbes représentant les filtres optimaux déduits des deux différentes conditions initiales se réduit au cours du temps, cf figure D4. Cet écart est exprimé par une intégrale, cf figure D5, définissant donc l'erreur entre deux conditions initiales. L'intégrale est de plus en plus aplatie. On a bien un oubli de la condition initiale Fig D6.

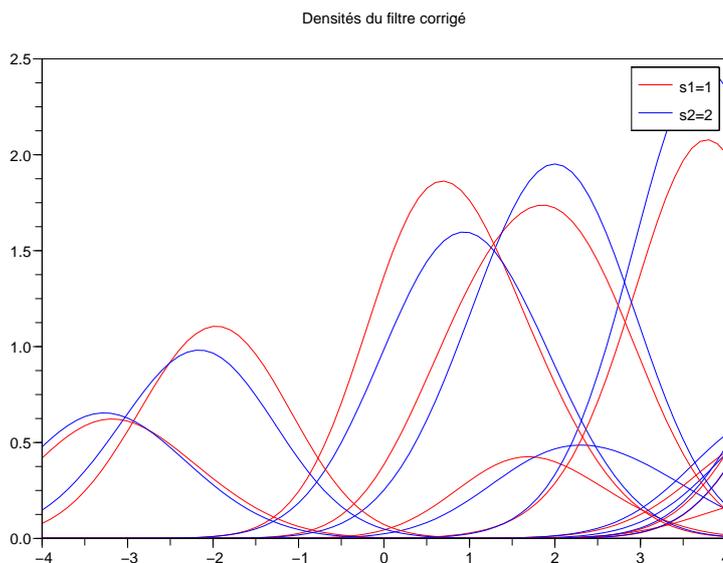


Fig 5 : filtres optimaux

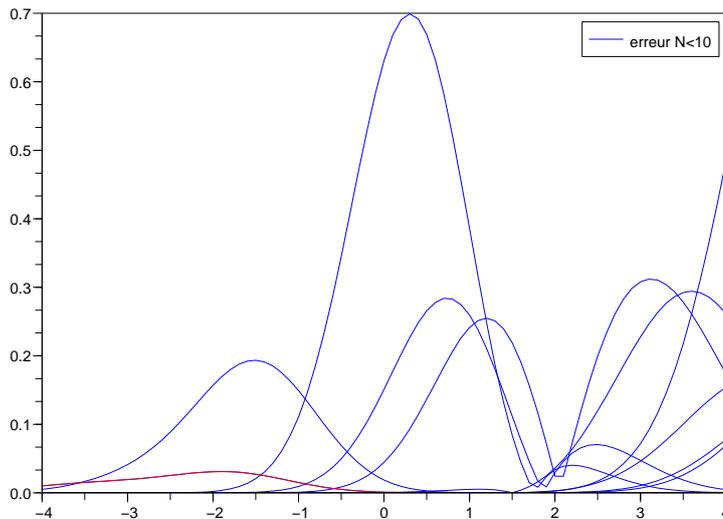


Fig 6 : Intégrales de l'erreur

Dans cette simulation, toutes les intégrales sont pratiquement nuls. L'erreur ϵ_n à chaque pas de temps a une trajectoire suivant l'axe horizontale.

Le filtre de Kalman a pour but d'analyser l'état d'un processus inobservé en général markovien, à partir d'un processus observable non nécessairement linéaire sous conditions de bruits gaussien [7].

Une autre application du filtre de Kalman se trouve en finance à propos des estimations et prévision de la volatilité stochastique et du rapport cours-bénéfice [6].

Conclusion

Cet article développe deux méthodes différentes dans la détermination du filtre optimal dans le filtrage non linéaire, l'une avec la construction d'une mesure de distribution avec une théorie probabiliste et l'autre avec le filtre de Kalman étendu. Un des intérêts du filtre de Kalman est de pouvoir estimer X_0 , la valeur initiale, dans le problème balistique.

Le résultat montre qu'avec une loi initiale erronée, on obtient les mêmes résultats qu'avec le vrai filtre : c'est à dire que le filtre obtenu avec la loi initiale erronée est capable d'oublier sa condition initiale. Ce résultat peut être étendu pour les cas des systèmes à bruit d'observation uniforme.

L'hypothèse de la non-linéarité du modèle est très délicate dans le sens où la multimodalité du filtre peut apparaître et dans ce cas le filtre de Kalman est inadapté surtout lorsque le système est fortement non linéaire, le filtre de Kalman étendu peut diverger. Dans la simulation, le filtre de Kalman [6] est limité à des systèmes à bruits gaussiens. Le cas général peut être envisager en appliquant un autre algorithme comme le filtre particulaire qui est plus efficace mais exige plus d'hypothèses.

Références

[1] Dominique BAKRI. Martingales et Chaînes de Markov. Laboratoire de Statistique et Probabilités. Université Paul Sabatier, Toulouse 2003.

[2] M.JEANBLANC, T.SIMON.Éléments de calculs stochastique.IRBID Sept 2006.

[3] E.PARDOUX. Processus de Markov et applications : Algorithmes, Réseaux, Génomes et Finance. Cours photocopié Université de PROVENCE 2006

[4] E.PARDOUX. Filtrage non linéaire et équations aux dérivées partielles stochastiques associés. In P.L Hennequin, editor, *Ecole d'été de Probabilités de Saint-Flour XIX-1989*, volume 1464 of *Lecture Notes in Mathematics*, pages 67-163. Springer Verlag, Berlin, 1991.

[5] D.ALAZARD, Introduction au filtrage de Kalman, photocopié de cours, SUPAERO, 2005.

[6]F-E. RACICOT & R.THEORET, Quelques applications du filtre de Kalman en Finance : Estimation et Prévision de la volatilité Stochastique et du rapport cours-bénéfices. Université de Montréal, 2005.

[7] F. LAFARGE. Introduction au filtrage de Kalman, théorie du filtre de Kalman discret et applications, cours photocopié ENSG 2009.

[8] Frédéric ROTELLA. Filtrage Optimal. Ecole Nationale d'Ingénieurs de Tarbes 2003.

[9] Michel VALADIER. Sur le filtre de Kalman-Bucy en temps continu. Revue française d'automatique, informatique, recherche opérationnelle. Mathématique, tome 6, n°2, p 9-22, 1972.

D.2 Filtrage Non Linéaire en Temps Discret. Application du filtre particulaire avec un bruit d'observation uniforme

Article publié dans MADA-ENELSA vol1,2013 p.10-17

Résumé

Cet article développe la méthode de Monte Carlo séquentielle ou la méthode particulaire pour le filtrage non linéaire non nécessairement gaussien. Le filtre particulaire nécessite des hypothèses assez complexes dans son élaboration. Néanmoins, avec un bruit uniforme dans l'équation d'observation, on justifie aisément la nécessité de redistribution des particules pour avoir une approximation plus vraisemblable.

Mots-clefs Filtrage non linéaire, Filtre particulaire, Bootstrap, Approximation de Monte Carlo, Chaîne de Markov.

Abstract

This paper develops the method of sequential Monte Carlo particle method or for non-linear filtering is not necessarily Gaussian. The particle filter requires assumptions rather complex in its development. With a uniform noise in the observation equation, it easily justifies the need for redistribution particles to have a more plausible approximation.

Keywords Filtering nonlinear, filter particulate, Bootstrap, Monte Carlo Approximation, Markov Chain.

Introduction

La mesure de l'état d'un système est en général impossible, pour différentes raisons entre autres, l'absence de matériels adaptés aux mesures... On fait alors des mesures observables à partir des capteurs qui sont évidemment liées à l'état du système non observé au cours du temps.

Le signal obtenu est en général perturbé par d'autres signaux d'origines supposés aléatoires.

Le filtrage [1],[2] consiste à obtenir, avec l'information des observations, une estimation de la distribution de la variable d'état non-observable. Cette estimation repose sur deux étapes classiques, la prédiction et la correction. ces étapes permettent de construire un algorithme récursif dans la détermination du filtre en "temps réel".

Les bruits de perturbation sont considérés comme additifs.

Le modèle du problème de filtrage [3] est un système de deux équations : une sur la dynamique de l'état $X = \{X_n; n \geq 0\}$ et une autre sur l'observation $Y = \{Y_n; n \geq 0\}$.

Notons \dagger_n , la tribu des observations Y_1, \dots, Y_n , les filtres prédit $\mu_n^-(dx) = \mathbb{P}_{X_n}(dx|\dagger_{n-1})$ et corrigé $\mu_n(dx) = \mathbb{P}_{X_n}(dx|\dagger_n)$.

Dans cet article, on établit des relations de récurrence des filtres pour un système non linéaire et non nécessairement à bruits gaussiens.

Une des particularités des filtrages à bruits d'observation uniforme est le pouvoir de construire le filtre séquentiellement sans passer par la décomposition classique de prédiction-correction.

Dans le cas plus général, on introduit la méthode de résolution par filtrage particulière, une méthode la plus adaptée aux filtrages non linéaires basée sur la méthode d'approximation de Monte-Carlo.

1. Filtrage Non Linéaire

Considérons le système tel que l'état non observé $X = \{X_n; n \geq 0\}$ est une chaîne de Markov homogène à valeurs dans l'espace mesurable $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est de loi initiale μ_0 et de probabilité de transition π où pour toute fonction mesurable $f : \mathbb{R}^d \rightarrow \mathbb{R}^+$

$$\mathbb{E}[f(X_{n+1})|X_n = y] = \int_{\mathbb{R}^d} f(x)\pi(y; dx) \quad (\text{D.2})$$

$$\mathbb{E}[f(X_0)] = \int_{\mathbb{R}^d} f(x)\mu_0(dx) \quad (\text{D.3})$$

L'observation est un processus tel que :

$$Y = \{Y_n; n \geq 0\}, Y_n = h(X_n) + V_n, n \geq 0. \quad (\text{D.4})$$

où h est une fonction mesurable de \mathbb{R}^d dans \mathbb{R}^k et $V = (V_n; n \geq 0)$ est une suite de variables i.i.d de loi commune $\nu(dx)$.

Supposons que le bruit d'observation suit la loi de densité f_V . On admet les résultats suivants qui établissent des relations de récurrence :

La loi de Y_n sachant \mathcal{Y}_{n-1} est donnée par :

La loi conditionnelle de (X_n, Y_n) sachant \mathcal{Y}_{n-1} est donnée par :

$$\mathbb{P}_{(X_n, Y_n)}(dx, dy|\mathcal{Y}_{n-1}) = f_V(y - h(x))\mu_n^-(dx)dy \quad (\text{D.5})$$

On a la proposition suivante sur la prédiction et la correction d'après la formule de Bayes :

La loi prédite $\mathbb{P}_{X_n}(dx|\mathcal{Y}_{n-1})$ de X_n sachant les observations jusqu'à l'instant $n-1$ est :

$$\mu_n^-(dx) = \int_{\mathbb{R}^d} \pi(x'; dx)\mu_{n-1}(dx') \quad (\text{D.6})$$

La loi corrigée $\mathbb{P}_{X_n}(dx|\mathcal{Y}_n)$ de X_n sachant les observations jusqu'à l'instant n est :

$$\mu_n(dx) = C_n f_V(Y_n - h(x))\mu_n^-(dx) \quad (\text{D.7})$$

C_n est une constante de normalisation

telle que : $C_n = \int_{\mathbb{R}^d} f_V(Y_n - h(x'))\mu_n^-(dx')$

1.1 Cas des bruits à densité

Supposons que le bruit d'observation suit une loi de densité f_V et que la loi de la condition initiale X_0 admet une densité $p_0(x)$. On a :

Les lois conditionnelles prédites et corrigées admettent des densités :

$$\mu_n^-(dx) = p_n^-(x)dx, \quad \mu_n(dx) = p_n(x)dx$$

$$p_n^-(x) = \int_{\mathbb{R}^d} f_W(x - f(x'))p_{n-1}(x')dx' \quad (\text{D.8})$$

$$p_n(x) = \frac{f_V(Y_n - h(x))p_n^-(x)}{\int_{\mathbb{R}^d} f_V(Y_n - h(x'))p_n^-(x')dx'} \quad (\text{D.9})$$

1.2 Cas de bruit d'observation uniforme

On suppose que le bruit d'observation est uniforme sur $\mathcal{D} = [-a; a]^d$, $a > 0$:

$$f_V(x) = \frac{1}{m(\mathcal{D})} \mathbb{I}_{\mathcal{D}}(x)$$

m est la mesure de Lebesgue sur \mathbb{R}^d .

$$p_n^-(x) = \int_{\mathbb{R}^d} f_W(x - f(x'))p_{n-1}(x')dx'$$

$$p_n(x) = C_n \cdot \mathbb{I}_{\mathcal{D}}(Y_n - h(x))p_n^-(x)$$

$$C_n^{-1} = \int_{\mathbb{R}^d} \mathbb{I}_{\mathcal{D}}(Y_n - h(x))p_n^-(x)dx$$

On peut donc avoir $p_n(x)$ en fonction de p_{n-1} sans faire la décomposition filtre prédit et filtre corrigé.

Avec une hypothèse supplémentaire où h est un difféomorphisme, on peut même simplifier cette relation.

Prenons pour simplifier le cas linéaire où $h(x) = x$ et posons : $q_n(x) = p_n(Y_n - x)$. On obtient, en tenant compte que la mesure de Lebesgue est invariante par translation, :

$$q_n(x) = \int_{\mathbb{R}^d} g_1(x')q_{n-1}(x')dx' \mathbb{I}_{\mathcal{D}}(x)$$

où $g_1(x') = C_n f_W(Y_n - x - f(Y_{n-1} - x'))$ et

C_n est une constante de normalisation.

On obtient $p_n(x)$ en fonction de $p_{n-1}(x)$ par l'intermédiaire d'une densité $q_n(x)$ de support \mathcal{D} .

Dans le cas général, le problème de filtrage non linéaire peut se résoudre par le filtrage de Kalman, pour les bruits gaussien, et par le filtrage particulière ou SIR (Sampling with Importance Resampling) pour les bruits non gaussiens.

2. Filtrage particulière

L'idée de la simulation particulière [5],[6] est d'approximer une suite de loi $p_k(x)$ par une suite de loi de $p_k^N(x) = \sum_{i=1}^N \omega_k^i \delta_{\xi_k^i}(x)$, où δ est la mesure de Dirac : c'est la méthode

de Monte Carlo séquentielle.

ξ_k^i sont appelées des particules, et on a $\xi_k^i \sim p_k(x)$ et $p_k(\xi_k^i) = \omega_k^i$ est le poids de la particule ξ_k^i .

Il se peut dans la simulation, les poids ω_k^i sont presque nuls et ne contribuent plus à l'approximation : on dit qu'il y a dégénérescence des particules. Ces particules de poids négligeables devront être remplacées par les particules de poids plus importants non négligeables. On a donc un regroupement de particules dans une région d'état pour ralentir la dégénérescence.

Dans ce cas, on redistribue les particules de poids plus importants pour compléter le nombre total des particules. Cette redistribution ne doit pas se faire sur un petit nombre de particules sinon on perd l'idée de convergence de la méthode.

On peut considérer un rééchantillonnage multinomial : à partir des particules ξ_k^1, \dots, ξ_k^N avec les probabilités $\omega_k^1, \dots, \omega_k^N$, on fait un échantillonnage de N particules de poids $\frac{1}{N}$. Cela revient à multiplier les particules les plus importantes.

Le filtrage particulaire est une méthode numérique qui est une approche adaptée au cas de modèle d'observation non linéaire avec des bruits non-gaussiens.

Considérons une variable aléatoire X à valeurs dans \mathbb{R}^d et de densité p_X , i.e

$$\mathbb{E}\phi(X) = \int_{\mathbb{R}^d} \phi(x)p_X(x)dx \quad \forall \phi \in \mathcal{C}_b(\mathbb{R}^d)$$

De même $p_{X,Y}$, la densité conjointe de X et Y par est telle que $\forall \phi \in \mathcal{C}_b(\mathbb{R}^d \times \mathbb{R}^d)$

$$\mathbb{E}\phi(X, Y) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \phi(x, y)p_{X,Y}(x, y)dx dy$$

On définit la densité conditionnelle de X sachant $Y = y$, notée $p_{X|Y=y}$:

$$p_{X|Y=y}(x) := \frac{p_{X,Y}(x, y)}{p_Y(y)} = \frac{p_{X,Y}(x, y)}{\int_{\mathbb{R}^d} p_{X,Y}(x, y)dx}$$

$$p_{X|Y} := \frac{p_{X,Y}}{p_Y} = \frac{p_{X,Y}}{\int_{\mathbb{R}^d} p_{X,Y}dX}$$

L'espérance conditionnelle associée à cette densité conditionnelle est :

$$\mathbb{E}\phi(X|Y = y) = \int_{\mathbb{R}^d} \phi(x)p_{X|Y=y}(x)dx$$

$$= \frac{\int_{\mathbb{R}^d} \phi(x)p_{X,Y}(x, y)dx}{\int_{\mathbb{R}^d} p_{X,Y}(x, y)dy}$$

$$\mathbb{E}\phi(X|Y) = \int_{\mathbb{R}^d} \phi(x)p_{X|Y}dX$$

$$= \frac{\int_{\mathbb{R}^d} \phi(X)p_{X,Y}dX}{\int_{\mathbb{R}^d} p_{X,Y}dY}$$

Echantillonnage d'importance

Considérons une loi $p(x)$ absolument continue par rapport à une autre loi $\tilde{p}(x)$
 $p(x) = 0 \Rightarrow \tilde{p}(x) = 0$.

$$\mathbb{E}\varphi(X) = \int_{\mathbb{R}^d} \varphi(x) p(x) dx = \int_{\mathbb{R}^d} \varphi(x) \frac{p(x)}{\tilde{p}(x)} \tilde{p}(x) dx$$

Si on simule un N -échantillon $\tilde{\xi}^1, \dots, \tilde{\xi}^N$ à partir de la loi $\tilde{p}(x)$, alors on peut avoir l'approximation par la méthode de Monte Carlo

$$\mathbb{E}\varphi(X) \simeq \sum_{i=1}^N \omega^i \phi(\tilde{\xi}^i)$$

où $\tilde{\xi}^i \sim \tilde{p}(x)$ et $\omega^i = p(\tilde{\xi}^i)/\tilde{p}(\tilde{\xi}^i)$.

Filtre SIR ou bootstrap

La base du filtrage particulière [4] est de déterminer des approximations du filtre prédit et du filtre corrigé sous la forme :

$$\mu_{n-}^{-}(dx) = \mathbb{P}_{X_n}(dx|\mathcal{Y}_{n-1}) \simeq \sum_{i=1}^N \omega_{n-}^i \times \delta_{\xi_{n-}^i}(x) dx$$

$$\mu_n(dx) = \mathbb{P}_{X_n}(dx|\mathcal{Y}_n) \simeq \sum_{i=1}^N \omega_n^i \times \delta_{\xi_n^i}(x) dx$$

À chaque instant k , on détermine des suites $(\omega_{n-}^i, \xi_{n-}^i)$ et (ω_n^i, ξ_n^i) , $i = 1, \dots, N$.

$$\text{Avec le modèle : } \begin{cases} Y_n & = h(X_n) + V_n \\ X_{n+1} & = f(X_n) + W_n \end{cases}$$

$X_0 \sim \mu_0(dx)$, $W_n \sim p_W(w)dw$ et
 $V_n \sim p_V(v)dv$ et notons :

$$p_{n-}(x)dx = \sum_{i=1}^N \omega_{n-}^i \times \delta_{\xi_{n-}^i}(x) dx$$

$$p_n(x)dx = \sum_{i=1}^N \omega_n^i \times \delta_{\xi_n^i}(x) dx$$

On peut considérer le filtre SIS³

3. Sequential Importance Sampling

Filtre SIS

Dans l'étape de **prédiction** avec la formule de Bayes [5],

$$\xi_k^i = f(\xi_{k-1}^i) + W_n^i \text{ avec } W_n^i \sim p_W(w)dw$$

et dans l'étape de **correction** :

$$\omega_n^i = \frac{\omega_{n-1}^i p_V(Y_n - h(\xi_n^i))}{\sum_{j=1}^N \omega_{n-1}^j p_V(Y_n - h(\xi_n^j))}$$

Par contre, le filtre SIR utilise le rééchantillonnage.

Filtre SIR

On commence par un **rééchantillonnage** [4], c'est l'étape de selection, c'est à dire, on choisit $\hat{\xi}_{n-1}^i$ parmi $(\xi_{n-1}^1, \dots, \xi_{n-1}^N)$ en fonction des poids $(\omega_{n-1}^1, \dots, \omega_{n-1}^N)$.

On peut vérifier la valeur du paramètre

$$N_n^{eff} = \frac{1}{\sum_{i=1}^N (\omega_n^i)^2} \in [1, N]$$

qui représente le nombre efficace des particules. Quand N_n^{eff} est proche de N , alors les particules sont d'égale importance.

En général, ce n'est pas le cas et on doit redistribuer les particules pour avoir des poids très proches de $\frac{1}{N}$.

La redistribution se fait selon la loi multinomiale qui élimine les particules de poids négligeables en faveur de celles qui apportent plus de contributions dans l'approximation.

La **prédiction** est une étape de mutation,

$$\xi_n^i = f(\hat{\xi}_{n-1}^i) + W_n^i \text{ avec } W_n^i \sim p_W(w)dw$$

La **correction** est une étape de pondération, telle que

$$\omega_n^i = \frac{p_V(Y_n - h(\xi_n^i))}{\sum_{j=1}^N p_V(Y_n - h(\xi_n^j))}$$

La dégénérescence des poids peut être résolue par une redistribution de Kitagawa, consistant à construire une suite arithmétique de raison $1/N$ de N nombres telle que le premier terme est simulé selon la loi uniforme $\mathcal{U}[0, \frac{1}{N}]$. Cet algorithme a une complexité de l'ordre de N mais n'est pas toujours efficace.

3. Applications

Supposons que

$$\begin{cases} X_{n+1} &= X_n + W_n \\ Y_n &= X_n + V_n \end{cases}$$

$n \geq 0$ et W_n et V_n sont des processus i.i.d de densité f_W et f_V . W_n et V_n sont indépendantes entre elles et indépendantes de la condition initiale X_0 .

Supposons que $d = 1$, le bruit de mesure est gaussien centré de matrice de variance Q et le bruit d'observation est uniforme sur un domaine $\mathcal{D} \in \mathcal{B}(\mathbb{R})$ borné : $\mathcal{D} = [-a, a]$.

$$f_W(w) = g_Q(w), \quad f_V(v) = \frac{1}{m(\mathcal{D})} \mathbf{1}_{\mathcal{D}}(v)$$

où g_Q est la densité gaussienne $\mathcal{N}(0, Q)$ sur \mathbb{R} avec $Q > 0$.

$m(\mathcal{D})$ est la mesure de Lebesgue de \mathcal{D} , on suppose également que $\int_{\mathcal{D}} v \, dv = 0$.

On obtient :

$$\begin{aligned} p_n^-(x) &= \int_{\mathbb{R}} g_Q(x - x') p_{n-1}(x') dx' \\ p_n(x) &= C_n \mathbf{1}_{\mathcal{D}+Y_n}(x) p_n^-(x) \\ C_n^{-1} &= \int_{\mathbb{R}} \mathbf{1}_{\mathcal{D}+Y_n}(x) p_n^-(x') dx' \end{aligned}$$

où $\mathcal{D} + Y_n := \{y \in \mathbb{R}; y - Y_n \in \mathcal{D}\}$

C_n est la constante de normalisation.

La densité $p_n(x)$ est portée par le domaine $\mathcal{D} + Y_n$ pour tout $n \geq 1$ et posons $q_n(x) = p_n(Y_n - x)$.

On a sur \mathcal{D} :

$$q_n(x) = C_n \int_{\mathbb{R}} g_Q(Y_n - Y_{n-1} - x + x') q_{n-1}(v') dx'$$

D'où l'algorithme de calcul du filtre suivant :

Algo filtre $N, N1$: entiers ; g fonction

rendre q_n

Debut

initialisation à q_0 à p_0

pour $n = 1$ à N faire

pour $k = 1$ à $N1$ faire

simuler X_k par q_{n-1}

calculer $g_Q(Y_n - Y_{n-1} - x + X_k)$

fin pour k

q_n

fin pour n

Fin

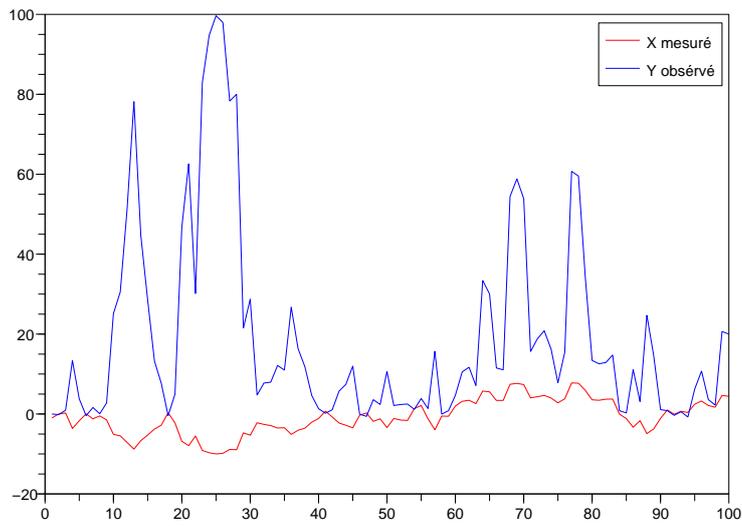


Fig 1 :Signal mesuré et signal observé

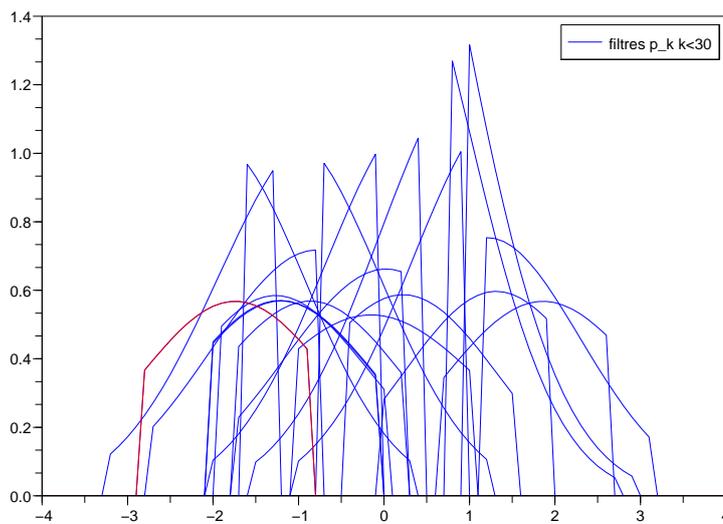


Fig 2 :Filtres $p_n, n \leq 30$

La particularité des densités des filtres est d'avoir des supports compacts bornés [8]

due aux bruits d'observation uniforme. La méthode d'approximation des intégrales de Monte Carlo est nécessaire. La solution théorique est assez compliquée voire difficile à mettre en œuvre dans la simulation quand la dimension est supérieure à 2. Par contre, le filtre particulière est plus adapté dans la généralisation.

Considérons un mobile qui se déplace dans un plan où sa position est le vecteur d'état à l'instant n , $X_n = \begin{bmatrix} X_n^1 \\ X_n^2 \end{bmatrix}$.

On a le système d'équations d'état :

$$\begin{cases} X_{n+1}^1 &= X_n^1 + W_n^1 \\ X_{n+1}^2 &= X_n^2 + W_n^2 \end{cases}$$

W^1 et W^2 sont des bruits blancs gaussiens.

La loi de transition est une loi gaussienne centrée.

On dispose de k -capteurs pour les observations. À chaque instant n , on dispose de la mesure $Y_n^i = h_i(X_n) + V_n^i$ pour chaque capteur i .

h_i^1 mesure la distance euclidienne entre le capteur i et le mobile h_i^2 est l'angle polaire entre l'horizontal et le segment capteur-signal. V_n^i est l'erreur de mesure suivant une loi uniforme.

On a le système d'équations d'observations :

$$\begin{cases} Z_{i,n}^1 &= \sqrt{(X_n^1)^2 + (X_n^2)^2} + V_{i,n}^1 \\ Z_{i,n}^2 &= \arctan\left(\frac{X_n^2}{X_n^1}\right) + V_{i,n}^2 \end{cases}$$

$V_{i,n}^1$ et $V_{i,n}^2$ sont des bruits d'observations uniformes du capteur i .

On reproduit le système d'observation par le système :

$$\begin{cases} Y_{i,n}^1 &= Z_{i,n}^1 \cdot \cos(Z_{i,n}^2) \\ Y_{i,n}^2 &= Z_{i,n}^1 \cdot \sin(Z_{i,n}^2) \end{cases}$$

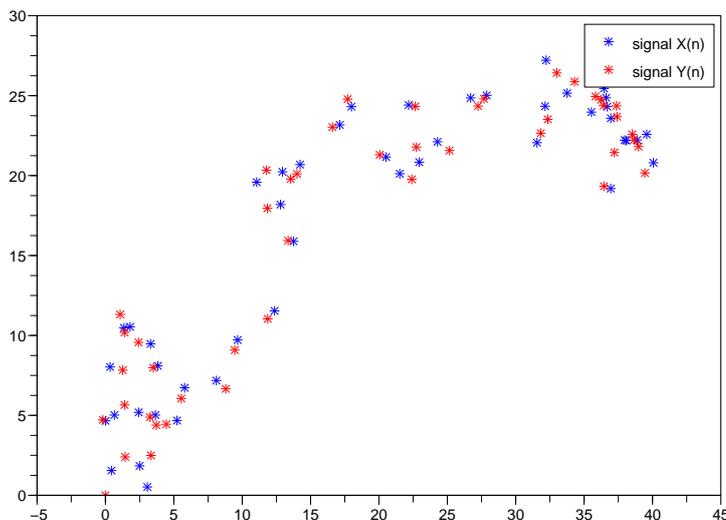


Fig 3 : Signaux mesurés et observés par un capteur

Algo filtre particulaire (p_{X_0} , h et f des fonctions, $\sigma_W > 0$, p_V densité.

Debut

Initialisation des particules : $(\hat{\xi}_1^0, \dots, \hat{\xi}_N^0)$

$$\xi_n^0 \sim p_{X_0} \text{ avec les poids } \omega_n^0 = \frac{1}{N}$$

Pour $i = 1 : N_1$ faire

 Pour $n = 1 : N$ faire

$$\text{mutation : } \xi_n^i = f(\hat{\xi}_n^{i-1}) + W_n^i$$

 pondération :

$$\omega_n^i = \frac{p_V(Y_i - h(\xi_n^i))}{\sum_{j=1}^N p_V(Y_i - h(\xi_n^j))}$$

 fin pour n

$$N^{eff} \leftarrow \left(\sum_{i=1}^N (\omega_n^i)^2 \right)^{-1}$$

 Test si $N^{eff}/N \leq 0.75$ alors,

 Selection : redistribution des particules : $(\hat{\xi}_1^i, \dots, \hat{\xi}_N^i)$ avec $\omega_n^i \leftarrow 1/N$

 Sinon continuer

$$(\hat{\xi}_1^i, \dots, \hat{\xi}_N^i) = (\xi_1^i, \dots, \xi_N^i)$$

fin pour i

Fin

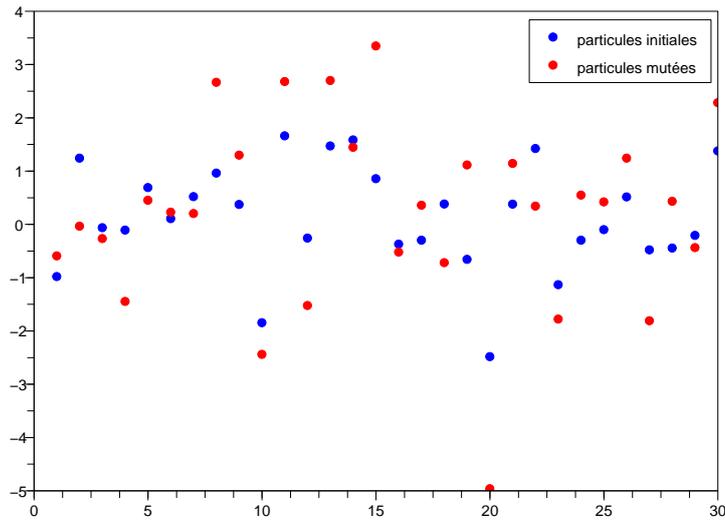


Fig 4 : Processus de mutation

Dans cette application : $p_V(x) = \mathbb{I}_{[-a;a]}(x)$ avec $a > 0$.

La mutation des particules est assez simple à élaborer dans la figure 4. La vraisemblance est une forme de distribution uniforme et la normalisation entraîne une distribution de particules équiprobables. Les particules de poids nuls sont éliminées et remplacées par une redistribution des particules de poids non nuls (Figure 5).

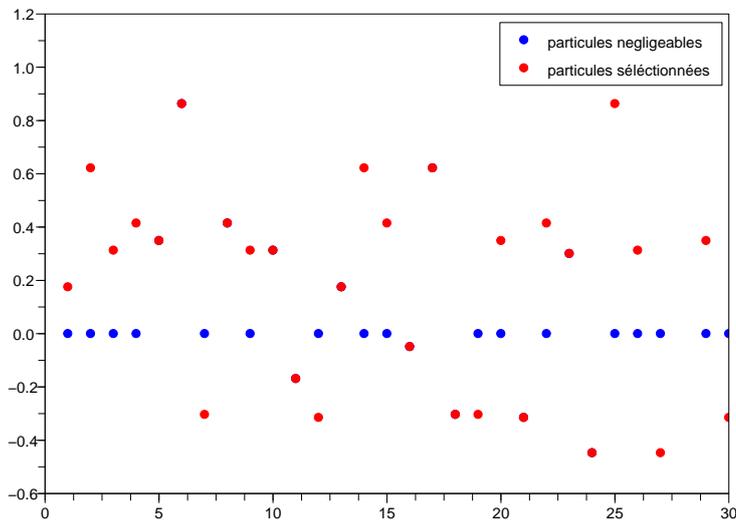


Fig 5 : Processus de selection

Le filtre est représenté par son espérance mathématique à chaque instant. Ces moyennes approchent à chaque instant d'observation la trajectoire réelle.

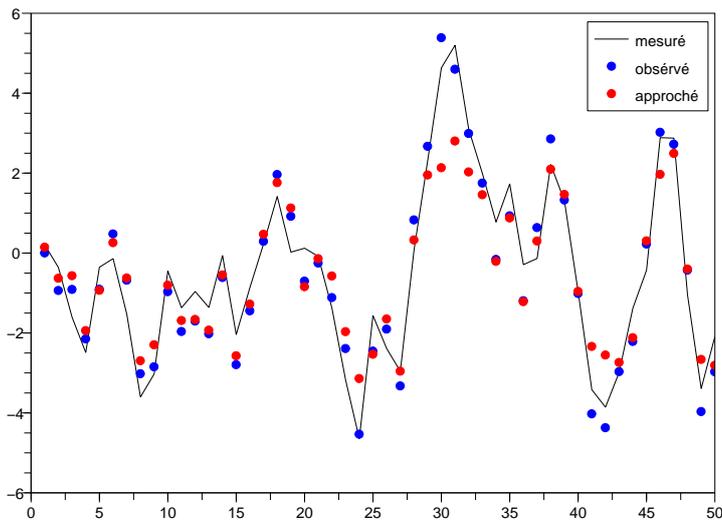


Fig 6 : Filtrage particulaire

Conclusion

Le filtrage particulaire [6] est une des méthodes adaptées à la résolution d'un problème de filtrage non linéaire. Sa mise en œuvre est conditionnée par la connaissance de la loi de distribution initiale, de la loi de transition de la chaîne de Markov, de la loi d'émission et d'un programme de redistribution. Pour un système à bruit d'observation uniforme, les particules sont distribuées de façon équiprobable. Une sélection parmi les particules importantes est nécessaire pour avoir le nombre suffisant pour approximer le filtre. La convergence est justifiée, comme dans toutes les méthodes de Monte Carlo par le théorème central limite et la loi des grands nombres. La complexité des algorithmes est assez raisonnable et la stabilité du filtre par rapport à la mesure initiale peut être étudiée.

Un des inconvénients du filtrage particulaire se trouve dans le rééchantillonnage des particules si le bruit d'observation est trop faible. Dans ce cas, on peut considérer la méthode MCMC pour générer les particules [7].

On étend l'étude pour un système où l'équation d'évolution est une équation différentielle stochastique c'est à dire une équation de la dynamique dans le temps continu.

Références

[1] E.PARDOUX. Mathématiques pour l'ingénieur. Cours photocopié DESS Université de PROVENCE 1991.

[2] F-E. RACICOT & R.THEORET, Quelques applications du filtre de Kalman en Finance : Estimation et Prédiction de la volatilité Stochastique et du rapport cours-bénéfices. Université de Montréal, 2005.

[3] E.PARDOUX. Filtrage non linéaire et équations aux dérivées partielles stochastiques associés. In P.L Hennequin, editor, *Ecole d'été de Probabilités de Saint-Flour XIX-1989*, volume 1464 of *Lecture Notes in Mathematics*, pages 67-163. Springer Verlag,

Berlin, 1991.

[4] Fabien CAMPILLO. Filtrage particulaire et modèles de Markov cachés. Cours polycopié Master 2 Université de Toulon ,2006.

[5] François Le GLAND. Filtrage Bayésien et Approximation Particulaire. Cours Ecole Nationale Supérieure de Techniques Avancées, module Commande des Systèmes, 2011.

[6] Nicolas PAUL. Filtrage Particulaire.Seminaire de Statistique, Conservatoire National des Arts et Métiers, 2006.

[7] Nicolas WIDYNSKI. Filtre particulaire : suivi multi-objets dans les séquences multimodales. Lip6-Télécom Paris Tech, 2008.

[8] M.JEANBLANC, T.SIMON.Éléments de calculs stochastique. IRBID Septembre 2006.

D.3 Estimation d'une commande de signal non observé par filtrage non linéaire par noyau de convolution

Conférence à l'ESPA le 30 Novembre 2014.

Résumé :

Les méthodes classiques de filtrage non linéaires exigent des hypothèses assez lourdes. Cet article développe une méthode d'approximation à la fois du filtre optimal et de la commande régissant l'évolution du signal non observé. La méthode de filtrage non linéaire par noyau de Parzen-Rosenblatt est la mieux adaptée surtout pour de bruits de mesure et d'observation faibles. Elle introduit à la fois la méthode particulaire et la régularisation des filtres obtenus. L'estimation à chaque instant de la commande permet de mieux contrôler la trajectoire du signal avec le système de particules déterminé sur une grille fixe avec des poids proportionnels à leurs apparitions dans l'échantillonnage.

Mots-clefs :

- Filtrage non linéaire, Filtre particulaire, Noyau de Parzen-Rosenblatt, paramètre de lissage, commande optimale.

Abstract :

-The nonlinear approved methods of filtering require heavy enough hypotheses. This article develops a method of approximation at a time of the optimal filter and the order governing the evolution of the non observed signal. The nonlinear filtering method by core of Parzen-Rosenblatt is better especially the adapted for weak noises of measure and observation. She/it introduces the particulate method and the regularization of the filters gotten at a time. The evaluation to every instant of the order permits to control the trajectory of the signal better with the particule system determined on a stationary grid with proportional weights to their apparition in the sampling.

Keywords :

- Nonlinear filtering, particulate Filter, Kernel of Parzen-Rosenblatt, parameter of smoothing, optimal order.

1.Introduction

Un signal (X_n) , $(n \in \mathbb{N})$ est contrôlé par une commande (U_n) , $(n \in \mathbb{N})$ mais inobservable. L'équation d'état de (X_n) , $(n \in \mathbb{N})$ est perturbée par un bruit blanc (W_n) , $(n \in \mathbb{N})$. Le signal observé (Y_n) , $(n \in \mathbb{N}^*)$ est régi par une équation non linéaire et aussi perturbée par un bruit blanc (V_n) , $(n \in \mathbb{N})$.

$$\begin{cases} X_{n+1} &= f_n(X_n, U_n) + W_n \\ Y_n &= h_n(X_n) + V_n \end{cases}$$

Le filtrage non linéaire consiste à estimer le signal $(X_n), (n \in \mathbb{N})$ au vu des observations $(Y_n), (n \in \mathbb{N}^*)$. Il permet aussi d'estimer la commande $(U_n), (n \in \mathbb{N})$ régissant le signal $(X_n), (n \in \mathbb{N})$ selon un critère d'approximation des estimés $\mathbb{E}(X_n|Y_1, \dots, Y_n)$ et des observations en tout instant.

Dans le filtrage, la méthode de Kalman est bien adaptée aux systèmes linéaires gaussiens pour la détermination du filtre optimal. Une extension du filtre de Kalman est applicable pour des systèmes non linéaires présentant des fonctions dérivables mais encore avec des bruits gaussiens. Les méthodes particulières peuvent résoudre les problèmes de système linéaire avec des bruits quelconques non nécessairement gaussiens. Mais ces méthodes particulières sont mises en défaut pour les bruits d'état et d'observation faibles. Le filtrage par noyau de Parzen-Rosenblatt [2] est une solution efficace surtout pour l'estimation de paramètres et en particulier l'estimation des commandes.

Le problème avec le filtrage par noyau est la lenteur de l'algorithme. Le choix d'une grille fixe support des particules pourrait améliorer cette vitesse.

1.1 Filtrage non linéaire

La méthode de filtrage non linéaire avec les phases prédiction-correction n'est adaptée que pour des états markoviens. Ces phases sont incontournables mais rencontrent des situations complexes dans leurs applications. En supposant que la loi conditionnelle prédite admet une densité telle que :

$$p_n^-(x) = \int_{\mathbb{R}^d} f_w(x - f(x')) p_{n-1}(x') dx' \quad (\text{D.10})$$

la loi conditionnelle corrigée est déterminée par la densité :

$$p_n(x) = \frac{f_v(Y_n - h(x)) p_n^-(x)}{\int_{\mathbb{R}^d} f_v(Y_n - h(x')) p_n^-(x') dx'} \quad (\text{D.11})$$

En application du filtrage particulière, l'étape de mutation correspond à la prédiction telle que : $\xi_n^i = f(\hat{\xi}_{n-1}^i) + W_n^i$ et l'étape de pondération correspond à la correction telle que les poids

$$\omega_n^i = \frac{p_V(Y_n - h(\xi_n^i))}{\sum_{j=1}^N p_V(Y_n - h(\xi_n^j))}$$

peuvent être trop faibles et entraîneraient une redistribution des particules en utilisant la loi multinomiale par exemple à l'instant $n = 1$. Les nouvelles particules redistribuées $\hat{\xi}_n^i$ déterminent le filtre optimal à l'instant n . Une régularisation par noyau de Parzen-Rosenblatt permet d'avoir la densité du filtre optimal.

Une régularisation du filtre prédit peut être considérée avant l'étape de correction et un échantillonnage à partir de la densité du filtre corrigé permet de déterminer l'approximation particulière du filtre optimal.

1.2 Modélisation du problème

Étant donné un signal $\{X_t, t \in \mathbb{R}\}$ muni d'un contrôle u_t et perturbé par un bruit blanc $\{W_n, n \in \mathbb{N}\}$. Le signal est régi par l'équation dynamique d'état

$$\frac{dx_t}{dt} = f_t(x_t, u_t, W_t)$$

Le contrôle u_t est déterminé par un critère d'optimisation [11] tel qu'à partir d'un temps initial t_0 :

$$J(t, u(t)) = \mathbb{E} \left[\int_{t_0}^t F(s, x(s), u(s)) ds + G(t, x(t)) \right]$$

Le contrôle optimal est défini par :

$$u^*(t) = \underset{u \in U}{\operatorname{argmin}} J(t, u(t))$$

U est l'ensemble des contrôles admissibles dans la situation.

À partir d'un signal observé $\{y_t, t \in \mathbb{R}^{*+}\}$ perturbé par un autre bruit blanc $\{V_t, t \in \mathbb{R}^{*+}\}$, on établit l'équation d'état $y_t = h_t(x_t, V_t)$. On peut supposer que les bruits sont additifs. Comme les observations sont mesurées en temps discret, on discrétise le système d'équations par :

$$\begin{cases} X_{n+1} &= f_n(X_n, U_n) + W_n \\ Y_n &= h_n(X_n) + V_n \end{cases} \quad (1.1)$$

Dans ce cas, la fonction coût associé à chaque période est :

$$J(u_n) = \mathbb{E} \left[\sum_{k=j}^{n-1} F_k(X_k, u_k) + G_n(X_n) \right]$$

Avec $0 \leq j \leq n - 1$.

Plus précisément, pour maintenir X_n proche de la trajectoire nominale x_n à chaque instant n , la commande optimale est obtenue par exemple :

$$(u_j^*, \dots, u_n^*) = \underset{u_j, \dots, u_n}{\operatorname{argmin}} \sum_{k=j}^n \|x_k - \mathbb{E}(X_k | Y_1, \dots, Y_n, u_1, \dots, u_n)\|^2$$

Avec le système d'équations,

$$\frac{\partial}{\partial u_i} \sum_{k=j}^n \|x_k - \mathbb{E}(X_k | Y_1, \dots, Y_n, u_1, \dots, u_n)\|^2 = 0$$

on peut déterminer les contrôles u_j^*, \dots, u_n^* .

Dans cet article, nous développons la méthode de filtrage particulière par noyau de convolution. Ce filtrage est théoriquement bien adapté pour l'estimation de paramètre ou de contrôle. Dans la mise en œuvre, l'application est très lente vu le nombre considérable de particules pour avoir une assez bonne précision dans les estimations. Nous proposons des particules obtenues par une grille fixe qui sont les éléments principaux dans le filtrage. Les poids des particules varient selon la loi à approcher.

2. Méthodes d'estimation

2.1 Noyau de Parzen-Rosenblatt

La densité d'une variable aléatoire X peut être estimée par la méthode de noyau de Parzen-Rosenblatt telle que :

$$\tilde{f}_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{x - \xi_i}{h_n}\right) \text{ où } K \text{ est un noyau défini dans } \mathbb{R}^d, \text{ borné et intégrable par}$$

rapport à la mesure de Lebesgue et tel que $\lim_{\|n\| \rightarrow \infty} \|x\|^d K(x) = 0$, h_n est le paramètre de

lissage. (ξ_1, \dots, ξ_n) est un échantillon tiré de la variable aléatoire X . Le noyau d'Epanechnikov $K(y) = 1 - \frac{3}{2}y^2$ pour $y < 1$ et le noyau de Picard $K(y) = \exp(-\frac{|y|}{2})$ pour tout y de \mathbb{R} sont parmi d'autres des noyaux de Parzen-Rosenblatt.

2.2 Estimation de la densité conditionnelle

Considérons les estimations par noyau de la densité conjointe de 2 variables aléatoires X et Y où X_i et Y_i sont des particules générées des lois de X et de Y .

$$\hat{p}_{XY}(x, y) = \frac{1}{nh^2} \sum_{i=1}^n K_h^{XY}\left(\frac{x - X_i}{h}, \frac{y - Y_i}{h}\right), \text{ et de la densité marginale de } Y,$$

$$\hat{p}_Y(y) = \frac{1}{nh} \sum_{i=1}^n K_h^Y\left(\frac{y - Y_i}{h}\right). \text{ La loi conditionnelle de } X \text{ sachant } Y = y \text{ est définie,}$$

d'après le théorème de Bayes, par

$$\hat{p}_{X|Y=y}^{(x)} = \frac{\frac{1}{nh^2} \sum_{i=1}^n K_h^{XY}\left(\frac{x - X_i}{h}, \frac{y - Y_i}{h}\right)}{\frac{1}{nh} \sum_{i=1}^n K_h^Y\left(\frac{y - Y_i}{h}\right)}.$$

K_h est un noyau de Parzen-Rosenblatt de paramètre de lissage h en une dimension ou deux.

$K_h(x, y) = (1 - 3/2x^2)(1 - 3/2y^2)$ dans le rectangle $] - 1; 1[\times] - 1; 1[$ pour le noyau d'Epanechnikov et $K_h(x, y) = 1/4 \exp(-(|x| + |y|))$ pour le noyau de Picard dans \mathbb{R}^2 .

2.3 Résolution par filtrage particulière

La méthode de filtrage non linéaire muni d'un paramètre θ [1] avec les phases prédiction-correction n'est adaptée que pour des états markoviens. Ces phases sont incontournables mais rencontrent des situations complexes dans leurs applications.

Supposant que θ admet une loi à priori π_θ et on obtient une estimation θ_0 de θ à l'instant initial.

$$\text{Considérons le système dynamique } \begin{cases} X_{n+1} &= f_n(X_n, \theta_n) + W_n \\ Y_n &= h_n(X_n, \theta_n) + V_n \end{cases} \text{ où les bruits sont sup-}$$

posés additifs.

L'objectif est de déterminer à la fois une estimation de X_n et une estimation θ_n du paramètre inconnu θ à chaque instant n par filtrage particulière à noyau de convolution [8], [2].

Notons $X_n^\theta = (X_n, \theta_n)$ l'état caché augmenté, une nouvelle variable et on obtient un nouveau système :

$$\begin{cases} X_{n+1}^\theta &= f_n(X_n^\theta) + W_n \\ Y_n^\theta &= h_n(X_n^\theta) + V_n \end{cases} (2.2).$$

Théoriquement, en supposant que la loi conditionnelle prédite admet une densité telle que :

$$p_n^{\theta-}(x) = \int_{\mathbb{R}^d} f_w(x - f(x')) p_{n-1}^\theta(x') dx'$$

La loi conditionnelle corrigée est déterminée par la densité :

$$p_n^\theta(x) = \frac{f_v(Y_n - h(x)) p_n^{\theta-}(x)}{\int_{\mathbb{R}^d} f_v(Y_n - h(x')) p_n^{\theta-}(x') dx'} (2.4)$$

La densité du filtre du filtre optimal est obtenu par marginalisation telle que :

$$p_n(x) = \int_{\mathbb{R}^d} p_n^\theta(x) d\theta.$$

En application du filtrage particulaire, on dispose de la densité initiale

$p_0^\theta(x) dx = \mathbb{P}(X_0^\theta \in dx)$, des densités de probabilités $p_{n,n-1}^\theta(x|y) dx = \mathbb{P}(X_n^\theta dx | X_{n-1}^\theta = y)$ de transition correspondant à l'équation d'évolution et des densités d'émission :

$p_{n,Y}^{X,\theta}(y|x) dy = P(Y_n \in dy | X_n^\theta = x)$ correspondant à l'équation d'observation ou la densité de Y sachant $X_n = x$ est : $p_{n,Y}^{X,\theta}(y|x) = f_V(y - h(x))$.

Les particules sont initialisées à partir de $p_0^\theta(x)$. À l'étape $n-1$, $\hat{\xi}_{n-1}^{\theta(i)}$ sont les M -particules tirées de l'estimation de la loi de densité $\hat{p}_{n-1}^\theta(x|y_1, \dots, y_{n-1})$ de l'état augmenté sachant les observations y_1, \dots, y_{n-1} . L'étape de mutation correspond à la prédiction en propageant les particules à partir de l'équation d'évolution : $\xi_n^{\theta(i)-} = f(\hat{\xi}_{n-1}^{\theta(i)}) + W_n^{(i)}$. Avec la loi d'émission, on établit les $p_{n,Y}^{X,\theta}(y|\xi_n^{\theta(i)-})$ pour déterminer une estimation de la loi conjointe de (X_n^θ, Y_n) sachant les observations y_1, \dots, y_{n-1} par convolution à un noyau de Parzen-Rosenblatt :

$$\hat{p}_{XY,n}^\theta(x, y | y_1, \dots, y_{n-1}) = \frac{1}{M} \sum_{i=1}^M K_h(x - \xi_n^{\theta(i)-}) p_{n,Y}^{X,\theta}(y | \xi_n^{\theta(i)-})$$

L'étape de correction correspond à l'estimation de la densité du filtre optimal :

$$\hat{p}_n^\theta(x | y_1, \dots, y_n) = \frac{\sum_{i=1}^M K_h(x - \xi_n^{\theta(i)-}) p_{n,Y}^{X,\theta}(y_n | \xi_n^{\theta(i)-})}{\sum_{i=1}^M p_{n,Y}^{X,\theta}(y_n | \xi_n^{\theta(i)-})}$$

$\frac{p_{n,Y}^{X,\theta}(y_n | \xi_n^{\theta(i)-})}{\sum_{i=1}^M p_{n,Y}^{X,\theta}(y_n | \xi_n^{\theta(i)-})} = \omega_n^{\theta(i)}$ est le poids de la particule $\xi_n^{\theta(i)}$.

Les particules peuvent être trop faibles et entraîneraient une redistribution en utilisant la loi multinomiale par exemple à l'instant $n = 1$. Les nouvelles particules redistribuées $\hat{\xi}_n^{\theta(i)}$ déterminent le filtre optimal à l'instant n .

À chaque itération, les estimateurs du paramètre et de l'état caché sont déterminés à

$$\text{partir de } \hat{X}_n^\theta = (\hat{\xi}_n, \hat{\theta}_n) \text{ où } \hat{X}_n^\theta = \sum_{i=1}^M \omega_n^{\theta(i)} \hat{\xi}_n^{\theta(i)} = \sum_{i=1}^M \omega_n^{\theta(i)} (\hat{\xi}_n^{(i)}, \hat{\omega}_n^{(i)}),$$

$$\hat{\theta}_n = \sum_{i=1}^M \omega_n^{\theta(i)} \hat{\theta}_n^{(i)} \text{ et } \hat{\xi}_n = \sum_{i=1}^M \omega_n^{\theta(i)} \hat{\xi}_n^{(i)}$$

2.4 Algorithme de filtrage par noyau

On a deux formes d'algorithme de filtrage par noyau : le filtrage par noyau pré-régularisé et le filtrage par noyau post-régularisé [7].

Dans le premier cas, le filtre prédit est régularisé par un noyau. L'étape de correction (2.3) permet d'obtenir le filtre optimal. Une sélection de particules permet de passer à l'étape suivante.

Dans le second cas, après l'étape de mutation des particules, l'étape de pondération permet d'obtenir les particules correspondantes au filtre optimal.

Une régularisation par un noyau détermine la densité du filtre optimal.

L'algorithme suivant donne le filtre optimal par régularisation à partir des particules sélectionnées.

Algo filtrage par noyau

Données : p_{X_0} densité de la loi initiale, h et f fonctions du modèle, σ_W^2 variance du bruit d'état, p_V densité du bruit d'observation, p_θ densité à priori de θ .

Résultats : particules ξ_n^i du filtre et \hat{p}_n densité du filtre; $\hat{\theta}$ estimation de θ .

DEBUT

Initialisation des particules

$$(\hat{\xi}_1^0, \theta_1^0), \dots, (\hat{\xi}_N^0, \theta_N^0) \text{ où } \hat{\xi}_k^0 \sim p_{X_0} \text{ avec les poids } \omega_n^0 = \frac{1}{N}$$

$$\theta_k^0 \sim p_\theta$$

Génération des trajectoires

Pour $k = 1 : n$, faire

Pour $i = 1 : N$, faire

$$\text{Mutation : } \xi_k^{\theta(i)} = f(\hat{\xi}_k^{i-1}, \theta_k^{i-1}) + W_k^i$$

$$\theta_k^i = \theta_k^{i-1}$$

Régularisation du filtre

$$\hat{p}_k^\theta(x|y_1, \dots, y_k) = \frac{\sum_{j=1}^N K_h \left(x - \xi_k^{\theta(j)-} \right) p_V \left(Y_k - h \left(\xi_n^{\theta(i)-} \right) \right)}{\sum_{j=1}^N p_V \left(Y_k - h \left(\xi_k^{\theta(j)-} \right) \right)}$$

$$\text{Pondération : } \omega_k^{\theta(i)} = \frac{p_V \left(Y_k - h \left(\xi_k^{\theta(i)-} \right) \right)}{\sum_{j=1}^N p_V \left(Y_k - h \left(\xi_k^{\theta(j)-} \right) \right)}$$

Fin pour i

$$N^e f f \leftarrow \left(\sum_{j=1}^N (\omega_k^j) \right)^{-1}$$

Test si $\frac{N^e f f}{N} \leq 0.75$, alors

Sélection : redistribution des particules

$$(\hat{\xi}_1^{\theta(i)}, \dots, \hat{\xi}_N^{\theta(i)}) \text{ de poids } \omega_n^{\theta(i)} = \frac{1}{N}$$

Sinon

$$\hat{X}_k^\theta = \sum_{i=1}^M \omega_k^{\theta(i)} \hat{\xi}_k^{\theta(i)} = \sum_{i=1}^M \omega_k^{\theta(i)} \left(\hat{\xi}_k^{(i)}, \hat{\omega}_k^{(i)} \right)$$

$$\hat{\theta}_k = \sum_{i=1}^M \omega_k^{\theta(i)} \hat{\theta}_k^{(i)}$$

$$\hat{\xi}_k = \sum_{i=1}^M \omega_k^{\theta(i)} \hat{\xi}_k^{(i)}$$

Fin pour k

FIN

Dans le filtrage non linéaire d'un système contrôlé,

$$\begin{cases} X_n &= f(X_{n-1}, U_n) + W_n \\ Y_n &= h(X_n) + V_n \end{cases}$$

le paramètre θ est substitué par une variable de contrôle U_n déterministe en chaque instant.

Plus précisément, dans la poursuite de trajectoire, la variable de contrôle à chaque instant est déterministe définie par :

$$u_n^* = \arg_u \|f(\hat{\xi}_{n-1}, u) - x_n\|^2$$

3. Applications

Considérons le système d'équation d'état et d'équation d'observation :

$$\begin{cases} X_{n+1} &= af(X_n) + W_n \\ Y_n &= h(X_n) + V_n \end{cases}$$

pour $n > 0$ et $X_0 = W_0$ où W_n est un bruit blanc de variance σ^2 inconnue perturbateur de l'équation d'état, a est un inconnu.

À l'instant initial, l'état du système est représenté par le bruit.

V_n est un bruit blanc perturbateur de l'équation d'observation de variance supposée connue.

En prenant $a = 0$. avec les écart-types égaux à 0.8, par le système dynamique (1), on représente les états et les observations par :

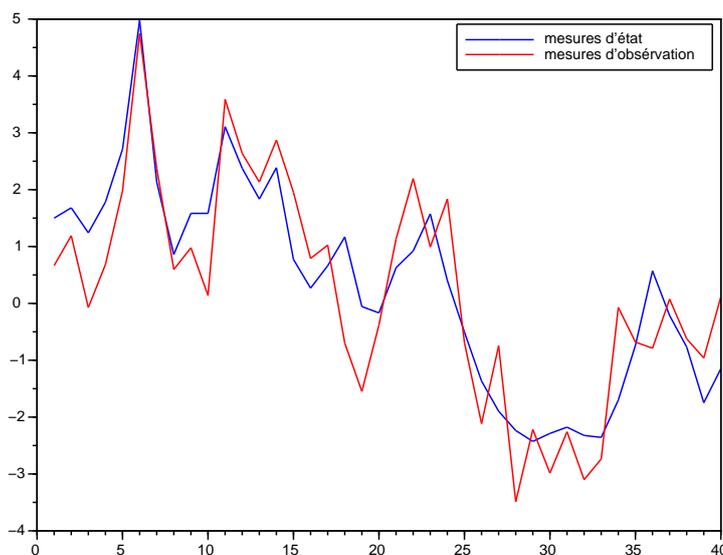


Fig 1 : Trajectoires de la variable d'état et de la variable d'observation.

On veut estimer les paramètres a et σ^2 et déterminer le filtre par filtrage par noyau.

Pour un échantillon (X_1, \dots, X_n) , la vraisemblance est :

$$L(a, \sigma, X_1, \dots, X_n) = \frac{1}{(\sqrt{2\pi}\sigma^2)^n} \prod_{k=1}^n \exp \left[-\frac{(X_k - aX_{k-1})^2}{2\sigma^2} \right]$$

Et le Log-vraisemblance par :

$$\ln L(a, \sigma, X_1, \dots, X_n) = \frac{n}{2} \ln(2\pi\sigma^2) - \sum_{k=1}^n \frac{(X_k - af(X_{k-1}))^2}{2\sigma^2}$$

Par la méthode du maximum de vraisemblance, on a les estimations suivantes :

$$\frac{\partial \ln L(a, \sigma, X_1, \dots, X_n)}{\partial a} = 0$$

$$\hat{a}_n = \frac{\sum_{k=1}^n f(X_{k-1})X_k}{\sum_{k=1}^n f(X_{k-1})^2}$$

$$\frac{\partial \ln L(a, \sigma, X_1, \dots, X_n)}{\partial \sigma^2} = 0$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \hat{a}_k f(X_{k-1}))^2.$$

Comme le signal X_n n'est pas observable, les estimations des paramètres doivent se faire récursivement en fonction des observations (Y_1, \dots, Y_n) .

Dans l'algorithme, les particules définissant l'estimé de θ sont échantillonnées à partir de la densité p_θ . On est dans le cas informatif.

Dans le cas non informatif, les particules initiales définissant la loi de θ sont tirées d'une loi uniforme. Dans ce cas, à la n -ième itération, les particules ξ_k^{i-1} sont mutées en $\xi_n^i = f(\hat{\xi}_n^{i-1}) + W_n^i$ et $\theta_k^i = \theta_k^{i-1}$ $\zeta_n^i = h(\hat{\xi}_n^{i-1}) + V_n^i$ sont les particules définissant la loi de Y_n .

L'estimation de la densité du filtre est :

$$p_n^M(x|\mathcal{Y}_n) = \frac{\sum_{i=1}^M K_{h_M}^y(y_n - \bar{y}_n^i) \times K_{h_M}^y(x - \bar{\xi}_n^i)}{\sum_{i=1}^M K_{h_M}^y(y_n - \bar{y}_n^i)}$$

L'estimation de la densité conditionnelle de θ

$$p_n^M(t|\mathcal{Y}_n) = \frac{\sum_{i=1}^M K_{h_M}^y(y_n - \bar{y}_n^i) \times K_{h_M}^y(t - \bar{\zeta}_n^i)}{\sum_{i=1}^M K_{h_M}^y(y_n - \bar{y}_n^i)}$$

$K_{h_M}^y, K_{h_M}^x, K_{h_M}^\theta$ sont des noyaux de Parzen-Rosenblatt relatives à Y, X et θ . L'estimé du paramètre θ est tel que :

$$\hat{\theta}_n = \mathbb{E}(\theta|Y_n) = \int_{\mathbb{R}} t.p_n^M(t|\mathcal{Y}_n)dt$$

Applications numériques

(ξ_1, \dots, ξ_N) est un échantillon issu d'une densité de probabilité $f(x)$. Avec une approximation particulière de la densité de probabilité, $p_x = \sum_{i=1}^n \omega_i \delta_{\xi_i}(x)$ où ξ_i sont des particules de poids $\omega_i = 1/N$.

La régularisation de $p(x)$ est obtenue par convolution d'un noyau avec $p(x)$. En considérant les noyaux suivants : le noyau d'Epanechnikov tel que $f(x) = 1 - 3/2x^2$ pour $x \in [-1/2; 1/2]$ et 0 ailleurs et le noyau de Picard tel que $f(x) = \exp(-|x|/2)$, la longueur de la fenêtre est de 0.8 et $N=100$. Pour avoir une meilleure approximation, le nombre des particules doit être assez importante.

Une suite déterministe de points est établie dans l'intervalle $[-4; 4]$ avec un pas de 0.1.

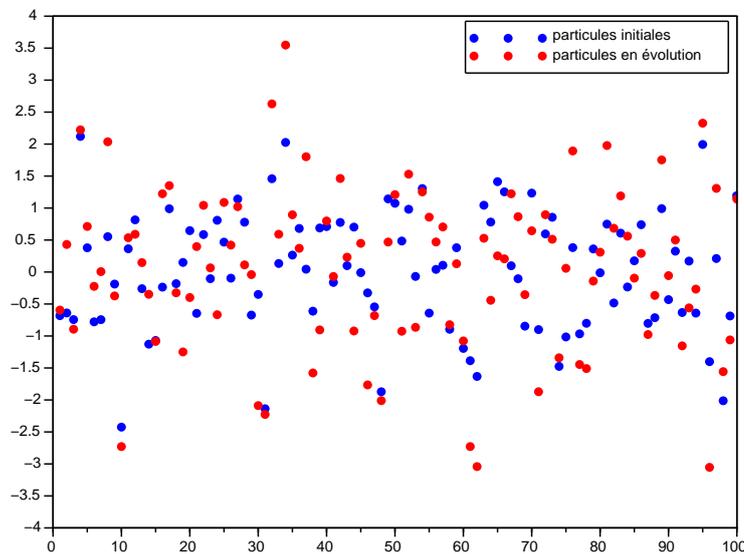


Fig 2 : Particules initiales et particules en évolution

Les particules évoluent selon l'équation d'état.

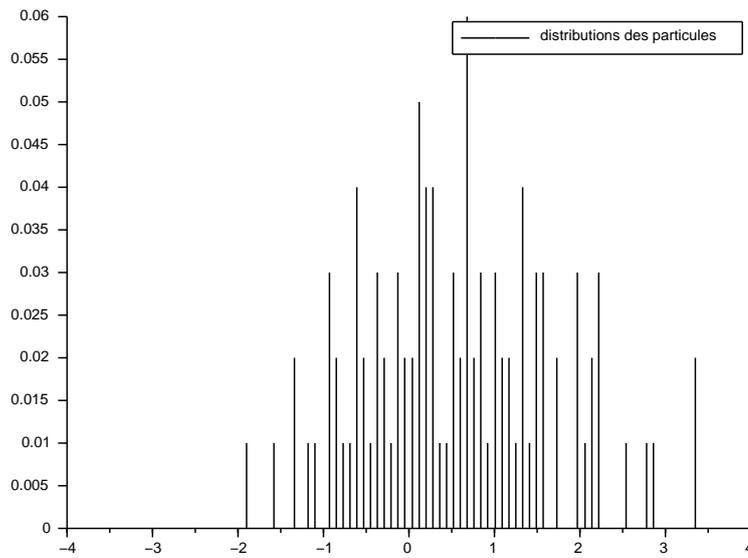


Fig 3 : Distribution des particules

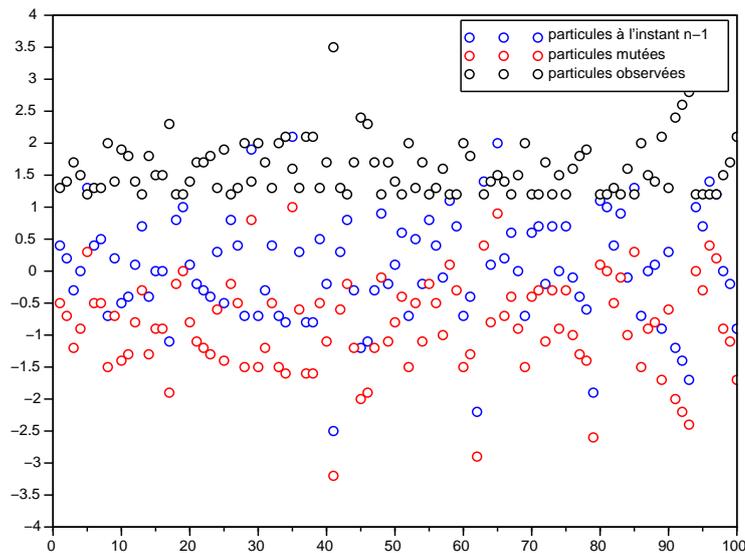


Fig 4 : Particules obtenues par maillage déterministe

Un histogramme de distribution montre les poids des particules suivant la fréquence. Une sélection des particules peut être introduite pour avoir une bonne répartition. Les particules de poids faibles sont éliminées et remplacées par multiplications des particules de poids forts.

Un échantillonnage de la loi à priori de θ permet d'obtenir des particules $(\theta_1, \dots, \theta_N)$. À l'étape suivante, on peut supposer que ces particules n'évoluent pas.

Avec l'équation d'observation, on obtient des particules représentant la loi de la variable d'observation à l'étape. Les estimations du filtre et du paramètre sont déterminées par convolution à un noyau de ces systèmes de particules.

Le choix du noyau [6] n'est pas absolument nécessaire. Avec un test d'approximation de la densité exponentielle avec le noyau d'Epanechnikov et le noyau de Picard, il n'y a pas de remarques particulières.

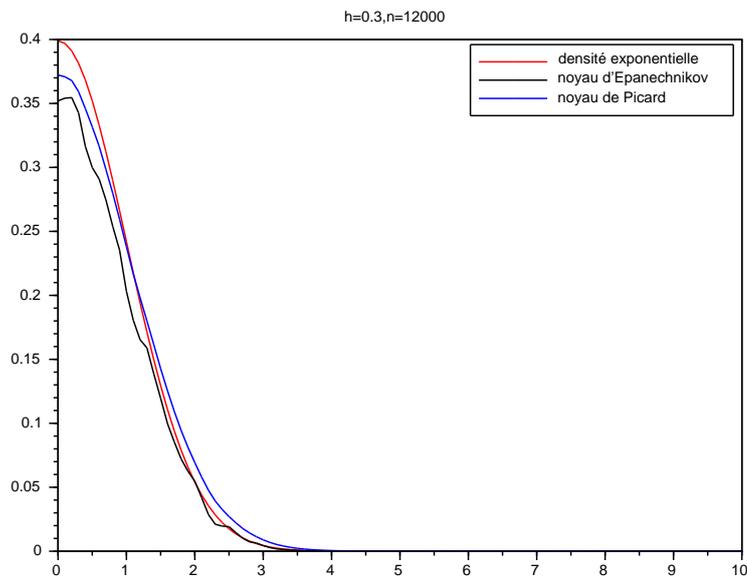


Fig 5 : Estimations par noyaux d'Epanechnikov et de Picard.

Une régularisation par convolution à un noyau détermine une approximation du filtre prédit.

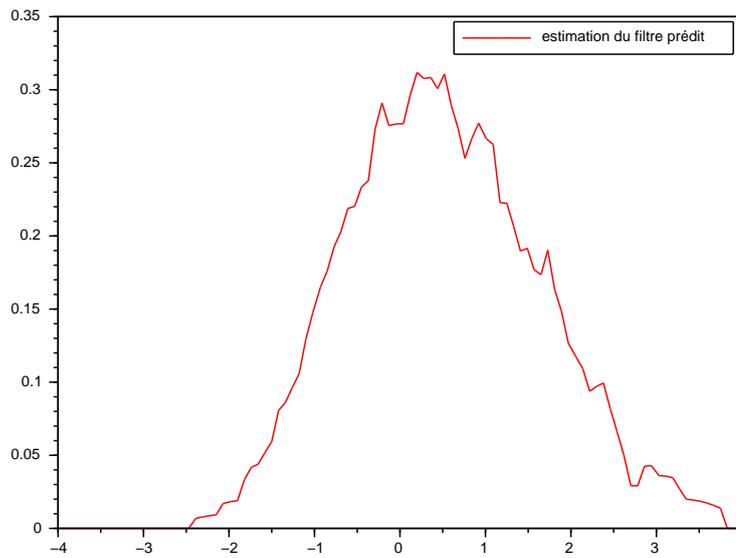


Fig 6 : Estimations de la densité du filtre prédit

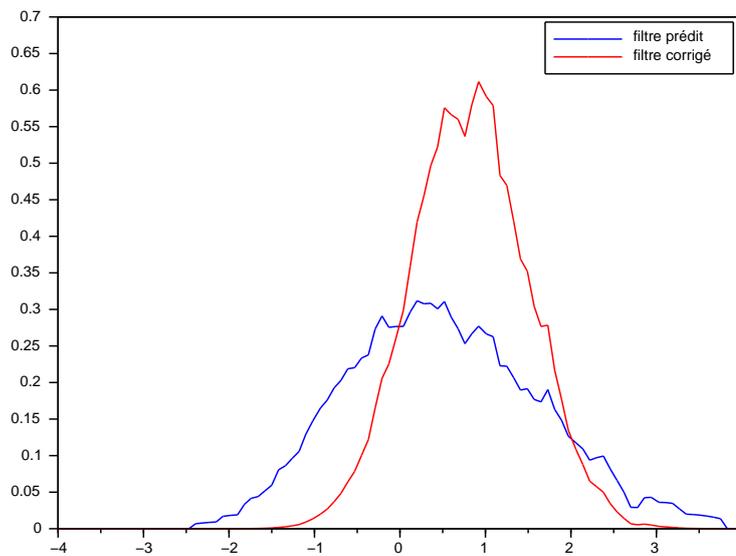


Fig 7 : Filtres prédit et corrigé.

Le filtre prédit est corrigé à partir de l'observation en cet instant. Un échantillonnage de ce filtre optimal définit les particules dans l'itération suivante. Dans la grille fixe, le nombre d'apparition de chaque particule est supposé son poids. L'estimation du paramètre est obtenue par l'espérance de la densité estimée du filtre. Dans le cas de système commandé, le paramètre estimé est la commande u .

$$\begin{cases} X_n &= u \cdot f(X_{n-1}) + W_n \\ Y_n &= h(X_n) + V_n \end{cases}$$

La commande est déterminée à chaque instant par

$$u_n^* = \underset{1 \leq i \leq N}{\operatorname{argmin}} \|f(\hat{\xi}_{n-1}, u_{n-1}^i) - x_n\|^2$$

x_n est la trajectoire nominale connue à l'instant n et u_{n-1}^i sont les particules obtenues par l'estimation de la densité de u à l'instant $n - 1$.

4. Conclusion

Le paramètre d'un signal non observable peut être estimé par un filtrage non linéaire par noyau de convolution. Les particules ont des poids différents selon l'estimation. Elles peuvent être absentes un moment et revivre avec de certains poids à l'itération suivante. L'exploration de l'espace, l'idée naturelle du filtrage particulaire, se fait dans une grille fixe. Ce choix peut accélérer la mise en œuvre du filtrage et on peut appliquer comme dans la poursuite de trajectoire, sans prendre trop de retard.

Le filtrage particulaire par noyau de convolution peut être utilisé dans un lissage pour la construction d'un modèle dynamique non observable. Dans le filtrage particulaire, un bruit quelconque d'observation [9] peut être décomposé en mélange de bruits gaussiens.

Références

[1] Aboubacar AMIRI. Estimateurs fonctionnels récurrents et leurs applications à la prévision. Thèse de doctorat en Mathématiques, spécialité Statistiques. Université d'Avignon et des Pays de Vaucluse, déc 2010.

[2] Yuting CHEN et al. Filtrage par noyaux de convolution itératif. Les 44 journées de statistique, Bruxelles Belgique, 2012.

[3] Smail ADJABI et al. Méthode du noyau dans l'évaluation de performance des systèmes d'attente. 41èmes Journées de Statistique. SFdS, Bordeaux, 2009.

[4] Smail ADJABI, Mouloud CHERFAOUI. Estimation non paramétrique de la fiabilité et du taux de hasard par la méthode du noyau adapté. Laboratoire LAMOS, univ. Béjaia, Algérie, 2012.

[5] Thomas BREHARD, Jean-Pierre Le Cadre. Un problème de filtrage non linéaire avec variance d'état inconnue : le pistage mono-cible par mesure d'angle seul. IRISA, Rennes 2006.

[6] Paul DEHEUVELS. Estimation non paramétrique de la densité par histogrammes généralisés. Revue de statistique appliquée, tome 25, n°3 p.5-42, 1977.

[7] Vivien ROSSI, Jean-Pierre VILA. Approche non paramétrique du filtrage de système non linéaire à temps discret et à paramètres inconnus. ENSAM-INRA, univ. Montpellier II, 2005.

[8]Elie YOUNDJÉ. Contribution à l'estimation non-paramétrique par la méthode du noyau. Mémoire d'Habilitation à Diriger des Recherches, spécialité mathématiques appliquées et statistique. Univ. De Rouen et du Havre, INSA de Rouen, oct 2011.

[9]Serge PROSPERI. Filtrage non linéaire en poursuite. Une approche par mélanges gaussiens. Seizième colloque GRETSI, Grenoble, sept 1997.

[10] François CARON. Méthodes particulières et fusion de données. Journées d'études Météo-SMAI-DSNA, Toulouse, 2010.

[11]François Le Gland. Filtrage Bayésien et Approximation Particulaire. Ecole Nationale de Techniques Avancées, INRIA Rennes, sept 2011.

[12]J.-M BROSSIER, P.-O AUBLAND. Particulaire et communication numériques. LIS/ENSIEG, Saint Martin d'Hères, SEE, juin 2004.

Bibliographie

- [ABDOUS et al,1994] Belkacem ABDOUS, Alain BERLINET. Correction locale de l'estimateur à noyau de la densité d'une loi de probabilité. The Canadian Journal of Statistics, Vol. 22, n°2, P.295-308, 1994.
- [ASJAKA et al.,2009] Smail Adjaka, Karima Lagha, Madjid Hassani. Méthode du noyau dans l'évaluation de performance des systèmes d'attente. 41èmes journées de Statistique SFdS, Bordeaux 2009.
- [ADJABI et al.,2012] Smail ADJABI, Mouloud CHERFAOUI. Estimation non paramétrique de la fiabilité et du taux de hasard par la méthode du noyau adaptée. Laboratoire LAMOS, Université de Béjaia, Algérie, 2012.
- [ANDERSON,MOORE,1979] B. ANDERSON, J.MOORE. Optimal Filtering. Prentice Hall. 1979
- [ALAZARD,2006] D. ALAZARD. Introduction au filtre de Kalman. SUPAERO, oct 2006.
- [AMIRI, 2010] Aboubacar AMIRI. Estimateurs fonctionnels récursif et leurs applications à la prévision. Thèse de doctorat, Université d'Avignon et des Pays de Vaucluse, 2010.
- [ARNAUD et al, ?] E.ARNAUD, E.MEMIN, B.CERNUSCHI-FRIAS. Filtrage Conditionnel pour le suivi de Point dans des Séquences d'Images; IRISA, Univ. de Rennes I, Univ. de Buenos Aires. ?
- [BAEHR,2008] Christophe BAEHR. Modélisation et Filtrage non-linéaire des mesures discrètes sur un fluide turbulent. Application au débruitage des mesures du vent atmosphérique. Atelier de Modélisation de l'Atmosphère Météo-France, Université Paul Sabatier 2008.
- [BAEHR,2009] Christophe BAEHR. Filtrage stochastique non linéaire et approximation particulière pour des modèles météorologiques, LEFE-ASSIM. Météo-France, IMT-LSP Univ. Paul Sabatier, Dec 2009.
- [BAKRI,2003] Dominique BAKRY. Martingales et Chaînes de Markov. Laboratoire de Statistiques et Probabilités, Université Paul Sabatier, Toulouse

France, Juin 2003.

- [BENSOUSSAN,1984] Alain BENSOUSSAN et al. Commande optimale de systèmes stochastiques. RAIRO Automatique/ Systems Analysis and Control, vol 18, n°2, p.225 à 250, INRIA. Voluceau-Rocquefort., 1984.
- [BERNIER et al.,1994] H.BERNIER, F.CAMPILLO, F.CEROU, F.Le GLAND, R.RAKOTOZAFY. Parallélisme de données et filtrage non linéaire, Analyse de performance. Rapport technique INRIA n°0167, nov 1994.
- [BERCU,2008] B. BERCU. Estimation fonctionnelle et contrôle adaptif. Université de Bordeaux, INRIA Bordeaux Sud-Ouest, Equipe CQFD, oct 2008.
- [BEROZZI,2003] Tanya BEROZZI, Didier Le RUYET, Gilles RIGAL et Han VU-THIEN. Application du filtrage particulaire à l'estimation conjointe données-canal. DIGINEXT Aix en Provence et CNAM Paris, 2003.
- [BILAL,2010] Bilal Nehme. Techniques non-additives d'estimation de la densité de probabilité. Thèse de doctorat, Université de Montpellier II, 2010.
- [BOUGEROL,2013] Philippe BOUGEROL. Modèles Stochastiques et Applications à la Finance, Univ. Pierre et Marie Curie, Master de Mathématiques M1, 2013.
- [BOURGOIS et al.,2011] L.BOURGOIS, G.ROUSSEL, M.BENJELLOUN. Kalman d'ensemble état-paramètres appliqué a modèle de Lorentz. ULCO, LISIC, Univ. de Lille Nord de France, 2011.
- [BREHARD, 2006] Thomas BREHARD, Jean-Pierre Le CADRE. Un problème de filtrage non linéaire avec variance d'état inconnue : le pistage mono-cible par mesure d'angle seul. IRISA Rennes, 2006.
- [F.CAMPILLO,2006] Fabien CAMPILLO. Filtrage particulaire et modèles de Markov cachés. Master 2 Université de Toulon ,2006.
- [CAMPILLO,1986] Fabien CAMPILLO. La méthode d'approximation de Gauss-Garlekin en filtrage non linéaire. RAIRO- Modélisation mathématique et analyse numérique, tome 20, n°2, p.203-223, 1986.
- [CAMPILLO,2004] Fabien CAMPILLO. Quelques applications des processus de diffusion : Filtrage-Statistique, Contrôle, Homogénéisation. HDR Univ. de RennesI, 2004.
- [CAMPILLO,2006] Fabien CAMPILLO. Filtrage particulaire : Quelques exemples. INRIA, janv 2006.

- [CARON,2010] François CARON. Méthodes particulières et fusion de données. INRIA Bordeaux. Journées d'étude Météo-SMAI-DSNA, Toulouse, Mars 2010.
- [CAYLUS et al.,?] N. CAYLUS, A.GUYADER, F.Le GLAND et N.OUDJANE. Application du filtrage particulaire à l'inférence statistique de HMM, IRISA Univ. Rennes I et II, INRIA Rennes. ?
- [CHEN et al.,2012] Yuting CHEN, Benoît BAYOL, Cedric LOI, Samis TREVEZAS, Paul Henry COURNEDE. Filtrage par Noyaux de Convolution Itératif. Ecole Centrale de Paris, INRIA Saclay-Île de France. Les 44 journées de STATISTIQUE, Bruxelles 2012.
- [CHONAVEL,2011] Thierry CHONAVEL. Amélioration du filtrage non linéaire dans les modèles d'état par ré-estimation de l'état passé. 23ième édition du colloque GRETSI, Bordeaux, 2011.
- [CLIMESCU, 2004] Adriana CLIMESCU-HAULICA. Filtrage stochastique non linéaire par la théorie de représentation des martingales. Laboratoire de Modelisation et Calcul. Institut d'Informatique et Mathématiques Appliquées de Grenoble 2004.
- [COLLIN et al.,1999] L. COLLIN, S. AZOU, K. YAO, G. BUREL. Base longue de surface et filtrage non-linéaire pour une localisation sous-marine précise. RTCLF, Metz France, oct 1999.
- [COLLIN et al.,2001] L. COLLIN, P. ROSTANG, O. BERDER, G. BUREL. Application de l'estimateur de Julier et Uhlmann à l'évaluation rapide du Taux Erreur Bit dans un système de transmission MIMO/OFDM. GRETSI, Toulouse FRANCE, sept.2001.
- [CURN et al.,2012] J. CURN, D.MARINESCU, G.LACEY, V.CAHILL. Estimation with Non-White Gaussian Observation, Noise using a Generalised Ensemble Kalman Filter. 2012.
- [DAHIA,2005] Karim DAHIA. Nouvelles méthodes en filtrage particulaire. Application au recalage de navigation inertielle par mesures altimétriques. Thèse de doctorat de l'université Joseph Fourier, spécialité Mathématiques et Applications, Janv 2005.
- [DALY et al.,?] M. DALY, J. REILLY, M.MORELANDE. Rao-blackwellised particle filtering for blind system identification. McMaster University, Electrical and Computer Engineering Hamilton, Ontario, Canada & The University of Melbourne, Electrical and Electronic Engineering Melbourne, Victoria, Australia.

- [DEHEUVELS, 1977] Paul DEHEUVELS. Estimation non paramétrique de la densité par histogrammes généralisés. *Revue de statistique appliquée*, tome 25, n°3, p.5-42, 1977.
- [P.DEL MORAL,2008] P. DEL MORAL. Une théorie champ moyen du filtrage non linéaire. INRIA Bordeaux-Sud Ouest. Colloquium du lab MAP5, Université Paul Descartes, Mars 2008.
- [DEL MORAL,2008] P. DEL MORAL. Une introduction aux méthodes particulières en ingénierie stochastique. INRIA Centre Bordeaux-Sud Ouest, Journée de Proba-Stat de bordeaux, oct 2008.
- [DEL MORAL et al.,2006] P.DEL MORAL, J-C.NOYER, G.RIGAL, G.SALUT. Résolution particulière et traitement non-linéaire du signal : Applications RADAR-SONAR. Lab d'Analyse et Architecture des systèmes du CNRS, Toulouse et Digilog Aix en Provence, 2006.
- [DEL MORAL, 2009] P. DEL MORAL. Filtrage particulière : une introduction avec applications. INRIA Centre Bordeaux-Sud Ouest, Journées GDR Statistique et Santé, univ. Paris-Descartes, oct. 2009.
- [DUFLOS,2011] Emmanuel DUFLOS. Modélisation et Inférence Bayésienne : Filtrage particulière. Ecole Centrale de Lille-G3-ISD, Jan 2011.
- [DUBOIS et DAVY,2005] Corentin DUBOIS et Manuel DAVY. Suivi de Trajectoires Temps-Fréquence par Filtrage Particulaire. Colloque GRETSI 2005.
- [ENZO et al.,2010] Casasola ENZO, Clanzig KELLYA et Deville THIBAUT. Etude de cas : Filtre de Kalman. Mars 2010.
- [FAURE,1992] Olivier FAURE. Simulation du mouvement brownien et des diffusions. Thèse de doctorat Mathématiques Appliquées, Ecole Nationale des Ponts et Chaussées, Fev 1992.
- [FROGERAIS,2008] Paul FROGERAIS. Modélisation et identification en épilepsie : de la dynamique des populations neuronales aux signaux EEG. Thèse de doctorat, traitement de signal et télécommunications. Université de Rennes I, Juillet 2008.
- [FUJISAKI,1980] Masatoshi FUJISAKI. Contrôle stochastique et martingales. *Seminaire de probabilités (Strasbourg)*, Springer Verlag tome 14, p.256-281, 1980.
- [FUJISAKI,1981] Masatoshi FUJISAKI. Unicité des lois optimales du contrôle stochastique continu dans le cas complètement observable. *Annales scientifiques de l'Université de Clermont-Ferrand 2*, tome 69,série

Mathématiques, n°19, p.37-45, 1981.

- [GEIST et al.,2008] M. GEIST, O.PIETQUIN, G . FRICOUT. Filtrage bayésien de récompense. Article Supélec Groupe de recherche IMS Metz France, Arcelor Mittal Research département MCE, Maizières-lès-Metz, France 2008
- [GIREMUS,2005] Audrey GIREMUS. Apports de techniques de filtrage non linéaire pour la navigation avec les systèmes de navigation inertielle et le GPS. Thèse de doctorat de l'ENSAE spécialité : signal, Image, Acoustique, 2005.
- [GUECHI, 2010] El-Hadi GUECHI. Suivi de trajectoires d'un robot mobile non holonome :approche par modèle flou de Takagi-Sugeno et prise en compte des retards. Thèse de doctorat, Univ. de Valenciennes et du Hainaut Cambrésis, specialité Automatique et Génie Informatique, juin 2010.
- [GUYON,1999] Xavier GUYON. Méthodes numériques par chaînes de Markov. Université Paris 1 et Universidad de Los Andes, Ecole d'été de Mathématiques Mérida, Venezuela, sept 1999.
- [HAMLAOUI et al.,2004] S.HAMLAOUI, M.DANG, F.DAVOINE. Suivi des variations de pose et d'apparence de visages dans des séquences vidéo. Laboratoire HEUDIASYC-CNRS UTC, Université de Technologie de Compiègne, 2004.
- [HAYKIN,1991] S. HAYKIN. Adaptive filter theory. Prentice Hall. 1991
- [JEANBLANC et SIMON,2006] M.JEANBLANC, T.SIMON.Éléments de calculs stochastique. IRBID Septembre 2006.
- [JULIER et UHLMANN, 2004] Simon J.JULIER, Jeffrey K. UHLMANN. Unscented Filtering and Nonlinear Estimation. Proceedings of the IEEE, vol.92,n°3, march 2004.
- [JUTTEN,2010] Christian JUTTEN. Filtrage linéaire optimal. Options Images et Images et Automatique. Université Joseph Fourier- Polytech' Grenoble. Août 2010.
- [KJETIL, 2007] Kjetil Bergh ANONSEN and Oddvar HALLINGSTAD. Sigma point Kalman Filter underwater terrain-based navigation. Departement of Engineering Cybernetics, Norwegian University of Science and Technology and University Graduate Center (UniK), Norway, 2007.
- [GIREMUS et al.,2006] A.GIREMUS, J.Y. TOURNERET, V.CALMETTES et A.C.ESCHER. Apport des méthodes de filtrage particulaire pour

la navigation avec le GPS. SUPAERO-ENSEEIH-ENAC, journée navigation 2006.

[Le GLAND, 2003] François Le GLAND. Filtrage Particulaire. IRISA-INRIA RennesI, GRETSI 2003.

[Le GLAND, 2011] François Le GLAND. Filtrage Bayésien et Approximation Particulaire. Cours Ecole Nationale Supérieure de Techniques Avancées, module Commande des Systèmes, 2011.

[LAFARGE,2009] Florent LAFARGE. Introduction au filtrage de Kalman. Théorie du filtre de Kalman discret et applications. 2009.

[LAPEYRE,1993] B. LAPEYRE, E. PARDOUX, R.SENTIS. Introduction aux Méthodes de Monte Carlo. 25 i-ème Congrès d'Analyse Numérique, Laboratoire de Calcul Scientifique de Besançon, 22 et 23 mai 1993.

[MASCARILLA,2004] Laurent MASCARILLA. Filtre de Kalman pour le suivi. Univ. La Rochelle, 2004.

[MERWE et WAN, 2001] Rudolph Van Der MERWE, Eric A. WAN. The square-root unscented Kalman filter for state and parameter-estimation. Oregon Graduate Institute of Science and Technology 20000 NW Walker Road, Beaverton, Oregon 97008, USA, 2001.

[NOYER,1996] Jean-Charles NOYER. Traitement non-linéaire du signal radar par filtrage particulaire/ Thèse de doctorat en Automatique et Traitement du signal. Université Paul Sabatier Toulouse, 1996.

[OUVRARD,1973] Jean-Yves OUVRARD. Martingales locales et théorème de Girsanov dans les espaces de Hilbert réels séparables. Annales de l'I.H.P, section B, tome 9, n°4, p.351-368, 1973.

[PAPADAKIS,2007] Nicolas PAPADAKIS. Assimilation de données images : application au suivi de courbes et de champs de vecteurs. Thèse de doctorat en Mathématiques et Applications. Université de Rennes I, Novembre 2007.

[PARDOUX,1981] E.PARDOUX. Équations aux dérivées partielles associées à un problème de filtrage non linéaire. Annales scientifiques de l'Université de Clermont-Ferrand 2, tome 69, série Mathématiques, n°19, p.141-147, 1981.

[PARDOUX,1990] E.PARDOUX. Mathématiques pour l'ingénieur. DESS Université de PROVENCE 1991.

- [PARDOUX,1991] E.PARDOUX. Filtrage non linéaire et équations aux dérivées partielles stochastiques associés. In P.L Hennequin, editor, *Ecole d'été de Probabilités de Saint-Flour XIX-1989*, volume 1464 of *Lecture Notes in Mathematics*, pages 67-163. Springer Verlag, Berlin, 1991.
- [PARDOUX,2006] E.PARDOUX. Processus de Markov et applications, Algorithmes, Réseaux, Génome et Finance. Univ de Provence, sept 2006.
- [PAUL,2006] Nicolas PAUL. Filtrage Particulaire.Seminaire de Statistique, Conservatoire National des Arts et Métiers, 2006.
- [PETIT,2011] Nicolas PETIT. Filtre de Kalman. Centre Automatique et Systèmes, Unité Mathématiques et Systèmes MINES Paris Tech, avril 2011.
- [PROSPERI, 1997] Serge PROSPERI. Filtrage non linéaire en poursuite. Une approche par mélanges gaussiens. Seizième colloque GRETSI, Grenoble, sept 1997.
- [RACICOT et THEORET,2005] F-E. RACICOT & R.THEORET, Quelques applications du filtre de Kalman en Finance : Estimation et Prévision de la volatilité Stochastique et du rapport cours-bénéfices. Université de Montréal, 2005.
- [ROSSI et VILA,2003] V. ROSSI, J.P. VILLA. Filtrage non linéaire en temps discret par convolution de particules. Actes des XXXVèmes Journées de Statistique, 823-826, 2003.
- [ROSSI et VILA,2005] V. ROSSI, J.P. VILLA. Approche non paramétrique du filtrage de système non linéaire à temps discret et à paramètres inconnus. C.R ACAD. Sci. Paris. Ser I 340, 759-764, 2005.
- [ROSSI et VILLA, 2004] V.ROSSI, J.P. VILLA. Filtrage Non Linéaire par Convolution pour des Modèles Incertains. Laboratoire d'Analyse des systèmes et Biométrie UMR INRA-ENSAM, 2004.
- [ROSSI, 2004] Vivien ROSSI. Filtrage non linéaire par noyaux de convolution. Application à un procédé de dépollution biologique. Thèse de doctorat en Science : Information, Structures, Systèmes, 2004.
- [ROSSI, 2005] Vivien ROSSI.Utilisation du filtrage particulaire en dynamique des populations et sur bio-procédés. IRISA-INRIA Rennes, Equipe ASPI, Journées de Statistique et applications en biologie, univ. Rennes II, 2005.
- [ROTELLA,2003] Frédéric ROTELLA. Filtrage Optimal. Ecole National d'Ingénieurs de Tarbes, 2003.

- [RUBENTHALER,2010] Sylvain RUBENTHALER. Aspects théoriques et applications en filtrage non linéaire, système de particules et processus stochastiques. Thèse de doctorat HDR Mathématiques Appliquées, Univ. Nice-Sophia Antipolis, 2010.
- [SZPIRGLAS, 1979] J.SZPIRGLAS, G.MAZZIATTO. Modèle général de filtrage non linéaire et équations différentielles stochastiques associées. Annales de l'I.H.P., section B, tome 15, n°2, p.147-173, 1979.
- [STRAUSS,2008] O. STRAUSS, B.NEHME, K.LOQUIN. Estimation imprécise de la densité de probabilité. LIRMM, Université de Montpellier II, Avril 2008.
- [TORTRAT,1965] A. TORTRAT. Lois de probabilité sur un espace topologique complètement régulier et produits infinis à termes indépendants dans un groupe topologique. Annales de l'I.H.P., section B, tome 1, n°3, p.217-237, 1965.
- [TOURNERET,2007] Jean-Yves TOURNERET. Méthodes de Monte-Carlo par chaînes de Markov. Institut de recherche en informatique de Toulouse (IRIT), ENSEEIHT, Toulouse, 2007.
- [TRAMINI et al.,1997] S.TRAMINI, M.ANTONINI, M.BARLAUD et G.AUBERT. Filtrage non-linéaire dynamique pour la compression des images. Seizième colloque GRETSI, Grenoble, sept 1997.
- [VALADIER,1972] Michel VALADIER. Sur le filtre de Kalman-Bucy en temps continu. RAIRO, Mathématique, tome 6, n°2, p.9-22, 1972.
- [VILA et al,2009] J.P. VILA, J.P. GAUCHI et C. BIDOT. Identification de systèmes dynamiques microbiologiques complexes par filtrage non linéaire. 41èmes journées de Statistique, SFdS Bordeaux, 2009.
- [WIDYNSKI,2008] Nicolas WIDYNSKI. Filtre particulaire : suivi multi-objets dans les séquences multimodales. Lip6-Télécom Paris Tech, 2008.
- [YOUNDJÉ,2011] Élie YOUNDJÉ. Contribution à l'estimation non-paramétrique par la méthode du noyau. Mémoire d'HDR, Université de Rouen et du Havre, INSA de Rouen, 2011.
- [ZIADI, 2007] Anis ZIADI. Particules gaussiennes déterministes en maximum de vraisemblance non-linéaire : Application au filtrage optimal des signaux radar et GPS. Thèse de doctorat, spécialité Automatique et Traitement du Signal, Université de Toulouse, 2007.

Fiche de renseignement



Nom ROBINSON
Prénoms Matio Hobihery
Adresse Lot VE 83 Ambatovinaky
Antananarivo 101
Téléphone +261 33 11 457 86
e-mail : mat robinson2000@yahoo.fr

Titre de la thèse :

” ESTIMATION ET STABILITÉ EN FILTRAGE NON LINÉAIRE :
APPROXIMATION PARTICULAIRE ”

Nombre de page : 213

Nombre de tableaux : 17

Nombre de figures : 40

Mots-clés :

Filtrage non linéaire, Filtrage de Kalman, Filtrage de Julier et Uhlmann, Filtrage particulaire, Noyau de convolution, Oubli de la condition initiale.

Directeur de thèse : **RADIMBINDRAINIBE Falimanana**

Téléphone : +261 33 07 056 20

e-mail : falimanana@mail.ru

Co-directeur de thèse : **ANDRIANAHARISON Yvon**

Téléphone : +261 33 15 156 27

e-mail : yvonkyo@gmail.com

Résumé

Cette thèse est développée dans un domaine des sciences cognitives en particulier dans le traitement de signal. Plus précisément, elle résout des problèmes de filtrages non linéaires à des cas particuliers. Le thème sur l'approximation particulière et la stabilité en filtrage non linéaire traite quelques problématiques avec des propositions de solutions. Les filtrages linéaires à bruits gaussiens sont résolus par le filtrage de Kalman. Le filtrage de Kalman étendu est appliqué au modèle non linéaire gaussien mais exige la dérivabilité de la fonction dans l'équation d'observation. Pour les modèles fortement non linéaires, le filtrage de Julier et Uhlmann est le mieux adapté. Le problème de régularisation des filtres pour déterminer l'estimation des paramètres ou des commandes est traité avec le filtrage par noyau de convolution. Un exposé intitulé " Estimation d'une commande de signal non observé par filtrage non linéaire par noyau de convolution " est réalisé à ce sujet à l'ESPA. Dans le cas plus général, le théorème de Bayes permet d'avoir le filtre avec la décomposition classique en filtre prédit et filtre corrigé. Dans la modélisation, la structure markovienne du signal étudié non observé est une propriété essentielle pour aboutir à cette décomposition en filtre prédit et filtre corrigé. Le modèle de diffusion est un modèle typique dans les problèmes de filtrage. L'existence des bruits blancs gaussiens comme perturbateurs, permet d'obtenir le filtre avec une forme explicite. Le filtrage bayésien est adapté au système non linéaire contrôlé et au système non linéaire paramétré.

Une approche sur la résolution du problème d'existence et de construction de solution théorique du filtre optimal est développée. Généralement, la résolution est axée sur le cas gaussien. Avec le théorème de Girsanov, une solution explicite est obtenue. Cette solution est construite à partir d'une mesure absolument continue.

L'efficacité des méthodes est mesurée par la stabilité du filtre optimal au sens de l'oubli de la condition initiale. Chaque méthode a ses propres hypothèses dans la justification de l'oubli de la condition initiale du filtre optimal obtenu. Une étude selon différentes normes permet de vérifier la stabilité avec les méthodes classiques de filtrage. Un article intitulé "Oubli de la condition initiale dans un filtrage non linéaire gaussien et Filtre de Kalman Etendu (FKE) " et un autre intitulé " Filtrage Non Linéaire en Temps Discret et Application du filtre particulière avec un bruit d'observation uniforme " sont publiés dans MADA-Enlsa à ce sujet.

Le problème de mise en œuvre des filtres prédits et corrigés est résolu à partir de simulation d'exemple sur Scilab. Après une introduction des filtrages particuliers, une amélioration avec la régularisation des filtres avec le filtrage par noyau de convolution est proposée.

Les modèles à bruits d'observation quelconque, en particulier des bruits uniformes sont traités avec l'approximation particulière. Le filtrage particulière peut s'adapter à n'importe quel bruit d'observation en particulier un bruit uniforme. Mais les bruits faibles mettent en défaut le filtrage particulière. Le filtrage par noyau de convolution complétant le filtrage particulière est une solution efficace dans ce cas. Il permet aussi de résoudre les problèmes d'estimation des paramètres inconnus dans le modèle.

Les applications sur les exemples de poursuite de trajectoire et d'estimation paramétrique permettent de vérifier les théories.

Finalement cette thèse entraîne des perspectives de recherche entre autres, la résolution des équations aux dérivées partielles par filtrage particulière ou encore la résolution de problème de contrôle optimal.

Abstract

This thesis is developed in particular in a domain of the cognitive sciences in the treatment of signal. More precisely, it solves problems of nonlinear filtering to particular cases. The theme on the particulate approximation and the stability in filtering non bill a few problematic with propositions of solutions.

The linear filtering with Gaussian noises is solved by the filtering of Kalman. The extended Kalman filtering is applied to the non Gaussian model but requires the derivability of the function in the equation of observation. For the greatly nonlinear models, the filtering of Julier and Uhlmann is adapted best.

The problem of regularization of the filters to determine the evaluation of the parameters or orders is dealt with filtering by core of convolution. An exposition titled "Evaluation of a non observed signal order by nonlinear filtering by core of convolution" is achieved to this topic to the ESPA.

In the more general case, the theorem of Bayes permits to have the filter with the classic decomposition in filter predicts and filter correct version. In the modeling, the Markovian structure of the non observed studied signal is an essential property to succeed to this decomposition in filter predicts and filter correct version. The model of diffusion is a typical model in the problems of filtering. The existence of the Gaussian white noises as hecklers permits to get the filter with an explicit shape. The filtering bayesien is adapted to the non controlled system and the non set system.

An approach on the resolution of the existence problem and constructional theoretical solution of the optimal filter is developed. Generally, the resolution is centered on the Gaussian case. With the theorem of Girsanov, an explicit solution is gotten. This solution is constructed from an absolutely continuous measure. The efficiency of the methods is measured by the stability of the optimal filter to the sense of the oblivion of the initial condition. Every method has its own hypotheses in the justification of the oblivion of the initial condition of the optimal filter gotten. A survey according to different norms permits to verify the stability with the approved methods of filtering. An article titled "Oblivion of the initial condition in a non Gaussian filtering and Filter of Ex-tended Kalman (FKE)" and another title "Nonlinear Filtering in Discreet Time and Applica-tion of the particulate filter with a "uniform observation noise is published in MADA-Enelsa to this topic.

The problem of setting in work of the filters predicted and correct versions are solved from simulation of example on Scilab. After an introduction of the particulate filtering, an improvement with the regularization of the filters with filtering by core of convolution is proposed.

The models to noises of any observation, in particular some uniform noises are dealt with the particulate approximation, the particulate filtering can adjust in particular at any noise of observation a uniform noise. But the weak noises put in defect the particulate filtering. Filtering by core of convolution completing the particulate filtering is an efficient solution in this case. it also permits to solve the problems of evaluation of the unknown parameters in the model.

The applications on the examples of pursuit of trajectory and parametric evaluation permit to verify the theories.

Finally this thesis entails perspectives of research among others, the resolution of the equations to the partial derivatives by particulate filtering or the resolution of optimal control problem.