Nouvelles approches en filtrage particulaire : application au recalage de la navigation inertielle

Achille Murangira

Thèse
de doctorat
de l'UTT

Achille MURANGIRA

Nouvelles approches en filtrage particulaire. Application au recalage de la navigation inertielle

Spécialité :
Optimisation et Sûreté des Systèmes

2014TROY0011 Année 2014

utt
université de technologie
Troyes

THESE

pour l'obtention du grade de

DOCTEUR de l'UNIVERSITE DE TECHNOLOGIE DE TROYES

Spécialité : OPTIMISATION ET SURETE DES SYSTEMES

présentée et soutenue par

Achille MURANGIRA

le 25 mars 2014

Nouvelles approches en filtrage particulaire. Application au recalage de la navigation inertielle

JURY

M. J.-Y. TOURNERET	PROFESSEUR DES UNIVERSITES	Président
M. J.-M. ALLARD	INGENIEUR DE RECHERCHE	Examinateur
M. K. DAHIA	DOCTEUR	Examinateur
M. F. LE GLAND	DIRECTEUR DE RECHERCHE INRIA	Rapporteur
M. C. MACABIAU	PROFESSEUR ENAC	Rapporteur
M. C. MUSSO	DOCTEUR	Examinateur
M. I. NIKIFOROV	PROFESSEUR DES UNIVERSITES	Directeur de thèse
M. H. SNOUSSI	PROFESSEUR DES UNIVERSITES	Examinateur

Personnalité invitée

M. G. CLERC-RENAUD	INGENIEUR

Remerciements

Tout d'abord, je tiens à remercier Igor Nikiforov d'avoir dirigé cette thèse.

Merci à François Le Gland et Christophe Macabiau d'avoir accepter d'en être les rapporteurs, et à Jean-Yves Tourneret qui a assuré la présidence du jury. Je suis également reconnaissant envers Hichem Snoussi pour sa participation au jury.

Merci à la DGA et à l'Onera pour avoir co-financé la thèse et à Gilles Clerc-Renaud pour avoir manifesté beaucoup d'intérêt pour mes recherches.

Grand merci également à Jean-Michel Allard, Karim Dahia de m'avoir accompagné dans la réalisation de ces travaux. Leur expertise et leur disponibilité ont été essentielles au bon déroulement de ce doctorat. Je remercie également Christian Musso pour ses idées, son ouverture scientifique et sa sympathie qui n'ont jamais fait défaut tout au long de ces trois années de recherches.

Merci aux collègues du DCPS et du DTIM avec qui j'ai travaillé et échangé dans une atmosphère conviviale et détendue.

Enfin merci à mon père, ma mère et mes frères pour leur soutien et leurs encouragements répétés.

Résumé

Les travaux présentés dans ce mémoire de thèse concernent le développement et la mise en oeuvre d'un algorithme de filtrage particulaire pour le recalage de la navigation inertielle par mesures altimétriques. Le filtre développé, le MRPF (Mixture Regularized Particle Filter), s'appuie à la fois sur la modélisation de la densité a posteriori sous forme de mélange fini, sur le filtre particulaire régularisé ainsi que sur l'algorithme mean-shift clustering. Nous proposons également une extension du MRPF au filtre particulaire Rao-Blackwellisé appelée MRBPF (Mixture Rao-Blackwellized Particle Filter). L'objectif est de proposer un filtre adapté à la gestion des multimodalités dues aux ambiguïtés de terrain. L'utilisation des modèles de mélange fini permet d'introduire un algorithme d'échantillonnage d'importance afin de générer les particules dans les zones d'intérêt. Un second axe de recherche concerne la mise au point d'outils de contrôle d'intégrité de la solution particulaire. En nous appuyant sur la théorie de la détection de changement, nous proposons un algorithme de détection séquentielle de la divergence du filtre. Les performances du MRPF, MRBPF, et du test d'intégrité sont évaluées sur plusieurs scénarios de recalage altimétrique.

Mots clés : filtrage particulaire, recalage altimétrique, modèles de mélange, mean-shift clustering, échantillonnage d'importance, maximum a posteriori, test d'intégrité, détection de changement, algorithme CUSUM.

Abstract

This thesis deals with the development of a mixture particle filtering algorithm for inertial navigation update via radar-altimeter measurements. This particle filter, the so-called MRPF (Mixture Regularized Particle Filter), combines mixture modelling of the posterior density, the regularized particle filter and the mean-shift clustering algorithm. A version adapted to the Rao-Blackwellized particle filter, the MRBPF (Mixture Rao-Blackwellized Particle Filter), is also presented. The main goal is to design a filter well suited to multimodal densities caused by terrain amibiguity. The use of mixture models enables us to introduce an alternative importance sampling procedure aimed at proposing samples in the high likelihood regions of the state space. A second research axis is concerned with the development of particle filtering integrity monitoring tools. A novel particle filter divergence sequential detector, based on change detection theory, is presented. The performances of the MRPF, MRBPF and the divergence detector are reported on several terrain navigation scenarios.

Keywords : particle filtering, terrain navigation, mixture models, mean-shift clustering, importance sampling, maximum a posteriori, integrity monitoring, change detection, CUSUM algorithm.

Table des matières

Notations et acronymes

Notations

$\mathbb{P}$	mesure de probabilité
k	temps discret
x_k	vecteur d'état
y_k	vecteur d'observation
$x_{0:k}$	ensemble $\{x_0, x_1, \cdots x_k\}$
$y_{0:k}$	ensemble $\{y_0, y_1, \cdots y_k\}$
d	dimension de l'espace d'état
$\mathbb{E}(.)$	espérance
$\mathbb{E}_p(.)$	espérance par rapport à la densité p
$p_k, p_{k\|k}$	densité de x_k sachant $y_{0:k}$
$p_{k\|k-1}$	densité de x_k sachant $y_{0:k-1}$
Cov	covariance
Var	variance
$\hat{x}_{MAP}$	maximum a posteriori
Q_k	matrice de covariance du bruit de dynamique
R_k	matrice de covariance du bruit d'observation
N	nombre de particules/échantillons
$x \sim p$	échantillon issu de p
x^i	particule/échantillon
ω^i	poids de la particule x^i
N_{eff}	taille effective de l'échantillon
N_{th}	seuil de ré-échantillonnage pour le critère N_{eff}
K	noyau de régularisation
h_{opt}	facteur de dilatation optimal
$\mathcal{X}_k$	ensemble des particules à l'instant k : $\{x_k^i,\, i = 1, \cdots, N\}$
$\stackrel{\triangle}{=}$	égal par définition à
$\propto$	proportionnel à
$D_{KL}\left(f \parallel g\right)$	divergence de Kullback-Leibler de g à f
$\mathcal{U}([a,b])$	loi uniforme sur $[a,b]$
$\langle F, G \rangle$	$\int F(x)G(x)dx$

$[\![m, n]\!]$	segment des entiers compris entre m et n
$\lfloor x \rfloor$	partie entière de $x \in \mathbb{R}$
$\mathbb{1}_A$	fonction indicatrice pour un événement A
$\det(A)$	déterminant de la matrice A
$\mathrm{diag}(u_1, \ldots, u_n)$	matrice de diagonale d'éléments $u_1, \ldots, u_n$
$\varphi(.\|\mu, \Sigma)$	densité de probabilité de la distribution gaussienne de moyenne μ et de covariance Σ
dilog	fonction di-logarithme
plim	limite en probabilité
$\alpha_{j,k}$	coefficient de mélange de la composante d'indice j à l'instant k
$C_{j,k}$	ensemble des particules appartenant au cluster d'indice j à l'instant k
$I_{j,k}$	indices des particules appartenant au cluster d'indice j
$m_{h,G}$	vecteur mean-shift évalué avec la fenêtre de lissage h et le noyau G
R_m	rayon de fusion dans l'algorithme mean-shift
N_{eff}^j	taille effective de l'échantillon du cluster d'indice j
α_{min}	seuil de suppression d'un cluster
$\hat{x}_j^*$	estimation du j-ème mode de la densité conditionnelle
$N_{th,MAP}$	seuil d'utilisation de lois d'importance locales dans le MRPF
λ	latitude géographique
ϕ	longitude géographique
z	altitude géographique
R_{n2b}	matrice de rotation du repère de navigation (n) au repère engin (b)
ψ	lacet
θ	tangage
φ	roulis
$\boldsymbol{\gamma}_m$	force spécifique
$\boldsymbol{g}$	pesanteur terrestre
$\boldsymbol{\omega_{ie}}$	vitesse de rotation de la terre dans le repère TGL
h_{MNT}	fonction d'observation correspondant au MNT
H_0	hypothèse nulle
H_1	hypothèse alternative
D_k	fonction de décision
t_a	temps d'arrêt du test
t_0	vrai instant de changement
$\bar{T}$	temps moyen entre deux fausses alarmes
$\overline{\tau}^*$	pire retard moyen de détection
$\mathrm{ess\,sup}$	essentiel supremum
ϵ_k	innovation du filtre optimal
ϵ_k^N	innovation du filtre particulaire
r_k	processus d'innovation normalisé
τ	délai moyen de détection

Abbréviations

i.i.d.	indépendant(e)s et identiquement distribué(e)s
p.s.	presque sûrement
v.a.	variable aléatoire

Acronymes

AMISE	Asymptotic Mean Integrated Squared Error
ARL	Fonction ARL (Average Run Length)
BCR	Borne de Cramer-Rao classique ou a posteriori
CUSUM	Algorithme des sommes cumulées pour la détection de changement de moyenne
EKF	filtre de Kalman étendu
KF	filtre de Kalman
KPKF	Kalman Particle Kernel Filter
MAP	Maximum a posteriori
MNT	Modèle numérique de terrain
MRPF	Mixture Regularized Particle Filter
MRBPF	Mixture Rao-Blackwellized Particle Filter
MSE	Mean Squared Error
MISE	Mean Integrated Squared Error
RBPF	Rao-Blackwellized Particle Filter
RPF	Regularized Particle Filter
SIR	Sequential Importance Resampling
SIS	Sequential Importance Sampling
TFA	Taux de fausses alarmes
TGL	Trièdre géographique local
TND	Taux de non détection
UKF	filtre de Kalman sans parfum

Introduction générale

Contexte de la thèse

De manière générale, la navigation est la science et l'ensemble des techniques permettant de connaître la position d'un véhicule en mouvement. Les systèmes de navigation inertielle fournissent une estimation des coordonnées de position, vitesse et attitude par double intégration des mesures d'accélération et de vitesse angulaire délivrées par les accéléromètres et gyromètres de la centrale inertielle. Ces systèmes permettent une navigation autonome contrairement aux systèmes de radio-navigation qui reposent sur la transmission de signaux par un émetteur externe. Les capteurs de la centrale inertielle étant entachés d'erreurs de mesure, la solution de navigation subit une dérive. Il est nécessaire de recaler la navigation à l'aide de mesures auxiliaires indépendantes de la centrale inertielle.

Dans cette thèse, nous nous intéressons au recalage d'un aéronef via des mesures de hauteur-sol fournies par un radio-altimètre, et un modèle numérique de terrain (MNT). L'estimation de la position de l'aéronef se fait alors en comparant le profil de terrain obtenu à partir des mesures altimétriques avec le profil de terrain issu de la carte embarquée. Les algorithmes de recalage fonctionnent soit par batch, c'est-à-dire par corrélation de blocs de mesures consécutives avec le MNT, soit récursivement en mettant à jour l'estimation de la position à chaque mesure reçue. Les systèmes batch comme le TERCOM (Terrain Contour Matching) supposent généralement une bonne connaissance de la vitesse et de l'altitude ce qui restreint leur domaine d'utilisation. Les méthodes récursives telles que le SITAN (Sandia Inertial Terrain Navigation) reposent sur l'utilisation d'un ou plusieurs filtres de Kalman étendus (EKF). Cependant lorsque l'incertitude sur la position initiale est grande ou lorsque le terrain est très accidenté, leur efficacité est limitée. Dans les deux dernières décennies, les filtres particulaires ont connu un intérêt croissant car ils permettent d'estimer récursivement l'état d'un système dynamique, sans imposer de contraintes sur la linéarité du modèle. Ils reposent sur l'approximation de la loi de l'état du système, conditionnellement aux mesures passées et sont particulièrement adaptés lorsque cette loi est multimodale. Les caractéristiques de non-linéarité et de multimodalités sont justement typiques de la problématique du recalage par mesures altimétriques :

- en début de recalage, du fait de l'ambiguïté de terrain, il existe plusieurs positions compatibles avec les mesures. Cela se traduit par une loi conditionnelle multimodale.
- l'équation d'observation du radio-altimètre reliant la position de l'aéronef à la hauteur-sol est non-linéaire.

Le filtre particulaire est une méthode de Monte Carlo séquentielle dont les fondements datent probablement des années 70 [53, 52], mais la plupart des travaux ont eu lieu à partir des années 90 avec les contributions de Gordon et. al [50], Kong et al. [69], Del Moral [30], Del Moral et. al [33], ou encore Pitt et Shepard [105]. Les algorithmes particulaires standards reposent sur la succession d'étapes de prédiction, de correction et de ré-échantillonnage qui maintiennent un système de particules pondérées approchant la loi a posteriori. Lors de l'étape de prédiction, les particules sont propagées selon l'équation de dynamique, puis leur poids sont ajustés selon leur adéquation avec l'observation courante via l'évaluation de la fonction de vraisemblance. Le ré-échantillonnage consiste à favoriser les particules dont les poids sont significatifs afin de conserver une approximation de la loi a posteriori non-dégénérée.

Bergman [6] est à l'origine des premiers travaux sur le filtrage particulaire pour le recalage altimétrique. Dans son étude, il suppose que l'altitude et les composantes de vitesse sont connues et cherche à estimer la position horizontale du véhicule. Les filtres implémentés se montrent robustes aux divergences et sont capables d'atteindre la borne d'erreur minimale (Borne de Cramer-Rao a posteriori), pour une incertitude initiale en position modeste. Des travaux ultérieurs ont permis d'étudier l'apport de nouveaux filtres à la problématique du recalage altimétrique. Le filtre particulaire régularisé de Oudjane et Musso [96], initialement développé pour le pistage de cible, a permis de rendre l'étape de ré-échantillonnage plus robuste. Dans ses travaux sur la navigation par corrélation de terrain, Nordlund [94] utilise Le Rao-Blackwellized Particle Filter. Ce filtre exploite la nature linéaire de l'équation de mesure altimétrique, conditionnellement à la position horizontale, ce qui permet de restreindre l'échantillonnage des particules aux coordonnées de position horizontale, les composantes d'altitude et de vitesse étant estimées grâce à une batterie de filtres de Kalman. Ceci permet en principe de réduire le nombre de particules utilisées pour des performances identiques. On peut également citer le développement du KPKF de Dahia [28] qui permet d'avoir des taux de convergence plus importants lorsque l'incertitude initiale est importante tout en réduisant le nombre de particules utilisées.
Néanmoins, il y a deux situations spécifiques pouvant provoquer la divergence du filtre :

1. Lorsque l'approximation particulaire est multimodale, on observe [67, 118] que les étapes de ré-échantillonnage successives peuvent provoquer la perte prématurée d'un ou plusieurs modes. Lorsque le mode perdu correspond à la vraie position de l'aéronef, le filtre diverge automatiquement. Ce risque est d'autant plus important que le terrain est ambigu car, dans ce cas le filtre est multimodal.
2. Lorsque le terrain connait une variation brusque sous forme d'un fort gradient de terrain (par exemple un canyon), les méthodes d'échantillonnage standards intervenant dans l'implémentation du filtre, sont mises en difficulté. Cette situation survient également lorsque la fonction de vraisemblance, qui intervient lors de l'étape de correction, est très localisée : c'est le cas si la variance du bruit de mesure est très petite. Lorsque cette situation survient, le risque est de n'avoir que quelques particules avec un poids significatif, ce qui peut faire diverger le filtre.

Une autre problématique soulevée par le filtrage particulaire est la détection automatique et séquentielle de la divergence : cela permet de contrôler l'intégrité de la solution de recalage fournie

par le filtre et, éventuellement, de le ré-initialiser. Dans ces travaux nous considérons que la divergence correspond à une situation où l'erreur d'estimation est incompatible avec le domaine de confiance fourni par le filtre particulaire.

Nous proposons dans cette thèse, une approche permettant de conserver le plus longtemps possible, les modes de la loi conditionelle, jusqu'à ce que l'ambigüité soit levée : les filtres particulaires développés à cet effet sont le MRPF (Mixture Regularized Particle Filter) et son extension au filtre Rao-Blackwellisé, noté MRBPF (Mixture Rao-Blackwellized Particle Filter). Ces filtres s'appuient sur le MPF (Mixture Particle Filter) [118] où la densité cible est exprimée sous forme d'un mélange de densités fini. Idéalement, chaque composante du mélange est monomodale et approximée par un sous-ensemble de particules pondérées. Nous utilisons dans le MRPF et le MRBPF l'algorithme de clustering mean-shift permettant de grouper automatiquement les particules associées au même mode. Par ailleurs, nous introduisons une procédure de régularisation locale adaptée aux multimodalités.
Le MRPF et le MRBPF sont ensuite combinés avec une méthode d'échantillonnage d'importance basée sur les maxima locaux de la loi conditionnelle, l'objectif étant de propager efficacement les particules dans les zones d'intérêt. Les algorithmes résultants sont le MRPF-MAP et le MRBPF-MAP.
Enfin, nous introduisons un algorithme de détection séquentielle de divergence que nous testons sur plusieurs scénarios de recalage : la première difficulté est de définir un test d'hypothèse permettant de décider entre la divergence et la non-divergence du filtre. Classiquement, la divergence en filtrage particulaire est souvent confondue avec le phénomène de dégénérescence, qui est évalué empiriquement grâce à des indicateurs comme la taille effective de l'échantillon ou l'entropie des poids. Pourtant en pratique, la divergence du filtre particulaire - au sens d'une "grande" erreur d'estimation - n'implique pas systématiquement la dégénérescence, d'où la nécessité d'introduire un autre critère.

Organisation du mémoire

Ce document est organisée en 5 chapitres :

- Le chapitre 1 est un aperçu général sur la problématique du filtrage linéaire et non-linéaire. Les algorithmes classiques (filtre de Kalman et ses variantes, méthodes de maillages) et leur cadre d'applicabilité y sont présentés.
- Le chapitre 2 présente les méthodes standards de filtrage particulaire avec un accent sur leur utilisation dans le cadre du recalage altimétrique. Les difficultés soulevées par le maintien des modes de la densité conditionnelle dans un filtre particulaire y sont détaillées.
- Le chapitre 3 présente brièvement les systèmes de navigation inertielles et la problématique du recalage radio-altimétrique.
- Le chapitre 4 introduit plusieurs algorithmes de filtrage particulaire (MRPF, MRBPF et leurs variantes), développés afin d'approcher plus efficacement la loi a posteriori en début de recalage lorsque celle-ci est fortement multimodale. Une méthode d'échantillonnage d'importance permettant de propager les particules dans les zones de forte probabilité est

également proposée. Nous présentons également les simulations expérimentales permettant de quantifier l'apport de chaque algorithme sur plusieurs scénarios de recalage altimétrique.
– Le chapitre 5 est consacré aux tests d'intégrité pour le filtrage particulaire. Nous formulons le problème de la détection de divergence sous forme d'un test d'hypothèses portant sur l'innovation du filtre. L'algorithme CUSUM (CUmulative SUM) est utilisé pour contrôler la statistique de test. Les performances du test développé sont étudiées dans le cadre du recalage altimétrique en terme de fausses alarmes, de délai de détection et de non détection.

Liste de publications

Le travail de recherche a donné lieu à trois publications dans des congrès internationaux :

▷ Murangira, A., Musso, C., *Proposal Distribution for Particle Filtering Applied to Terrain Navigation*, Proceedings of the 21st European Signal Processing Conference (EUSIPCO), Marrakech, September 9–13, 2013

▷ Murangira, A., Musso, C., Nikiforov, I., *Particle filter divergence monitoring with application to terrain navigation*, Proceedings of the 15th International Conference on Information Fusion (FUSION), Singapore, July 9–12, 2012

▷ Murangira, A., Musso, C., Dahia, K., Allard, J., *Robust regularized particle filter for terrain navigation*, Proceedings of the 14th International Conference on Information Fusion (FUSION), Chicago, July 5–8, 2011

Chapitre 1

Filtrage statistique : méthodes analytiques et numériques

1.1 Introduction

Ce chapitre introduit la notion de filtrage d'un point de vue général. Après avoir exposé le filtre optimal comme solution au problème de filtrage, nous nous décrivons les méthodes permettant de calculer de manière exacte ou approchée le filtre optimal. Nous présentons d'abord les méthodes analytiques (le filtre de Kalman et ses dérivées) avant de passer au méthodes de résolution numériques.

1.2 Estimation bayésienne et filtrage en temps discret

Le filtrage consiste à estimer les variable d'états d'un système dynamique que l'on observe partiellement à travers une série de mesures. Les états $(x_k)_{k\geq 0}$ du système sont soumis à une évolution dans le temps décrite par une équation de dynamique du type :

$$x_k = f_k(x_{k-1}, w_k) \tag{1.1}$$

où le vecteur des états x_k prend ses valeurs dans $\mathbb{R}^d$, $d \geq 1$ étant un entier désignant le nombre d'états. La suite de variables aléatoires (v.a.) $(w_k)_{k\geq 0}$ à valeurs dans $\mathbb{R}^s$, $s \geq 1$, modélise les perturbations aléatoires de la dynamique et l'imperfection du modèle. f_k est une fonction à valeurs dans $\mathbb{R}^d$.
Ce système est partiellement observé à travers une série de mesures y_k fournies par un instrument. Ce processus d'observation est modélisé par l'équation suivante :

$$y_k = h(x_k) + v_k \tag{1.2}$$

h est une fonction d'observation connue à valeurs dans $\mathbb{R}^m$, $m \geq 1$ et $(v_k)_{k\geq 0}$ une suite de v.a. décrivant les erreurs de mesures dont l'intensité est fonction des imperfections des capteurs et du rapport signal à bruit. Étant donné la série d'observations bruitées $\{y_0, y_1, y_2, \cdots y_k\}$, on souhaite estimer l'état du système x_k.

1.2.1 Estimation bayésienne

Le problème consistant à déterminer l'état du système grâce aux mesures $(y_k)_{k\geq 0}$ s'appelle estimation. L'estimation de x_k se fait souvent selon un critère de minimisation d'un risque. Par exemple, l'estimateur $\hat{x}_k$ minimisant l'erreur quadratique moyenne (ou estimateur MMSE pour *minimum mean squared error*)

$$\mathbb{E}\left(\|x_k - \hat{x}_k\|_2^2\right) = \operatorname{trace}\left(\mathbb{E}\left(x_k - \hat{x}_k\right)\left(x_k - \hat{x}_k\right)^T\right)$$

est l'espérance conditionnelle [73]

$$\hat{x}_k^{MMSE} = \mathbb{E}\left(x_k \mid y_0, \ldots, y_k\right) \stackrel{\triangle}{=} \int x_k p(x_k \mid y_0, \ldots, y_k)\, dx_k \tag{1.3}$$

Un deuxième estimateur, utilisé notamment lorsque la loi de densité conditionnelle $p(x_k \mid y_0, \ldots, y_k)$ est multimodale, est le maximum a posteriori [29], que nous notons ici $\hat{x}_k^{MAP}$ et qui est défini par

$$\hat{x}_k^{MAP} = \arg\max_{x_k} p(x_k \mid y_0, \ldots, y_k) \tag{1.4}$$

Dans tous les cas, la formulation de ces estimateurs nécessite de déterminer la loi de probabilité conditionnelle de x_k sachant les mesures passées $\{y_0, y_1, y_2, \cdots y_k\}$: c'est la problématique du filtrage. De façon plus générale, l'estimation de la loi de x_k sachant $y_{0:p} = \{y_0, y_1, y_2, \cdots y_p\}$ s'appelle :

- prédiction si $p < k$
- filtrage si $p = k$
- lissage si $p > k$

Pour résoudre le problème du filtrage non-linéaire en temps discret, on se place dans le cadre des modèles à espace d'état :

- $(w_k)_{k\geq 0}$ est une suite de v.a. mutuellement indépendantes et indépendante de x_0. x_k est donc markovien, sa loi est entièrement définie par sa loi initiale $\mu_0(dx) = \mathbb{P}(x_0 \in dx)$ et le noyau de transition Q_k tel que

$$\mathbb{P}(x_k \in A | x_{k-1} = x') = \int_A Q_k(x', dx) \tag{1.5}$$

- Le bruit v_k est une suite de v.a. mutuellement indépendantes et indépendante du processus x_k.
- L'état initial x_0 est indépendant de v_k et du bruit de dynamique w_k

Dans toute la suite, on supposera que la loi conditionnelle de x_k sachant $y_{0:k}$ admet une densité par rapport à la mesure de Lebesgue. Par abus de langage, on confondra parfois la loi conditionnelle de x_k avec sa densité.

1.2.2 Le filtre optimal

On note $p_k = p(x_k \mid y_{0:k})$ la densité conditionnelle de x_k sachant les observations passées $y_{0:k} = \{y_0, y_1, \cdots y_k\}$. La suite de densités de probabilité $(p_k)_{k\geq 0}$ est appelée filtre optimal. Une

remarque importante pour la résolution du problème du filtrage est la suivante : le calcul de p_k peut se faire de manière récursive. En effet, p_k est entièrement déterminé par la donnée de p_{k-1} et de la mesure y_k sous réserve que l'on puisse évaluer la densité de probabilité $\mathbb{P}(y_k \mid x_k) = g_k(x_k)$ appelée fonction de vraisemblance.
Le calcul récursif de p_k peut alors se décomposer en une étape dite de prédiction et une étape de correction.

Prédiction L'étape de prédiction calcule la loi prédite définie par $p_{k|k-1}(dx) = \mathbb{P}(x_k \in dx \mid y_{0:k-1})$ à partir de p_{k-1} et de la dynamique du système, caractérisée par le noyau de transition Q_k.

$$p_{k|k-1}(dx) = \int_{\mathbb{R}^d} p_{k-1}(dx')Q_k(x', dx) \tag{1.6}$$

En effet,

$$\begin{aligned} p_{k|k-1}(dx) &= p(x_k \in dx|y_{0:k-1}) \\ &= \int p(x_k \in dx, x_{k-1}|y_{0:k-1})dx_{k-1} \\ &= \int p(x_k \in dx|x_{k-1}, y_{0:k-1})p(x_{k-1}|y_{0:k-1})dx_{k-1} \end{aligned}$$

et étant donné le caractère markovien de x_k, $p(x_k \in dx|x_{k-1}, y_{0:k-1}) = p(x_k \in dx|x_{k-1})$, d'où

$$\begin{aligned} p_{k|k-1}(dx) &= \int p(x_k \in dx|x_{k-1})p(x_{k-1}|y_{0:k-1})dx_{k-1} \\ &= \int Q_k(x_{k-1}, dx)p_{k-1}(dx_{k-1}) \end{aligned}$$

Correction Cette étape permet d'obtenir le filtre p_k en utilisant la mesure y_k pour corriger la loi prédite $p_{k|k-1}$ grâce à la formule de Bayes.

$$p_k = \frac{g_k(x_k)p_{k|k-1}}{\int_{\mathbb{R}^d} g_k(x')p_{k|k-1}(dx')} \tag{1.7}$$

En effet,

$$\begin{aligned} p_k &= p(x_k|y_{0:k}) \\ &= \frac{p(y_k|x_k)p(x_k|y_{0:k-1})p(y_{0:k-1})}{p(y_k)} \\ &= \frac{g_k(x_k)p(x_k|y_{0:k-1})}{p(y_k|y_{0:k-1})} \\ &= \frac{g_k(x_k)p_{k|k-1}}{\int_{\mathbb{R}^d} g_k(x')p_{k|k-1}(dx')} \end{aligned}$$

La réalisation des étapes de prédiction et de correction nécessite le calcul d'intégrales multi-dimensionnelles dont le calcul analytique n'est possible que sous certaines conditions. Par exemple dans le cas d'un modèle linéaire gaussien, on dispose d'expressions exactes grâce à l'algorithme du filtre de Kalman présenté ci-après. De manière générale, il est nécessaire de recourir à des approximations soit par des méthodes analytiques faisant intervenir une simplification du modèle (linéarisation pour le filtre de Kalman étendu par exemple, cf. section 1.3.1.2), soit par des méthodes numériques fondées sur une discrétisation de l'espace d'état (méthodes de maillage), soit par des méthodes de Monte Carlo (cf. chapitre 2).

1.3 Méthodes de résolution du problème de filtrage

1.3.1 Méthodes analytiques

1.3.1.1 Le filtre de Kalman (KF)

Le filtre de Kalman a été développé par Rudolf Emil Kalman en 1960 pour les modèles discrétisés [65]. Une version en temps continu a été élaborée une année après par Kalman et Bucy. C'est le filtre optimal pour les modèles linéaires gaussiens de la forme suivante :

$$\begin{cases} x_k &= F_k x_{k-1} + B_k + w_k \\ y_k &= H_k x_k + D_k + v_k \end{cases}$$

avec les hypothèses :

- $F_k \in \mathcal{M}_{d,d}(\mathbb{R})$, $B_k \in \mathcal{M}_{d,1}(\mathbb{R})$, $H_k \in \mathcal{M}_{m,d}(\mathbb{R})$ et $D_k \in \mathcal{M}_{m,1}(\mathbb{R})$ sont des matrices déterministes et connues
- les bruits w_k et v_k à valeurs dans $\mathbb{R}_d$ et $\mathbb{R}_m$ sont des bruits blancs gaussiens de matrices de covariance respectives Q_k et R_k. w_k et v_k sont mutuellement indépendants et indépendants de la condition initiale x_0.
- La condition initiale x_0 suit une loi gaussienne de moyenne m_0 et de matrice de covariance P_0.

Sous ces hypothèses, on montre que la loi jointe de (x_k, y_k) est gaussienne ce qui implique que la loi a posteriori p_k l'est également. Le filtre p_k est donc entièrement déterminé par la donnée de sa moyenne $\hat{x}_k$ et de sa matrice de covariance P_k. Le filtre de Kalman est un algorithme qui permet de calculer analytiquement et récursivement ces 2 moments via des étapes de prédiction (1.6) et de correction (1.7).

1. *Prédiction*

$$\begin{cases} \hat{x}_{k|k-1} &= F_k \hat{x}_{k-1} + B_k \\ P_{k|k-1} &= F_k P_{k-1} F_k^T + Q_k \end{cases} \tag{1.8}$$

2. *Correction*

$$\begin{cases} \hat{x}_k &= \hat{x}_{k|k-1} + K_k[y_k - (H_k \hat{x}_{k|k-1} + D_k)] \\ K_k &= P_{k|k-1} H_k^T [H_k P_{k|k-1} H_k^T + R_k]^{-1} \\ P_k &= [I - K_k H_k] P_{k|k-1} \end{cases} \tag{1.9}$$

Le terme K_k désigne la matrice de gain de Kalman. Il permet d'obtenir la moyenne a posteriori $\hat{x}_k = \mathbb{E}(x_k|y_{0:k})$ et la covariance a posteriori $P_k = \mathrm{Cov}(x_k|y_{0:k})$ en faisant intervenir la mesure y_k via le terme dit d'innovation $\epsilon_k = y_k - (H_k\hat{x}_{k|k-1} + D_k)$. Les équations de Kalman peuvent être obtenues en appliquant directement les équations (1.6) et (1.7) afin de calculer successivement la moyenne et la covariance prédites $\left(\hat{x}_{k|k-1}, P_{k|k-1}\right)$, puis la moyenne et la covariance corrigées $(\hat{x}_k, P_k)$. Notons que les matrices $(K_k)_{k\geq 0}$, $(P_k)_{k\geq 0}$ et $\left(P_{k|k-1}\right)_{k\geq 0}$ ne dépendent pas des mesures $(y_k)_{k\geq 0}$ et peuvent être calculées hors ligne.
Lorsque le bruit de dynamique ou de mesure n'est plus gaussien, le filtre de Kalman est le filtre linéaire à variance minimale.

Filtre de Kalman informationnel Le filtre de Kalman informationnel est une implémentation qui propage l'inverse J_k de la matrice de de covariance P_k, appelée matrice d'information.
Introduisons $J_{k|k-1} = P_{k|k-1}^{-1}$ et $J_k = P_k^{-1}$. Le calcul récursif de $J_{k|k-1}$ et J_k s'obtient en utilisant le lemme d'inversion matricielle (cf. annexe A.1) :

1.

$$J_{k|k-1} = P_{k|k-1}^{-1} = \left[F_k P_{k-1} F_k^T + Q_k\right]^{-1} \tag{1.10}$$
$$= \left[F_k J_{k-1}^{-1} F_k^T + Q_k\right]^{-1} \tag{1.11}$$
$$= Q_k^{-1} - Q_k^{-1} F_k [F_k^T Q_k F_k + J_{k-1}]^{-1} F_k^T Q_k^{-1} \tag{1.12}$$

2.

$$J_k = P_k^{-1} = [P_{k|k-1} - P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1} H_k P_{k|k-1}]^{-1} \tag{1.13}$$
$$= (P_{k|k-1}^{-1} + H_k^T R_k^{-1} H_k) \tag{1.14}$$
$$= (J_{k|k-1} + H_k^T R_k^{-1} H_k) \tag{1.15}$$

La moyenne prédite se calcule de la même façon :

$$\hat{x}_{k|k-1} = F_k \hat{x}_{k-1} + B_k$$

La moyenne corrigée est obtenue en exprimant le gain de Kalman K_k en fonction de $J_{k|k-1}$:

$$\begin{aligned} K_k &= P_{k|k-1} H_k^T [H_k P_{k|k-1} H_k^T + R_k]^{-1} \\ &= [H_k R_k^{-1} H_k^T + P_{k|k-1}^{-1}]^{-1} H_k^T R_k^{-1} \end{aligned}$$

d'après le lemme d'inversion matricielle A.2. D'où

$$\begin{aligned} K_k &= [H_k R_k^{-1} H_k^T + P_{k|k-1}^{-1}]^{-1} H_k^T R_k^{-1} \\ &= J_k^{-1} H_k^T R_k^{-1} \end{aligned}$$

Finalement,

$$\hat{x}_k = \hat{x}_{k|k-1} + J_k^{-1} H_k^T R_k^{-1} [y_k - (H_k \hat{x}_{k|k-1} + D_k)]$$

D'après (1.12) et (1.15), on observe que dans le filtre de Kalman informationnel, on inverse essentiellement des matrices $d \times d$ (exception faite de la matrice de covariance de mesure R_k) contrairement au filtre de Kalman standard qui nécessite l'inversion de matrices $m \times m$, où m est la dimension de la mesure et d celle de l'état. Dans le cas où $m >> d$, la version informationnelle peut se révéler avantageuse.
D'autre part, le filtre de Kalman informationnel permet de traiter le cas d'une incertitude initiale P_0 infinie, qui correspond à une information initiale nulle i.e. $J_0 = 0$, là où l'écriture $P_0 = \infty$ n'aurait pas de sens numériquement dans le filtre de Kalman standard.

1.3.1.2 Le filtre de Kalman étendu (EKF)

On considère un système dynamique dont les fonctions d'évolution et d'observation sont non-linéaires.

$$\left\{ \begin{array}{rcl} x_k & = & f_k(x_{k-1}) + w_k \\ y_k & = & h_k(x_k) + v_k \end{array} \right. \tag{1.16}$$

w_k et v_k sont des bruits blancs gaussiens mutuellement indépendants et indépendant de x_0. On ne suppose plus que la condition initiale x_0 est gaussienne.
Dans le cas où les fonctions d'observations sont dérivables, on peut linéariser les équations de prédictions et de corrections respectivement autour de l'état corrigé $\hat{x}_{k-1}$ et autour de l'état prédit $\hat{x}_{k|k-1}$:

$$x_k \approx f_k(\hat{x}_{k-1}) + \nabla f_k\left(\hat{x}_{k-1}\right)\left(x_{k-1} - \hat{x}_{k-1}\right) + w_k$$
$$y_k \approx h(\hat{x}_{k|k-1}) + \nabla h_k\left(\hat{x}_{k|k-1}\right)\left(x_k - \hat{x}_{k|k-1}\right) + v_k$$

où ∇f_k et ∇h_k représentent les matrices jacobiennes de f_k et h_k. Soit $F_{k-1} = \nabla f_k(\hat{x}_{k-1})$ et $H_k = \nabla h_k(\hat{x}_{k|k-1})$. Le modèle linéarisé est donc :

$$\left\{ \begin{array}{rcl} x_k & = & F_k x_{k-1} + f_k(\hat{x}_{k-1}) - \nabla f_k\left(\hat{x}_{k-1}\right)\hat{x}_{k-1} + w_k \\ y_k & = & H_k x_k + h_k(\hat{x}_{k|k-1}) - \nabla h_k\left(\hat{x}_{k|k-1}\right)\hat{x}_{k|k-1} + v_k \end{array} \right. \tag{1.17}$$

L'algorithme du filtre de Kalman étendu est alors le suivant :

1. *Prédiction*

$$\left\{ \begin{array}{rcl} F_{k-1} & = & \nabla f_k(\hat{x}_{k-1}) \\ \hat{x}_{k|k-1} & = & f_k(\hat{x}_{k-1}) \\ P_{k|k-1} & = & F_{k-1} P_{k-1} {F_k}^{\mathrm{T}} + Q_k \end{array} \right. \tag{1.18}$$

2. *Correction*

$$\left\{ \begin{array}{rcl} H_k & = & \nabla h_k(\hat{x}_{k|k-1}) \\ K_k & = & P_{k|k-1} {H_k}^T \left(H_k P_{k|k-1} {H_k}^{\mathrm{T}} + R_k\right)^{-1} \\ \hat{x}_k & = & \hat{x}_{k|k-1} + K_k\left[y_k - h(\hat{x}_{k|k-1})\right] \\ P_k & = & \left(I - K_k H_k\right) P_{k|k-1} \end{array} \right. \tag{1.19}$$

Le filtre de Kalman étendu est extrêmement populaire en estimation de systèmes dynamiques. Il est notamment utilisé en pistage de cible [100], en navigation par mesures GPS [64] ou en robotique [74]. Il est avantageux par la simplicité de sa mise en oeuvre et sa faible complexité algorithmique. Cependant, lorsque les fonctions d'évolution et d'observation sont fortement non-linéaires ou lorsque l'initialisation est mauvaise, l'EKF est susceptible de diverger.

1.3.1.3 Le filtre de Kalman sans parfum (UKF)

Le filtre de Kalman sans parfum - Unscented Kalman Filter (UKF) en anglais - a été introduit par Julier et Uhlmann [63] afin d'éviter l'étape de linéarisation de l'EKF qui peut poser des problèmes numériques lorsque l'on souhaite calculer les jacobiennes du modèles. L'UKF fonctionne en calculant successivement la moyenne et la covariance a posteriori de l'état à l'aide d'un nombre fini d'échantillons, appelés sigma-points caractéristiques de la loi conditionnelle. L'algorithme suppose implicitement que cette loi est monomodale. Le calcul des sigma points se fait grâce à une transformation sans parfum (Unscented Transformation) qui permet de calculer les moments d'une variable aléatoire Y ayant subi une transformation non linéaire sous la forme $Y = g(X)$.

Transformation sans parfum (UT) Etant donné une variable aléatoire X de dimension d de moyenne $\bar{X}$ et matrice de covariance P_X. Soit $Y = g(X)$ une v.a. dont on veut calculer la moyenne et la covariance et g une fonction non linéaire. Le principe de la transformation UT est de choisir $2d+1$ points pondérés, caractéristiques de la distribution de X, de telle sorte que leurs moyenne et covariance valent respectivement $\bar{X}$ et P_X. Les sigma-points $\{\chi_i\}_i^{2d+1}$ et leur poids correspondants $\{\omega_i\}_i^{2d+1}$ sont choisis de manière déterministe selon le schéma suivant :

$$\begin{cases} \chi_0 &= \bar{X} \\ \chi_i &= \bar{X} + (\sqrt{(d+\kappa)P_X})_i, & i = 1, ..., d \\ \chi_i &= \bar{X} - (\sqrt{(d+\kappa)P_X})_i, & i = d+1, ..., 2d \\ \omega_0 &= \kappa/(d+\kappa) \\ \omega_i &= \omega_i = 1/2(d+\kappa), & i = 1, ..., 2d \end{cases} \tag{1.20}$$

$\left(\sqrt{(d+\kappa)P_X}\right)_i$ est la i-ème colonne de la racine carrée de la matrice $(d+\kappa)P_X$. Pour une distribution gaussienne $\kappa = 3 - d$, cette valeur est en général utilisée lorsqu'on ne connait pas la forme de la distribution de X.

Le calcul de la moyenne $\bar{Y}$ et de la covariance P_Y de Y se fait alors de la manière suivante :

$$\begin{cases} Y_i &= g(\chi_i), & i = 0, ..., 2d \\ \bar{Y} &\approx \sum_{i=0}^{2d} \omega_i Y_i \\ P_Y &\approx \sum_{i=0}^{2d} \omega_i (Y_i - \bar{Y})(Y_i - \bar{Y})^T \end{cases}$$

Julier et Uhlmann montrent que ces approximations de la moyenne et de la covariance sont correctes au troisième ordre [62].

L'algorithme UKF L'algorithme de filtrage récursif de l'UKF a été obtenu par Julier et Uhlmann en définissant l'état augmenté du système x_k^a, comme la concaténation de l'état x_k et des bruits w_k et v_k à l'instant k soit : $x_k^a = (x_k, w_k, v_k)^T$. x_k^a est donc un vecteur de dimension $d_a = 2d + m$, m étant la dimension du vecteur de mesure. L'UKF calcule donc à chaque instant

la moyenne conditionnelle $\bar{x}_k$ et la variance a posteriori P_k.

On pose $P_k^a = \mathrm{Cov}(x_k^a \mid y_{0:k}) = \begin{pmatrix} P_k & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & R \end{pmatrix}$

- *Initialisation*
  1. $\bar{x}_0 = \mathbb{E}(x_0)$
  2. $P_0 = \mathbb{E}\left[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T\right]$
  3. $P_0^a = \begin{pmatrix} P_0 & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & R \end{pmatrix}$

  Pour $k = 1 \dots n$
- *Calcul des sigma-points*
  1. $\chi_{k-1}^{a,i} = \left[\bar{x}_{k-1}^a \quad \bar{x}_{k-1}^a \pm \sqrt{(d_a + \kappa) P_{k-1}^{a,i}}\right] \qquad i = 1 \dots 2d_a$

     où $\sqrt{(d_a + \kappa) P_{k-1}^{a,i}}$ représente la $i^{ème}$ colonne de la décomposition de Cholesky de $(d_a + \kappa) P_{k-1}^{a,i}$.
- *Prédiction*
  1. $\chi_{k|k-1}^{x,i} = f_k\left(\chi_{k-1}^{x,i}, \chi_{k-1}^{w,i}\right)$
  2. $\bar{x}_{k|k-1} = \sum_{i=0}^{2d_a} \omega_i \chi_{k|k-1}^{x,i}$
  3. $P_{k|k-1} = \sum_{i=0}^{2d_a} \omega_i (\chi_{k|k-1}^{x,i} - \bar{x}_{k|k-1})(\chi_{k|k-1}^{x,i} - \bar{x}_{k|k-1})^T$
  4. $y_{k|k-1}^i = h_k\left(\chi_{k-1}^{x,i}, \chi_{k-1}^{v,i}\right)$
  5. $\bar{y}_{k|k-1} = \sum_{i=0}^{2d_a} \omega_i y_{k|k-1}^i$
- *Correction*
  1. $P_{k|k-1}^Y = \sum_{i=0}^{2d_a} \omega_i (y_{k|k-1}^i - \bar{y}_{k|k-1})(y_{k|k-1}^i - \bar{y}_{k|k-1})^T$
  2. $P_{k|k-1}^{XY} = \sum_{i=0}^{2d_a} \omega_i (\chi_{k|k-1}^{x,i} - \bar{x}_{k|k-1})(y_{k|k-1}^i - \bar{y}_{k|k-1})^T$
  3. $K_k = P_{k|k-1}^{XY} P_{k|k-1}^{Y,-1}$
  4. $\bar{x}_k = \bar{x}_{k|k-1} + K_k\left(Y_k - \bar{y}_{k|k-1}\right)$
  5. $P_k = P_{k|k-1} + K_k P_{k|k-1}^Y K_k^T$

Dans la plupart des applications, il a été observé que l'UKF permet d'estimer l'état du système avec plus de précision que l'EKF [63] avec un gain en robustesse appréciable. De plus, il ne nécessite pas de linéarisation du modèle, ce qui rend son implémentation plus simple que celle de l'EKF. Cependant son application se limite aux applications où la loi conditionnelle $p(x_k \mid y_{0:k})$ est monomodale.

1.3.2 Méthodes numériques

La plupart des systèmes physiques sont décrits par des modèles en temps continu. Par exemple, la dérive d'une centrale inertielle obéit à un système d'équations différentielles stochastiques (cf.3.3). Les systèmes dynamiques en temps continu consistent en un processus d'état caché $(x_t)_{t\geq 0}$ et un processus d'observation $(y_t)_{t\geq 0}$ à valeur respectivement dans $\mathbb{R}^d$ et $\mathbb{R}^m$, dont l'évolution est régie par le système d'équations suivant :

$$\left\{ \begin{array}{rcl} dx_t & = & b(x_t)dt + \sigma(x_t)dW_t \\ dy_t & = & h(x_t)dt + dV_t \end{array} \right. \tag{1.21}$$

où $(W_t)_{t\geq 0}$ et $(V_t)_{t\geq 0}$ sont des processus de Wiener standards mutuellement indépendants avec $(V_t)_{t\geq 0}$ indépendant de $(x_t)_{t\geq 0}$. b et h sont des fonctions à valeurs dans $\mathbb{R}^d$ et σ une application de $\mathbb{R}^d \mapsto \mathbb{R}^d \times \mathbb{R}^d$. Soit $\mathcal{Y}_t = \sigma(y_s,\, 0 \leq s \leq t)$ la sigma-algèbre engendrée par les observations. Le problème du filtrage en temps continu consiste à déterminer la densité conditionnelle de x_t sachant $\mathcal{Y}_t$ notée p_t, telle que pour toute fonction ϕ mesurable bornée

$$\mathbb{E}(\phi(x_t)|\mathcal{Y}_t) = \int_{\mathbb{R}^d} \phi(x) p_t(x) dx \tag{1.22}$$

Sous certaines hypothèses sur les fonctions b et σ, Zakai [125] a montré que le filtre optimal p_t est l'unique solution d'une équation différentielle stochastique, l'équation de Zakai.

$$dp_t = L^T p_t dt + p_t h^T R^{-1} dy_t \tag{1.23}$$

où L est l'opérateur infinitésimal associé au processus x_t, défini par

$$L = \frac{1}{2} \sum_{i,j=1}^{m} a_{i,j}(.) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i,j=1}^{m} b_i(.) \frac{\partial}{\partial x_i}$$

où $a = (a_{i,j}) = \sigma\sigma^T$. En pratique, les observations sont disponibles à des instants discrets $t_k, k = 0, 1, \ldots$ de sorte qu'il est nécessaire de discrétiser l'équation de Zakai en utilisant un schéma d'Euler par exemple [104]. Reste à discrétiser l'espace d'état par maillage de manière à résoudre numériquement (1.23) par différences finies. Le maillage peut se faire à pas constant [71] ou avec un pas variable ce qui permet de mailler finement les zones de grande probabilité [14] et mailler grossièrement les zones de faible probabilité : l'avantage du maillage adaptatif est qu'il permet généralement des gains en temps de calcul et en mémoire. Les méthodes de maillage sont efficaces en dimension $d \leq 3$ mais au delà, le temps de calcul est souvent prohibitif.

Conclusion du Chapitre 1

Nous avons introduit et défini le problème du filtrage et son application à l'estimation de l'état d'un système. Le cadre théorique nous a conduit à étudier les méthodes analytiques et numérique en filtrage. Les principales méthodes analytiques sont essentiellement le filtre de Kalman et ses variantes. Lorsque le système à espace d'état est faiblement non-linéaire, le filtre de Kalman étendu (EKF) donne généralement des résultats satisfaisants. Le filtre de Kalman sans parfum (UKF) améliore la précision de l'estimation de l'état par rapport à l'EKF tout en se montrant plus robuste. En revanche lorsque les non-linéarités sont fortes, ces méthodes analytiques sont peu fiables. Une possibilité est de recourir aux méthodes numériques, basées sur une résolution d'une équation aux dérivées partielles, par des schémas de discrétisation. Cependant, au delà de la dimension 3, le temps de calcul augmente exponentiellement avec la dimension de l'espace d'état, ce qui rend leur utilisation limitée.
Les méthodes de Monte Carlo sont a priori plus appropriées car moins sensibles à la dimension de l'espace d'état et indifférents aux non-linéarités et au caractère gaussien des bruits de dynamique et de mesure. En outre, elles sont adaptées aux lois conditionnelles multimodales.

Chapitre 2

Méthodes particulaires pour le filtrage non linéaire

Introduction

L'objectif de ce chapitre est d'introduire les briques de base du filtrage particulaire et d'exposer quelques algorithmes utilisés notamment pour le recalage de la navigation inertielle. Nous introduisons brièvement les méthodes de Monte Carlo usuelles avant d'illustrer comment leur utilisation dans un cadre séquentiel permet d'aboutir au filtre particulaire. Nous mettons en évidence les limites du filtre particulaire classique en présence de multimodalités. Dans ce contexte, nous présentons une méthodologie générale adaptée à l'estimation des lois a posteriori $p(x_k|y_{0:k})$ multimodales.

2.1 Méthodes de Monte Carlo

2.1.1 Echantillonnage Monte Carlo

Soit X une variable aléatoire à valeurs dans $\mathbb{R}^d$ de densité de probabilité $p(x)$. Soit ϕ une fonction bornée de $\mathbb{R}^d$ dans $\mathbb{R}$. On souhaite approcher numériquement l'espérance de $\phi(X)$ définie par :

$$\mathbb{E}\left(\phi(X)\right) = \int \phi(x)p(x)dx = \bar{\phi} \tag{2.1}$$

Soit X^i, $i = 1, 2, \ldots, N$ une suite de variable aléatoires i.i.d de même loi que X. Alors, d'après la loi des grands nombres, la moyenne empirique converge presque surement (p.s.) vers $\mathbb{E}(\phi(X))$, i.e.

$$\bar{\phi}^N \triangleq \frac{1}{N}\sum_{i=1}^{N}\phi(X^i) \xrightarrow[N\to+\infty]{} \mathbb{E}(\phi(X)) \tag{2.2}$$

De plus, la variance de l'estimateur $\bar{\phi}^N(X)$ vaut :

$$\mathrm{Var}\left(\bar{\phi}^N\right) = \frac{\sigma_\phi^2}{N}$$

avec

$$\sigma_\phi^2 = \int (\phi(x) - \bar{\phi})^2 p(x) dx$$

Par ailleurs, le théorème central limite caractérise l'erreur asymptotique de l'estimateur $\bar{\phi}^N$. En effet, la variable aléatoire $\frac{\sqrt{N}}{\sigma_\phi}\left(\bar{\phi}^N - \bar{\phi}\right)$ converge en loi vers une loi normale centrée réduite.

2.1.2 Echantillonnage pondéré

L'échantillonnage Monte Carlo pondéré (*importance sampling* en anglais) consiste à calculer l'espérance définie par l'équation (2.1) en simulant un échantillon issu d'une loi de densité q, dite loi d'importance ou loi instrumentale, et en pondérant cet échantillon de manière adéquate. Cette méthode est utile lorsque l'on souhaite réduire la variance de l'estimateur Monte Carlo classique ou quand échantillonner p est difficile (voir l'exemple ci-après). La densité de probabilité q doit vérifier les hypothèses suivantes pour toute fonction mesurable bornée ϕ :

1. Le support de $x \mapsto \phi(x)p(x)$ est inclus dans le support de q
2. $\int \phi^2(x)\frac{p(x)}{q(x)}dx < \infty$

L'estimateur obtenu par échantillonnage pondéré est obtenu en remarquant que :

$$\mathbb{E}_p\left(\phi(X)\right) = \int \phi(x)p(x)dx = \int \phi(x)\frac{p(x)}{q(x)}q(x)dx = \mathbb{E}_q\left(\phi(X)\frac{p(X)}{q(X)}\right) \tag{2.3}$$

Soit x^i, $i = 1, 2, \ldots, N$ un échantillon issu de q. Alors

$$\mathbb{E}\left(\phi(X)\right) \approx \bar{\phi}_{IS}^N = \frac{1}{N}\sum_{i=1}^N \phi(x^i)\frac{p(x^i)}{q(x^i)} \tag{2.4}$$

Lorsque la densité p n'est connue qu'à une constante de normalisation c près, i.e. $p(x) = c\tilde{p}(x)$ avec $c = \int p(x)dx$, il est possible de définir l'estimateur $\bar{\phi}_{IS}^N$ pondéré de la manière suivante :

$$\bar{\phi}_{IS}^N \approx \frac{\sum_{i=1}^N \tilde{\omega}^i \phi(x^i)}{\sum_{i=1}^N \tilde{\omega}^i} \tag{2.5}$$

avec $\tilde{\omega}^i = \frac{\tilde{p}(x^i)}{q(x^i)}$. Dans ce cas, la constante de normalisation c est approchée par $\frac{1}{N}\sum_{i=1}^N \tilde{\omega}^i$.

Exemple (Inférence Bayésienne) : Afin d'illustrer l'intérêt de l'échantillonnage pondéré, supposons que le paramètre inconnu $X \sim q$ où q est la densité d'une loi que l'on sait échantillonner. On observe X à travers une mesure $y = h(X) + v$ où v est un bruit de densité g_v et h une fonction non-linéaire. On souhaite estimer une quantité $\mathbb{E}(\phi(X)|Y) = \int \phi(x)p(x|y)dx$ où ϕ est une fonction bornée. Dans la plupart des cas, échantillonner $p = p(x|y)$ n'est pas aisé donc on ne peut recourir à l'estimateur Monte Carlo standard. En revanche, en prenant comme densité d'importance la densité a priori q et en constatant que $\frac{p(x)}{q(x)} \propto p(y|x) = g_v(y - h(x))$ en vertu de la loi de Bayes, on a accès à l'estimateur par échantillonnage pondéré $\bar{\phi}_{IS}^N = \frac{\sum_{i=1}^N \phi(x^i) g_v(y-h(x^i))}{\sum_{i=1}^N g_v(y-h(x^i))}$ où x^i, $i = 1, \ldots, N$ est un N-échantillon issu de q.

Les estimateurs (2.4) et (2.5) sont asymptotiquement non biaisés et convergent presque surement vers $\mathbb{E}(\phi(X))$ (voir, par exemple [48]). De plus,

$$\mathrm{Var}\left(\bar{\phi}_{IS}^N\right) = \frac{1}{N} \int \phi^2(x) \frac{p^2(x)}{q(x)}\, dx - \frac{1}{N} \left(\int \phi(x) p(x)\, dx \right)^2$$

De même que pour l'estimateur Monte Carlo standard, il existe un théorème central limite qui permet de caractériser la loi limite de l'erreur d'estimation. En effet, lorsque N tend vers l'infini la v.a. $\sqrt{N}\left(\bar{\phi}_{IS}^N - \bar{\phi}\right)$ converge en loi vers une loi normale centrée de variance $\sigma_{IS}^2 = \int \left(\phi(x) - \bar{\phi}\right)^2 \frac{p^2(x)}{q(x)} dx$

2.1.3 Échantillonnage par acceptation-rejet

Supposons que la densité p s'exprime comme le produit d'une densité q et d'une fonction g que l'on sait évaluer en tout x de $\mathbb{R}^d$.

$$p(x) = \frac{g(x)q(x)}{\int_{\mathbb{R}^d} g(x)q(x)dx}$$

La méthode d'acceptation permet [34] d'obtenir un échantillon suivant la loi de densité p lorsqu'on sait simuler la loi de densité q.
Pour cela, on suppose seulement que $M = \sup_{x \in \mathbb{R}^d} g(x) < \infty$. Afin d'obtenir un échantillon $x^i \sim p$, $i = 1, \ldots, N$, on applique l'algorithme suivant.

1. $i = 1$
2. Générer $\xi \sim q$ et $u \sim \mathcal{U}([0,1])$ où $\mathcal{U}([a,b])$ désigne la loi uniforme sur l'intervalle $[a,b]$, a et b étant des réels.
3. Si $Mu \leq g(\xi)$, poser $x^i = \xi$ et $i = i + 1$.
4. Aller en (2) tant que $i \leq N$.

L'efficacité de la méthode d'acceptation-rejet dépend essentiellement du recouvrement de g et q. En effet, la probabilité d'acceptation vaut $p_a = \frac{\int_{\mathbb{R}^d} g(x)q(x)dx}{M}$ et est d'autant plus grande que le recouvrement entre q et g est important. Cette méthode est donc potentiellement couteuse en temps de calcul puisqu'on risque de rejeter beaucoup de variables aléatoires lors de l'étape 3.

2.2 Filtrage particulaire

Soit un système dynamique non-linéaire défini par l'ensemble d'équations suivant :

$$\begin{cases} x_k &= f_k\left(x_{k-1}, w_k\right) \\ y_k &= h_k(x_k, v_k) \end{cases} \tag{2.6}$$

- x_k est le vecteur d'état de dimension d
- y_k est le vecteur d'observations de dimension m
- w_k et v_k sont des bruits blancs mutuellement indépendants non nécessairement gaussiens qui admettent des densités de probabilités. $(w_k)_{k\geq 0}$ et $(v_k)_{k\geq 0}$ sont indépendantes de la condition initiale x_0. x_0 suit une loi de densité μ_0.

Le filtre particulaire - également appelé bootstrap filter [50], Monte Carlo filter [68] ou encore condensation filter [79] - approche la densité conditionnelle $p(x_k \mid y_{0:k})$ par un ensemble d'échantillons pondérés appelés particules. Ces particules évoluent selon la dynamique de l'état et sont pondérés en fonction de leur adéquation avec l'observation courante grâce à la fonction de vraisemblance (cf. section 1.2.2). Dans toute la suite on supposera que :

- on sait générer des échantillons selon la densité de transition $p(x_k|x_{k-1})$,
- il est possible d'évaluer la fonction de vraisemblance $g_k(x_k) = p(y_k|x_k)$ en tout point x_k de l'espace d'état.

2.2.1 L'échantillonnage séquentiel pondéré

L'algorithme d'échantillonnage séquentiel pondéré [52] (*Sequential Importance Sampling (SIS)* en anglais) est fondé sur une approximation de la densité conditionnelle $p_{0:k} = p(x_{0:k}|y_{0:k})$ par la méthode d'échantillonnage pondéré abordée dans la section 2.1.2. On notera $x_{0:k} = (x_0, \ldots, x_k)^T$ la trajectoire de l'état de l'instant initial à l'instant k.
Dans le cas général, échantillonner directement $p_{0:k}$ est difficile et/ou couteux. En effet, la factorisation

$$p_{0:k}(x_{0:k}|y_{0:k}) = \frac{p(y_{0:k}|x_{0:k})p(x_{0:k})}{p(y_{0:k})}$$

suggère que si échantillonner selon la densité jointe $p(x_{0:k})$ est généralement possible, cela nécessiterait une méthode de type acceptation-rejet pour obtenir un échantillon suivant la loi $p_{0:k}$.

Soit $q(x_{0:k}|y_{0:k})$ une densité instrumentale pour laquelle la simulation directe est aisée. Soit $x^i_{0:k}$, $i = 1, \ldots, N$ un échantillon issu de q. On définit les poids non normalisés par $\tilde{\omega}^i_k = \frac{p(x^i_{0:k}|y_{0:k})}{q(x^i_{0:k}|y_{0:k})}$ et les poids normalisés $\omega^i_k = \frac{\tilde{\omega}^i_k}{\sum_{i=1}^N \tilde{\omega}^i_k}$. Alors,

$$p(x_{0:k}|y_{0:k}) \approx \sum_{i=1}^{N} \omega^i_k \delta_{x^i_{0:k}} \tag{2.7}$$

où δ_x désigne la masse de Dirac centrée au point x. Une approximation du filtre à l'instant k est

alors obtenue par marginalisation de $p(x_{0:k}|y_{0:k})$:

$$p(x_k|y_{0:k}) \approx \sum_{i=1}^{N} \omega_k^i \delta_{x_k^i} \tag{2.8}$$

De manière générale échantillonner $q(x_{0:k}|y_{0:k})$ a une complexité qui augmente au moins linéairement au cours du temps car la dimension du vecteur $x_{0:k}$ est linéaire en k. Il est donc intéressant d'avoir une méthode d'échantillonnage séquentielle.
Pour obtenir une formulation récursive de (2.7), on exprime $p(x_{0:k}|y_{0:k})$ en fonction de $p(x_{0:k-1}|y_{0:k-1})$.

$$p(x_{0:k}|y_{0:k}) = \frac{p(x_{0:k-1}|y_{0:k-1})p(x_k|x_{0:k-1}, y_{0:k-1})p(y_k|x_k)}{p(y_k|y_{0:k-1})}$$

Etant donné que x_k est Markovien, $p(x_k|x_{0:k-1}, y_{0:k-1}) = p(x_k|x_{k-1})$. D'où

$$p(x_{0:k}|y_{0:k}) = \frac{p(x_{0:k-1}|y_{0:k-1})p(x_k|x_{k-1})p(y_k|x_k)}{p(y_k|y_{0:k-1})} \tag{2.9}$$

Afin d'obtenir un algorithme séquentiel du filtre, considérons une densité d'importance q sous la forme :

$$q(x_{0:k}|y_{0:k}) = q(x_{0:k-1}|y_{0:k-1})q(x_k|x_{0:k-1}, y_{0:k}) \tag{2.10}$$

On observe que si l'on dispose d'un échantillon pondéré $\left\{\omega_{k-1}^i, x_{0:k-1}^i\right\}_{i=1}^N$ de taille N, il est possible d'obtenir un échantillon $\{\omega_k^i, x_{0:k}^i\}_{i=1}^N$ en simulant x_k^i selon $q(x_k|x_{0:k-1}^i, y_{0:k})$ et en concaténant cet échantillon au vecteur $x_{0:k-1}^i$.
Le calcul des poids ω_k^i se fait selon

$$\omega_k^i \propto \frac{p(x_k^i|x_{k-1}^i)p(y_k|x_k^i)p(x_{0:k-1}^i|y_{0:k-1})}{q(x_k^i|x_{0:k-1}^i, y_{0:k})q(x_{0:k-1}^i|y_{0:k-1})} \tag{2.11}$$

$$= \omega_{k-1}^i \frac{p(x_k^i|x_{k-1}^i)p(y_k|x_k^i)}{q(x_k^i|x_{0:k-1}^i, y_{0:k})} \tag{2.12}$$

Notons qu'en filtrage, le but est d'estimer la loi a posteriori $p(x_k^i|y_{0:k})$. On a donc uniquement besoin de l'ensemble des particules à l'instant précédent $\left\{x_{k-1}^i\right\}_{i=1}^N$. L'algorithme d'échantillonnage pondéré séquentiel peut alors être simplifié en choisissant une densité d'importance q telle que $q(x_k|x_{0:k-1}, y_{0:k}) = q(x_k|x_{k-1}, y_k)$. Dans ce cas,

$$\omega_k^i \propto \omega_{k-1}^i \frac{p(x_k^i|x_{k-1}^i)p(y_k|x_k^i)}{q(x_k^i|x_{k-1}^i, y_k)} \tag{2.13}$$

L'algorithme SIS se résume ainsi :

Algorithme 1 : SIS

pour $i = 1, \ldots, N$ **faire**

– tirer $x^i \sim q(x_0|y_0)$
– calculer $\omega_0^i \propto \frac{\mu_0(x_0^i)p(y_0|x_0^i)}{q(x_0^i|y_0)}$
– normaliser les poids ω_0^i
fin pour
pour $k = 1, \ldots, n$ **faire**
pour $i = 1, \ldots, N$ **faire**
– *[Prédiction]* tirer $x_k^i \sim q(x_k^i|x_{k-1}^i, y_k)$
– *[Correction]* calculer les poids $\omega_k^i \propto \omega_{k-1}^i \frac{p(x_k^i|x_{k-1}^i)p(y_k|x_k^i)}{q(x_k^i|x_{k-1}^i, y_k)}$
– normaliser les poids ω_k^i
fin pour
fin pour

Les estimées de la moyenne et de la covariance de l'état x_k sont alors respectivement :

$$\mathbb{E}(x_k|y_{0:k}) \approx \hat{x}_k = \sum_{i=1}^{N} \omega_k^i x_k^i \tag{2.14}$$

$$\mathbb{E}\left[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T |y_{0:k}\right] \approx \hat{P}_k = \sum_{i=1}^{N} \omega_k^i \left(x_k^i - \hat{x}_k\right)\left(x_k^i - \hat{x}_k\right)^T \tag{2.15}$$

2.2.2 Dégénérescence des poids

L'algorithme d'échantillonnage séquentiel pondéré est la brique de base du filtrage particulaire. Cependant, il souffre d'une limitation importante : après quelques itérations, toutes les particules sauf une ont un poids proche de zéro. L'estimation de la loi a posteriori est donc réduite à un dirac, ce qui entraine la divergence du filtre. Ce phénomène est caractérisé par une augmentation dans le temps de la variance des poids normalisés $\{\omega_k^i\}_{i=1}^N$. Kong et al. ont notamment montré que lorsque la loi d'importance de densité $q_{0:k} = q(x_{0:k}|y_{0:k})$ est différente de la loi cible de densité $p(x_{0:k}|y_{0:k})$, la variance des poids ne peut qu'augmenter [69]. Une illustration de la dégénérescence du système de particules est donnée par la figure 2.1. On considère ici l'évolution des particules d'un filtre SIS à 100 particules estimant le modèle linéaire gaussien suivant :

$$\begin{cases} x_{k+1} &= x_k + w_k \\ y_k &= x_k + v_k \end{cases} \tag{2.16}$$

avec $w_k \sim \mathcal{N}(0, 0.002^2)$, $v_k \sim \mathcal{N}(0, 0.1^2)$ et $x_0 \sim \mathcal{N}(0, 1)$.

La position des particules de poids significatifs[1] y est affichée en fonction du temps. On observe que le système perd rapidement en diversité et vers la fin de la trajectoire, très peu de particules contribuent à l'estimation.

Pour retarder ce phénomène il est nécessaire de choisir une densité instrumentale $q_{0:k}$ aussi proche de la loi conditionnelle $p_{0:k}$ que possible. Malgré cela, il est indispensable de recourir à une étape dite de ré-échantillonnage afin de conserver un système de particules non dégénéré.

1. càd dont les poids sont supérieurs à 10^{-6}

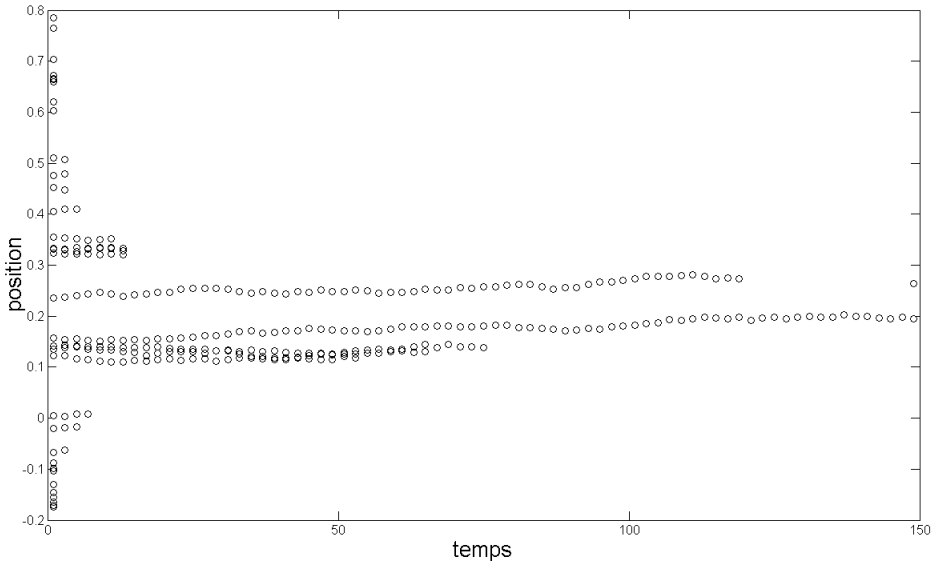


FIGURE 2.1 – Illustration du phénomène de dégénérescence des poids. Abscisse : temps discret, ordonnée : position des particules

2.2.3 Ré-échantillonnage des particules

Le ré-échantillonnage a pour objectif de dupliquer les particules dont le poids est suffisamment fort et d'éliminer les particules dont le poids est trop faible afin de ne conserver que les particules les plus significatives. Kong et al. [69] préconisent de contrôler l'évolution de la taille effective de l'échantillon (*effective sample size* en anglais) notée N_{eff} qui est un indicateur de la dégénerescence du filtre. Ils montrent que N_{eff} est relié à la variance des poids normalisés à l'instants k, $\{\omega_k^i\}_{i=1}^N$ via l'égalité :

$$N_{eff} = \frac{N}{1 + \mathrm{Var}_{p_{0:k}}\left(\omega_k(x_{0:k})\right)} \tag{2.17}$$

Ainsi, plus la variance des poids est élevée, plus la taille effective de l'échantillon est faible. Par ailleurs, N_{eff} est relié à la divergence du χ^2 entre $p_{0:k}$ et $q_{0:k}$ via l'égalité :

$$\frac{N}{N_{eff}} - 1 = \chi^2(p_{0:k}, q_{0:k}) = \int \frac{\left(p_{0:k}(x_{0:k}) - q_{0:k}(x_{0:k})\right)^2}{q_{0:k}(x_{0:k})}\, dx_{0:k} \tag{2.18}$$

On peut remarquer que $1 \leq N_{eff} \leq N$. Autrement dit, N_{eff} est un indicateur du nombre d'échantillons qui contribuent significativement à l'estimation de la loi $p(x_k|y_{0:k})$. Dans la pratique, lorsque N_{eff} est inférieur à un seuil prédéfini N_{th}, on applique le ré-échantillonnage. Pour cela, l'approximation suivante de N_{eff} est souvent utilisée :

$$N_{eff} \approx \frac{1}{\sum_{i=1}^N (\omega_k^i)^2} \tag{2.19}$$

Il existe un second critère relatif à l'entropie des poids normalisés, introduit par Pham [102] dans le cadre d'assimilation de données, qui s'exprime sous la forme suivante :

$$S_\omega = \log(N) + \sum_{i=1}^{N} \omega_k^i \log\left(\omega_k^i\right) \tag{2.20}$$

Une observation intéressante est que :

$$S_\omega \approx \log(N) + D_{KL}\left(p_{0:k} \parallel q_{0:k}\right)$$

où $D_{KL}\left(p_{0:k} \parallel q_{0:k}\right)$ désigne la divergence de Kullback-Leibler de $q_{0:k}$ par rapport à $p_{0:k}$: c'est une mesure de l'information perdue lorsque $q_{0:k}$ est utilisée pour approcher $p_{0:k}$ [13]. De façon générale, la divergence de Kullback-Leibler d'une loi de densité g par rapport à une loi de densité f est définie par :

$$D_{KL}\left(f \parallel g\right) = \int f(x) \log \frac{f(x)}{g(x)} dx \tag{2.21}$$

S_ω est nul lorsque tous les poids sont égaux à $\frac{1}{N}$ et $S_\omega = \log(N)$ si l'un des poids vaut 1 . Une heuristique est de ré-échantillonner le système de particules lorsque l'entropie S_w dépasse un certain seuil S_{th}.

Ré-échantillonner consiste à remplacer le système de particules $\{\omega_k^i, x_k^i\}_{i=1}^N$ par $\{\omega_k'^i, x_k'^i\}_{i=1}^N$ de sorte que la variance des poids après ré-échantillonnage soit réduite et que $p_k'^N = \sum_{i=1}^N \omega_k'^i \delta_{x_k'^i}$ soit une "bonne approximation" de $\hat{p}_k = \sum_{i=1}^N \omega_k^i \delta_{x_k^i}$. Les schémas qui correspondent à une bonne approximation supposent que $p_k'^N$ soit un estimateur non-biaisé de p_k^N. De manière générale, le ré-échantillonnage vérifie les conditions suivantes :

1. $\omega_k'^i = \frac{1}{N}$
2. $\mathbb{E}\left[\sum_{i=1}^N \mathbb{1}\left\{x_k'^i = x_k^j\right\} | x_{0:k}, y_{0:k}\right] = N\omega_k^j$

Les deux conditions assurent le caractère non-biaisé de l'approximation après ré-échantillonnage [76]. La seconde signifie qu'en moyenne une particule sera dupliquée d'autant plus qu'elle possède un poids fort.

La méthode de redistribution la plus courante est le **ré-échantillonnage multinomial**. Elle consiste à tirer N particules $\tilde{x}_k^i$ avec probabilités ω_k^i, $i = 1, \ldots, N$ et à leur affecter un nouveau poids égal à $\omega_k'^i = \frac{1}{N}$.

A titre d'exemple considérons un échantillon pondéré $x^1, x^2, \ldots, x^{10}$ de taille 10 dont les poids normalisés respectifs sont $\omega^1 = 0.02$, $\omega^2 = 0.3$, $\omega^3 = 0.15$, $\omega^4 = 0.07$, $\omega^5 = 0.09$, $\omega^6 = 0.07$, $\omega^7 = 0.1$, $\omega^8 = 0.12$, $\omega^9 = 0.05$ et $\omega^{10} = 0.03$. Le tirage multinomial peut se faire de la façon suivante :

1. Tirer N va indépendantes $u_i \sim \mathcal{U}([0,1])$, $i = 1, \ldots, N$
2. Pour $i = 1, \ldots, N$, si $u^i \in \left[\sum_{i=1}^{j-1} \omega_k^i, \sum_{i=1}^{j} \omega_k^i\right]$, sélectionner la particule d'indice j, i.e. poser $x_k'^i = x_k^j$.

Dans l'exemple précédent, un tirage multinomial nous donne l'ensemble de particules ré-échantillonnées $\{x^2, x^2, x^3, x^3, x^9, x^2, x^3, x^{10}, x^{10}, x^2\}$ où les particules x^2 et x^3 sont dupliquées tandis que x^1 et

x^6 sont éliminées.
Il est important de noter que le ré-échantillonnage augmente la variance de l'estimation. Il existe tout de même des schémas de redistribution qui induisent une variance plus faible que celle du ré-échantillonnage multinomial :

Ré-échantillonnage résiduel D'après [76], cette méthode permet de réduire efficacement la variance du ré-échantillonnage multinomial. L'un des problèmes du ré-échantillonnage multinomial est que le nombre de copies de la particule x_k^i peut varier fortement pour deux réalisations de l'algorithme de ré-échantillonnage. Pour limiter cet effet, le ré-échantillonnage résiduel attribue d'office $N^i = \lfloor N\omega_k^i \rfloor$ copies de l'échantillon x_k^i pour $i = 1, \ldots, N$. Il reste à allouer le nombre de particules restant $\bar{N} = N - \sum_{i=1}^N N^i$ afin de conserver un nombre constant de particules. Ces $\bar{N}$ particules sont alors échantillonnées selon un tirage multinomial de probabilités données par les poids résiduels $\{\bar{\omega}_k^i\}_{i=1}^N$ définis par :

$$\bar{\omega}_k^i = \frac{N\omega_k^i - N^i}{\bar{N}} \tag{2.22}$$

Afin d'illustrer notre propos, revenons à l'exemple précédent du ré-échantillonnage de $\{x^i, \omega^i\}_{i=1}^1 0$. Dans la première passe, chaque particule x^i est dupliquée $\lfloor N\omega^i \rfloor$ fois, ce qui donne l'ensemble $\{x^2, x^2, x^2, x^3, x^7, x^8\}$. Il reste maintenant à tirer 4 échantillons selon un tirage multinomial de probabilités $\bar{\omega}^1 = 0.05$, $\bar{\omega}^2 = 0$, $\bar{\omega}^3 = 0.125$, $\bar{\omega}^4 = 0.175$, $\bar{\omega}^5 = 0.225$, $\bar{\omega}^6 = 0.175$, $\bar{\omega}^7 = 0$, $\bar{\omega}^8 = 0.05$, $\bar{\omega}^9 = 0.125$, $\bar{\omega}^1 0 = 0.075$. Un tirage donne les particules additionnelles $\{x^6, x^4, x^5, x^3\}$. Un réalisation du ré-échantillonnage stratifié fournit donc le système de particules $\{x^2, x^2, x^2, x^3, x^7, x^8, x^6, x^4, x^5, x^3\}$.

Ré-échantillonnage stratifié L'intervalle $[0, 1]$ est partitionné en N sous-intervalles disjoints sous la forme $[0, 1] = \cup_{i=1}^N \left[\frac{i-1}{N}, \frac{i}{N}\right]$. Ensuite, N variables U_i, $i = 1, \ldots, N$ sont tirées selon une loi uniforme $\mathcal{U}\left(\frac{i-1}{N}, \frac{i}{N}\right)$. A chaque variable U_i est associé l'inverse de la fonction de répartition de la loi discrète $\sum_{i=1}^N \omega_k^i \delta_{x_k^i}$. Plus précisément, si $U_i \in \left[\sum_{i=1}^{j-1} \omega_k^i, \sum_{i=1}^j \omega_k^i\right]$, on sélectionne la particule d'indice j, i.e. $x_k'^i = x_k^j$. La variance induite par le ré-échantillonnage stratifié est également inférieure à celle du ré-échantillonnage multinomial [35].

Ré-échantillonnage systématique Cette méthode est similaire au ré-échantillonnage stratifié excepté que le choix des U_i se fait selon : $U_i = \frac{i-1}{N} + U$ où $U \sim \mathcal{U}\left(0, \frac{1}{N}\right)$. En revanche, les particules sélectionnées $\tilde{x}_k^i$ ne sont plus indépendantes conditionnellement aux poids $\{\omega_k^i\}_{i=1}^N$.

2.2.4 L'algorithme Sequential Importance Resampling (SIR)

L'ajout de l'étape de ré-échantillonnage dans l'algorithme SIS permet d'obtenir l'algorithme SIR [107] qui évite le phénomène de dégénérescence des poids et permet d'obtenir un filtre stable. Lorsque le ré-échantillonnage intervient à chaque pas de temps, le SIR correspond au *bootstrap filter* de Gordon, Salmond et Smith [50].

Algorithme 2 : SIR

pour $i = 1, \ldots, N$ **faire**
- tirer $x^i \sim q(x_0|y_0)$
- calculer $\omega_0^i \propto \frac{\mu_0(x_0^i)p(y_0|x_0^i)}{q(x_0^i|y_0)}$
- normaliser les poids ω_0^i

fin pour
pour $k = 1, \ldots, n$ **faire**
- *[Prédiction]* tirer $x_k^i \sim q(x_k^i|x_{k-1}^i, y_k)$, $i = 1, \ldots, N$
- *[Correction]* calculer les poids $\omega_k^i \propto \omega_{k-1}^i \frac{p(x_k^i|x_{k-1}^i)p(y_k|x_k^i)}{q(x_k^i|x_{k-1}^i, y_k)}$, $i = 1, \ldots, N$
- normaliser les poids ω_k^i, $i = 1, \ldots, N$
- Calculer N_{eff} d'après (2.19)
- *[Ré-échantillonnage]* Si $N_{eff} < N_{th}$, ré-échantillonner les particules et poser $\omega_k^i = \frac{1}{N}$, $i = 1, \ldots, N$

fin pour

Bien que le ré-échantillonnage permette de garantir la survie du système de particules, il introduit plusieurs effets indésirables. La duplication des particules de poids fort a tendance à appauvrir le système de particules, ce qui peut être problématique lorsque le bruit de dynamique est faible. Dans ce cas, lorsque la loi d'importance utilisée dans l'étape de prédiction est le noyau de transition $p(x_k|x_{k-1})$, les particules ont tendance à se concentrer autour de quelques zones resserrées provoquant une perte de diversité dans le système de particules (phénomène connu sous le nom de *sample impoverishment* dans la littérature). Cela augmente le risque qu'aucune des particules ne soit cohérente avec l'observation y_k courante, ce qui à terme peut provoquer la divergence du filtre. Pour limiter ce phénomène, une méthode peut-être de choisir une loi d'importance différente du noyau de transition ou de recourir aux méthodes de régularisation (cf. section 2.3).

Choix de la densité d'importance Le choix d'une densité d'importance appropriée permet de concevoir des filtres plus robustes. Dans la classe des algorithmes d'échantillonnage pondéré séquentiel, la densité d'importance $q_{opt} = p(x_k|x_{k-1}, y_k) = \frac{p(y_k|x_k)p(x_k|x_{k-1})}{p(y_k|x_{k-1})}$ est optimale. On peut montrer que la variance des poids normalisés correspondant ω_k^i est nulle, conditionnellement à x_{k-1} [37]. De manière équivalente, ce choix maximise la taille effective de l'échantillon N_{eff} introduite précédemment via l'équation (2.17) [37].
Malheureusement, échantillonner suivant cette loi n'est aisé que pour certaines classes de modèles. Le choix le plus simple est d'utiliser la densité a priori $q = p(x_k|x_{k-1})$. Échantillonner selon $p(x_k|x_{k-1})$ est immédiat pour les modèles que nous considérons mais cette loi ne prend pas en compte l'observation courante et peut déplacer les particules dans des zones de faible vraisemblance. Afin d'exploiter l'information apportée par la mesure y_k, il peut être souhaitable de recourir à des approximations de q_{opt}.

Parmi les méthodes d'approximations de la loi d'importance optimale, on peut citer le filtre particulaire auxiliaire [105] ainsi que les filtres basés sur une approximation gaussienne de q_{opt}

via l'utilisation d'un filtre de Kalman étendu [37] ou un filtre de Kalman sans parfum [117]. Ce dernier filtre est connu sous le nom d'*Unscented Particle Filter*. Une autre solution consiste à rechercher une approximation π de q_{opt} dans une classe de lois paramétrées. Le choix de π se fait alors en minimisant la divergence du chi-deux entre la densité cible q^{opt} et π ou bien la divergence de Kullback-Leibler (2.21) de π à q^{opt} [23].

2.2.5 Convergence des filtres particulaires

La littérature sur les méthodes Monte Carlo séquentielle est riche en résultats sur la convergence des filtres particulaires lorsque $N \to +\infty$. On peut citer notamment [25],[24], [26] ou [32]. On rappelle ci-dessous un résultat établi pour l'algorithme SIR.

Théorème 1 (Convergence faible [26]). *Soit ϕ une fonction continue bornée de $\mathbb{R}^d$. On suppose que la fonction de vraisemblance $x_k \mapsto g_k(x_k) = p(y_k|x_k)$ est bornée. Alors,*

$$\lim_{N\to+\infty} \sum_{i=1}^{N} \omega_k^i \phi(x_k^i) = \int \phi(x_k) p(x_k|y_{0:k}) dx_k \tag{2.23}$$

presque-surement.

Un autre résultat concernant la convergence en norme $\mathbb{L}^p$ s'énonce comme suit :

Théorème 2 (Convergence $\mathbb{L}^p$ [32]). *Sous les hypothèses précédentes, pour tout $p \geq 1$, il existe un réel $C_{k,p} > 0$, indépendant de N, tel que pour toute fonction ϕ continue bornée de $\mathbb{R}^d$,*

$$\mathbb{E}\left[|\langle \phi, p_k\rangle - \langle \phi, \hat{p}_k\rangle|^p\right]^{\frac{1}{p}} \leq C_{k,p} \frac{\sup_{x_k\in\mathbb{R}^d} |\phi(x_k)|}{\sqrt{N}}$$

avec $\langle \phi, p_k\rangle = \int \phi(x_k) p(x_k|y_{0:k}) dx_k$ *et* $\hat{p}_k = \sum_{i=1}^{N} \omega_k^i \delta_{x_k^i}$.

Ce dernier résultat montre que l'erreur $\mathbb{L}^p$ converge à la vitesse $\frac{1}{\sqrt{N}}$. Mais comme la constante $C_{k,p}$ dépend du temps, sans hypothèse supplémentaire, $C_{k,p}$ augmente généralement de manière exponentielle avec le temps. Ceci suggère que pour maintenir une erreur bornée, on a en général besoin d'un nombre croissant de particules.

2.3 Le filtre particulaire régularisé (RPF)

On a vu que le ré-échantillonnage classique appauvrissait dans certains cas le système de particules dans la mesure où les schémas de redistribution sont fondées sur la duplication de particules. En particulier lorsque le bruit de dynamique est faible, les particules explorent peu l'espace d'état et le filtre peut diverger. Afin d'éviter ceci, le filtre particulaire régularisé (RPF) [37, p. 247-271] [89, 72], effectue le rééchantillonnage en simulant un échantillon selon une distribution continue, ce qui évite d'avoir plusieurs copies d'une même particule. La distribution continue utilisée est obtenue en régularisant la densité discrète $\hat{p}_k = \sum_{i=1}^{N} \omega_k^i \delta_{x_k^i}$, obtenue après l'étape de correction.

2.3.1 Régularisation d'une loi de probabilité discrète

Soit K une fonction sur $\mathbb{R}^d$ symétrique, telle que :

$$\int K(x)\,dx = 1 \qquad \int xK(x)\,dx = 0 \qquad \int \|x\|^2 K(x)\,dx < +\infty \tag{2.24}$$

Pour tout $h > 0$ et tout $x \in \mathbb{R}^d$, on définit le noyau de régularisation

$$K_h = \frac{1}{h} K\left(\frac{x}{h^d}\right)$$

Le paramètre h s'appelle facteur de dilatation. Si ν est une densité de probabilité, on définit la densité régularisée de ν par :

$$\nu_h(x) = (K_h * \nu)(x) = \int K_h(x-u)\nu(u)du \tag{2.25}$$

Considérons le cas où ν représente une densité de probabilité discrète de la forme $\nu = \sum_{i=1}^{N} \omega^i \delta_{x^i}$ obtenue par échantillonnage d'importance i.e. en simulant N points i.i.d. x^i, $i = 1, \ldots, N$ selon une loi a priori q et en les pondérant selon une fonction de vraisemblance g avec $\omega^i = g(x^i)$. Autrement dit, ν est une approximation de la loi a posteriori $p \propto gq$.La densité de probabilité de la distribution régularisée de ν est

$$\nu_h(x) = \sum_{i=1}^{N} \omega^i K_h(x - x^i) \tag{2.26}$$

Le paramètre h contrôle le degré de lissage de la densité ν. Une valeur de h petite rapproche ν_h de ν tandis qu'une valeur plus élevée a tendance à fusionner les modes de la densité ν.

En estimation de densité par noyaux, le noyau et le facteur de dilatation sont choisis de manière à minimiser l'erreur quadratique intégrée moyenne $\mathbb{E}\,\|\nu_h - p\|^2$ entre ν_h et la loi a posteriori p avec $\|\nu_h - p\|^2 = \int (\nu_h(x) - p(x))^2\,dx$.
En particulier, lorsque $w^i = \frac{1}{N}$ pour $i = 1, \ldots, N$, le noyau optimal K_{opt} est le noyau d'Epanechnikov défini par [112] :

$$K_{opt}(x) = \begin{cases} \frac{d+2}{2c_d}\left(1 - \|x\|^2\right) & \text{si } \|x\| < 1 \\ 0 & \text{sinon} \end{cases} \tag{2.27}$$

De manière générale, le facteur de dilatation optimal h_{opt} dépend de la quantité $\left(\int |\nabla^2 p(x)|\,dx\right)^{1/2}$ qui n'est facilement calculable que pour certaines densités de probabilités. Par exemple, si p est une densité gaussienne standard, en utilisant le noyau d'Epanechnikov (2.27),

$$h_{opt} = A(K)N^{-\frac{1}{d+4}} \qquad A(K) = [8c_d^{-1}(d+4)(2\sqrt{\pi})^d]^{\frac{1}{d+4}} \tag{2.28}$$

où c_d est le volume de la sphère unité. Pour les applications usuelles, il est suffisant d'utiliser un noyau gaussien, dont le coût de simulation est moindre. Dans ce cas, h_{opt} est obtenu selon

$$h_{opt} = A(K)N^{-\frac{1}{d+4}} \qquad A(K) = (4/(d+2))^{\frac{1}{d+4}} \tag{2.29}$$

Pour une densité quelconque, le choix de h peut se faire selon des méthodes classiques d'estimation de densité par noyaux (validation croisée [108], règle du *plug-in* [122]) mais elles impliquent un coût de calcul non négligeable. Pour s'en affranchir, on peut par exemple approcher la densité sous-jacente p par une gaussienne de moyenne la moyenne empirique et de covariance la covariance empirique S donnée par l'échantillon $x^1, \ldots, x^N$ [37]. Pour tenir compte du rapport d'échelle entre les composantes du vecteur d'état, on blanchit le système de particule afin de se ramener à un système de particules de covariance unitaire. En supposant que $S = AA^T$, les particules blanchies sont les $A^{-1}x^i$. La valeur de h donnée par (2.29) peut alors être utilisée pour le système de particules blanchies. Cette procédure de régularisation est équivalente à utiliser le noyau de régularisation :

$$K_h(x) = \frac{1}{\det(A)h^d} K\left(\frac{A^{-1}x}{h}\right) \tag{2.30}$$

Lorsque la densité a posteriori est multimodale, la méthode de blanchiment (« data sphering ») a tendance à surlisser la densité d'origine. Silverman [112] conseille de poser $h = ch_{opt}$, avec $c < 1$ afin de limiter cet effet. Nous verrons dans le chapitre 4 qu'il est possible de régulariser les densités multimodales en s'affranchissant de ce réglage empirique.

2.3.2 Algorithme du filtre particulaire régularisé

Il existe deux versions du RPF qui diffèrent par le choix de l'instant de régularisation [37, p. 253-254].

Le filtre particulaire pré-régularisé : l'étape de prédiction est identique à celle de SIR. En revanche, l'étape de correction est obtenue en considérant la distribution régularisée de la loi prédite empirique $\hat{p}_{k|k-1} = \sum_{i=1}^N \omega_{k-1}^i \delta_{x_{k|k-1}^i}$ où $\left\{\omega_{k-1}^i, x_{k|k-1}^i\right\}_{i=1}^N$ est le système de particules obtenues après l'étape de prédiction.

$$\hat{p}_{k|k-1}^h(x) = \left(K_h * \hat{p}_{k|k-1}\right)(x) = \sum_{i=1}^N \omega_{k-1}^i K_h(x - x_{k|k-1}^i)$$

L'approximation de la densité corrigée est alors

$$\hat{p}_k^h(x) \propto \sum_{i=1}^N \omega_{k-1}^i g_k(x_{k|k-1}^i) K_h(x - x_{k|k-1}^i)$$

Simuler un échantillon $\{x_k^i\}_{i=1}^N$ issue de cette loi requiert généralement de recourir à une méthode d'acceptation/rejet en prenant comme loi instrumentale la densité prédite régularisée $\hat{p}_{k|k-1}^h(x)$. Cette méthode est donc généralement coûteuse et n'est pas recommandée.

Le filtre particulaire post-régularisé : La régularisation intervient après l'étape de correction du SIR. L'étape de ré-échantillonnage consiste cette fois à échantillonner la distribution régularisée de la loi a posteriori empirique $\hat{p}_{k|k} = \sum_{i=1}^{N} \omega_k^i \delta_{x_k^i}$. Le surcout du à la régularisation par rapport au ré-échantillonnage classique provient uniquement de la génération de N variables aléatoires selon le noyau K. Pour cette raison on préfèrera l'algorithme du filtre particulaire post-régularisé à sa version pré-régularisée.

Algorithme 3 : RPF (filtre particulaire régularisé)

---

pour $i = 1, \ldots, N$ **faire**
- tirer $x^i \sim q(x_0|y_0)$
- calculer $\omega_0^i \propto \frac{\mu_0(x_0^i)p(y_0|x_0^i)}{q(x_0^i|y_0)}$
- normaliser les poids ω_0^i

fin pour
pour $k = 1, \ldots, n$ **faire**
- *[Prédiction]* tirer $x_k^i \sim q(x_k^i|x_{k-1}^i, y_k)$, $i = 1, \ldots, N$
- *[Correction]* calculer les poids $\omega_k^i \propto \omega_{k-1}^i \frac{p(x_k^i|x_{k-1}^i)p(y_k|x_k^i)}{q(x_k^i|x_{k-1}^i, y_k)}$, $i = 1, \ldots, N$
- normaliser les poids ω_k^i, $i = 1, \ldots, N$
- Calculer N_{eff} d'après (2.17)
- *[Ré-échantillonnage/Régularisation]* Si $N_{eff} < N_{th}$
  - générer $I_1, \ldots, I_N$ avec $\mathbb{P}(I_j = i) = \omega_k^i$
  - générer ε^i suivant le noyau d'Epanechnikov ou un noyau gaussien
  - calculer A_k tel que $S_k = A_k A_k^T$ où S_k est la covariance empirique de $\{\omega_k^i, x_k^i\}_{i=1}^N$
  - calculer h_{opt} d'après (2.28) ou (2.29) et poser $x_k^j = x_k^{I_j} + h_{opt} A_k \varepsilon^j$, $j = 1, \ldots, N$

fin pour

Les applications du RPF dans le domaine du pistage de cible [96] ou du recalage d'aéronefs par mesures radio-altimétriques [28] ont montré que la régularisation améliore sensiblement les performances du filtre particulaire classique (SIR) notamment en terme de taux de non-divergence, lorsque la dynamique est faiblement bruitée.

2.4 Les filtres particulaires hybrides (RBPF, KPKF)

2.4.1 Le filtre particulaire Rao-Blackwellisé (RBPF)

Il existe une classe de modèles, appelée modèles conditionnellement linéaires gaussiens, pour laquelle l'estimation peut se faire grâce à un filtre particulaire pour une partie du vecteur d'état et un filtre de Kalman pour le reste du vecteur. Un tel filtre est appelé *Rao-Blackwellized Particle Filter* [15, 36], *Mixture Kalman Filter* [17] ou *Marginalized Particle Filter* [110, 94].
On suppose que l'on peut partitionner le vecteur d'état x_k telle que $x_k^T = \left[(x_k^n)^T, (x_k^l)^T\right]$. Le

vecteur x_k^n désigne la partie non-linéaire de x_k tandis que x_k^l représente la partie linéaire, conditionnellement à x_k^n et $y_{0:k}$.
En outre, on fait l'hypothèse que le modèle s'écrit sous la forme suivante :

$$\left\{ \begin{array}{rcl} x_k^n & = & f_{k-1}^n(x_{k-1}^n) + F_{k-1}^n(x_{k-1}^n)x_{k-1}^l + G_{k-1}^n(x_{k-1}^n)w_{k-1}^n \\ x_k^l & = & f_{k-1}^l(x_{k-1}^n) + F_{k-1}^l(x_{k-1}^n)x_{k-1}^l + G_{k-1}^l(x_{k-1}^n)w_{k-1}^l \\ y_k & = & h_k(x_k^n) + H_k(x_k^n)x_k^l + v_k \end{array} \right. \tag{2.31}$$

où

- w_k est un bruit blanc gaussien de matrice de covariance Q_k et $w_k^T = \left[(w_k^n)^T, (w_k^l)^T\right]$ $Q_k = \begin{pmatrix} Q_k^n & Q_k^{ln} \\ (Q_k^{ln})^T & Q_k^l \end{pmatrix}$
- v_k est un bruit blanc gaussien de matrice de covariance R_k
- La partie non linéaire suit une loi $p_0(x_0^n)$ connue
- La partie linéaire x_0^l est une gaussienne $\mathcal{N}(\bar{x}_0, \bar{P}_0)$

La densité jointe $p(x_k^l, x_{0:k}^n|y_{0:k})$ se factorise comme $p(x_k^l, x_{0:k}^n|y_{0:k}) = p(x_k^l|x_{0:k}^n, y_{0:k})p(x_{0:k}^n|y_{0:k})$. On montre que sous les hypothèses du modèle $p(x_k^l|x_{0:k}^n, y_{0:k})$ est une densité gaussienne : il suffit donc de calculer à chaque pas de temps sa moyenne et sa covariance grâce au filtre de Kalman.

Par ailleurs, La loi conditionnelle de la partie non-linéaire de l'état $p(x_{0:k}^n|y_{0:k})$ peut être estimée grâce un filtre particulaire selon le schéma suivant :

- Prédiction
$$p(x_{0:k}^n|y_{0:k-1}) = p(x_k^n|x_{0:k-1}^n, y_{0:k-1})p(x_{0:k-1}^n|y_{0:k-1}) \tag{2.32}$$
- Correction
$$p(x_{0:k}^n|y_{0:k}) = \frac{p(y_k|x_{0:k}^n, y_{0:k-1})}{p(y_k|y_{0:k-1})}p(x_{0:k}^n|y_{0:k-1}) \tag{2.33}$$

Définissons les variables z_k^1 et z_k^2 par :

$$\left\{ \begin{array}{rcl} z_k^1 & = & x_{k+1}^n - f_k^n(x_k^n) \\ z_k^2 & = & y_k - h_k(x_k^n) \end{array} \right.$$

Détaillons à présent la récursion de Kalman dans le cadre du RBPF. On suppose que la densité $p(x_{k-1}^l|x_{0:k-1}^n, y_{0:k-1})$ est une loi gaussienne $\mathcal{N}\left(\hat{x}_{k-1}^l, P_{k-1}\right)$ connue dont la moyenne et la covariance dépendent de $x_{0:k-1}^n$ et $y_{0:k-1}$. Alors la moyenne et la covariance prédite s'obtiennent comme suit [94] :

$$\left\{ \begin{array}{rcl} \hat{x}_{k|k-1}^l & = & \bar{F}_{k-1}^l\hat{x}_{k-1}^l + G_{k-1}^l(Q_{k-1}^{ln})^T(G_{k-1}^nQ_{k-1}^n)^{-1}z_{k-1}^1 + f_{k-1}^l(x_{k-1}^n) \\ & + & L_{k-1}(z_{k-1}^1 - A_{k-1}^n\hat{x}_{k-1}^l) \\ P_{k|k-1} & = & \bar{F}_{k-1}^lP_{k-1}(\bar{F}_{k-1}^l)^T + G_{k-1}^l\bar{Q}_{k-1}^l(G_{k-1}^l)^T - L_{k-1}N_{k-1}L_{k-1}^T \\ N_k & = & F_k^nP_k(F_k^n)^T + G_k^nQ_k^n(G_k^n)^T \\ L_k & = & \bar{F}_k^lP_k(F_k^n)^TN_k^{-1} \end{array} \right. \tag{2.34}$$

avec

$$\left\{ \begin{array}{rcl} \bar{F}_k^l & = & F_k^l - G_k^l(Q_k^{ln})^T(G_k^nQ_k^n)^{-1}F_k^n \\ \bar{Q}_k^l & = & Q_k^l - (Q_k^{ln})^T(Q_k^n)^{-1}Q_k^{ln} \end{array} \right.$$

De même, la loi corrigée $p(x_k^l|x_{0:k}^n, y_{0:k})$ est une densité gaussienne $\mathcal{N}\left(\hat{x}_k^l, P_k\right)$ de moyenne

$$\hat{x}_k^l = \hat{x}_{k|k-1}^l + K_k(y_k - h_k(x_k^n) - H_k^T(x_k^n)\hat{x}_{k|k-1}^l) \tag{2.35}$$

et de covariance

$$P_k = P_{k|k-1} - K_k M_k K_k^T \tag{2.36}$$

où

$$\begin{array}{rcl} M_k & = & H_k(x_k^n)P_{k|k-1}H_k^T(x_k^n) + R_k \\ K_k & = & P_{k|k-1}H_k^T(x_k^n)M_k^{-1} \end{array} \tag{2.37}$$

Si l'on suppose que $p(x_{0:k}^n|y_{0:k}) \approx \sum_{i=1}^N \omega_k^i \delta_{x_{0:k}^{n,i}}$, la loi conditionnelle $p(x_k^l|y_{0:k})$ s'obtient selon

$$\begin{aligned} p(x_k^l|y_{0:k}) &= \int p(x_k^l|x_{0:k}^n, y_{0:k}) p(x_{0:k}^n|y_{0:k})\, dx_{0:k}^n \\ &\approx \sum_{i=1}^N \omega_k^i p(x_k^l|x_{0:k}^{n,i}, y_{0:k}) \end{aligned} \tag{2.38}$$

où les densités $p(x_k^l|x_{0:k}^{n,i}, y_{0:k})$ sont gaussiennes et dont les paramètres sont donnés par les équations (2.35) et (2.36).

Revenons à l'approximation particulaire de $p(x_{0:k}^n|y_{0:k})$. L'étape de prédiction (2.32) nécessite de déterminer la loi de $p(x_k^n|x_{0:k-1}^n, y_{0:k-1})$.

$$p(x_k^n|x_{0:k-1}^n, y_{0:k-1}) = \int p(x_k^n|x_{k-1}^l, x_{k-1}^n) p(x_{k-1}^l|x_{0:k-1}^n, y_{0:k-1}) dx_{k-1}^l$$

Comme $p(x_k^n|x_{k-1}^l, x_{k-1}^n)$ est une gaussienne $\mathcal{N}\left(f_k^n(x_{k-1}^n) + F_k^n(x_{k-1}^n)x_{k-1}^l, G_k^n Q_k^n (G_k^n)^T\right)$ et $p(x_{k-1}^l|x_{0:k-1}^n, y_{0:k-1})$ est également une densité gaussienne $\mathcal{N}\left(\hat{x}_l^k, P_k\right)$, il s'en suit que $p(x_k^n|x_{0:k-1}^n, y_{0:k-1})$ suit une distribution gaussienne

$$\mathcal{N}\left(f_k^n(x_{k-1}^n) + F_k^n(x_{k-1}^n)\hat{x}_{k-1}^l, F_k^n P_k (F_k^n)^T + G_k^n Q_k^n (G_k^n)^T\right) \tag{2.39}$$

Il reste à déterminer le terme $p(y_k|x_{0:k}^n, y_{0:k-1})$ qui intervient dans l'étape de correction (2.33). On remarque que y_k est la somme des termes $h_k(x_k^n)$, $H_k(x_k^n)x_k^l$ et du bruit de mesure v_k. De plus :

- conditionnellement à $x_{0:k}^n$ et $y_{0:k-1}$, $H_k(x_k^n)x_k^l$ est gaussien de moyenne $H_k(x_k^n)\hat{x}_{k|k-1}^l$ et de covariance $H_k(x_k^n)P_{k|k-1}H_k^T(x_k^n)$
- $v_k \sim \mathcal{N}(0, R_k)$ est indépendant de $H_k(x_k^n)x_k^l$ et $h_k(x_k^n)$

Donc la loi de $p(y_k|x_{0:k}^n, y_{0:k-1})$ est une gaussienne

$$\mathcal{N}\left(h_k(x_k^n) + H_k^T(x_k^n)\hat{x}_{k|k-1}^l, H_k(x_k^n)P_{k|k-1}H_k^T(x_k^n) + R_k\right) \tag{2.40}$$

Nous donnons ci-dessous une version de l'algorithme du RBPF où la loi d'importance utilisée pour la partie non linéaire est la loi $p(x_k|x_{0:k-1}, y_{0:k-1})$.

Algorithme 4 : RBPF

pour $i = 1, \ldots, N$ **faire**

  - tirer $x_k^{n,i} \sim p(x_0^n)$
  - poser $x_{0|-1}^{l,i} = \bar{x}_0^l$ et $P_{0|-1}^i = \bar{P}_0$

fin pour

pour $k = 1, \ldots, n$ **faire**

  - *[Correction particulaire]* calculer les poids $\omega_k^i \propto \omega_{k-1}^i p(y_k|x_k^{n,i}, y_{0:k-1})$, $i = 1, \ldots, N$ où la vraisemblance $p(y_k|x_k^{n,i}, y_{0:k-1})$ est donnée d'après (2.40)
  - normaliser les poids ω_k^i, $i = 1, \ldots, N$
  - Calculer N_{eff} d'après (2.19)
  - *[Ré-échantillonnage]* Si $N_{eff} < N_{th}$ rééchantilloner le système de particules
  - *[Correction de Kalman]* Pour $i = 1, \ldots, N$, calculer $\hat{x}_k^{l,i}$ et $P_k^{l,i}$ d'après (2.35) et (2.36)
  - *[Prédiction particulaire]* Générer un échantillon $x_{k+1}^{n,i} \sim p(x_{k+1}^n|x_{0:k}^n, y_{0:k})$ (cf. (2.39))
  - *[Prédiction de Kalman]* Pour $i = 1, \ldots, N$, calculer $\hat{x}_{k+1|k}^{l,i}$ et $P_{k+1|k}^{l,i}$ d'après (2.34)

fin pour

Le filtre Rao-Blackwellisé produit des estimées en général meilleures que le SIR en terme d'erreur Monte Carlo. Cependant, à nombre de particules constant, il est plus coûteux puisqu'en toute généralité, il recquiert un filtre de Kalman par particule. Néanmoins, il est possible d'obtenir un compromis coût de calcul/performances en faveur du RBPF pour bien des applications. Par exemple, en recalage par mesures radio-altimétriques ou en pistage par mesure angulaire, la matrice de covariance du filtre de Kalman est indépendante de la partie non-linéaire ce qui réduit la complexité du RBPF étant donné qu'il il n'y a qu'une matrice de covariance à mettre à jour.

2.4.2 Un exemple de filtre particulaire à noyau : le Kalman Particle Kernel Filter (KPKF)

Le KPKF [28, 103] s'appuie sur la théorie d'estimation de densité par noyau en exprimant les densités prédites et corrigées comme une somme pondérée de noyaux gaussiens. Dans la théorie du KPKF, le bruit de dynamique w_k et le bruit d'observation v_k sont supposés additifs gaussiens de moyennes nulles et de matrices de covariance respectives Q_k et R_k.
Étant donné une densité de probabilité f et $\{x^i, i = 1 \ldots N\}$ un N-échantillon issu de f, un estimateur de densité par noyau est donné par :

$$\hat{f} = \frac{1}{N} \sum_{i=1}^{N} \varphi(x - x^i|P)$$

où l'application $x \mapsto \varphi(x|P)$ désigne un noyau gaussien de moyenne nulle et de covariance $P = h^2 \operatorname{Cov}(x^i) = \frac{h^2}{N-1} \sum_{i=1}^{N} (x^i - \bar{x})(x^i - \bar{x})^T$. h est le paramètre de dilatation du noyau et joue le même rôle que dans le filtre particulaire régularisé.
Conformément à l'approche ci-dessus, le KPKF formule la loi prédite $p(x_k|y_{0:k-1})$ comme un mélange de N noyaux gaussiens :

$$p(x_k|y_{0:k-1}) \approx \sum_{i=1}^{N} \omega_{k|k-1}^i \varphi(x_k - x_{k|k-1}^i|P_{k|k-1}^i) \tag{2.41}$$

où pour tout $i = 1, \ldots, N$, $P^i_{k|k-1}$ est une matrice de covariance égale à h^2 fois la covariance empirique des particules prédites, i.e. :

$$P^i_{k|k-1} = h^2 \sum_{i=1}^{N} \omega^i_{k|k-1}(x^i_{k|k-1} - \bar{x}_{k|k-1})(x^i_{k|k-1} - \bar{x}_{k|k-1})^T \tag{2.42}$$

$\bar{x}_{k|k-1} = \sum_{i=1}^{N} \omega^i_{k|k-1} x^i_{k|k-1}$ étant la moyenne prédite donnée par le système de particules. La particularité du KPKF est que, à l'étape de correction, les covariance $P^i_{k|k-1}$ sont suffisamment petites pour justifier une linéarisation de la fonction d'observation h_k autour des particules prédites $x^i_{k|k-1}$, $i = 1, \ldots, N$. Étant donné que le bruit est additif gaussien, on vérifie que la densité a posteriori $p(x_k|y_{0:k})$ s'exprime comme un mélange de noyaux gaussiens dont les moyennes et les matrices de covariance s'obtiennent en appliquant les équations de Kalman [28]. Autrement dit,

$$p(x_k|y_{0:k}) \approx \sum_{i=1}^{N} \omega^i_k \varphi(x_k - x^i_k|P^i_k) \tag{2.43}$$

avec

$$H^i_k = \nabla h_k(x^i_{k|k-1}) \tag{2.44}$$
$$x^i_k = x^i_{k|k-1} + K^i_k(y_k - h_k(x^i_{k|k-1})) \tag{2.45}$$
$$\Sigma^i_k = H^i_k P^i_{k|k-1}(H^i_k)^T + R_k \tag{2.46}$$
$$K^i_k = P^i_{k|k-1}(H^i_k)^T(\Sigma^i_k)^{-1} \tag{2.47}$$
$$P^i_k = P^i_{k|k-1} - P^i_{k|k-1}(H^i_k)^T(\Sigma^i_k)^{-1} H^i_k P^i_{k|k-1} \tag{2.48}$$
$$\omega^i_k = \frac{\omega^i_{k|k-1}\varphi(y_k - y^i_{k|k-1}|\Sigma^i_k)}{\sum_{j=1}^{N} \omega^i_{k|k-1}\varphi(y_k - y^i_{k|k-1}|\Sigma^i_k)} \tag{2.49}$$

L'étape de prédiction nécessite le calcul de

$$p(x_{k+1}|y_{0:k}) = \sum_{i=1}^{N} \omega^i_k \int_{\mathbb{R}^d} \varphi(x_{k+1} - f_{k+1}(x)|Q_{k+1})\varphi(x - x^i_k|P^i_k)du$$

En remarquant que $\|P^i_k\| \leq \left\|P^i_{k|k-1}\right\|$, on peut de nouveau effectuer une linéarisation de la fonction d'observation f_{k+1} autour des particules corrigées x^i_k.

$$f_{k+1}(x) \approx f_{k+1}(x^i_k) + F^i_{k+1}(x - x^i_k)$$

où $F^i_{k+1} = \nabla f_{k+1}(x^i_k)$. Dans de nombreuses applications, la dynamique considérée est linéaire et il n'est donc pas nécessaire de procéder à une linéarisation.
Dans tous les cas, la densité prédite est égale à une somme pondérée de noyaux gaussiens :

$$p(x_{k+1}|y_{0:k}) = \sum_{i=1}^{N} \omega^i_k \varphi(x_k - f_{k+1}(x^i_k)|P^i_{k+1|k})$$

avec $P^i_{k+1|k} = F^i_{k+1} P^i_k (F^i_{k+1})^T + Q_{k+1}$.
L'étape de prédiction soulève une difficulté : la matrice de covariance $P^i_{k+1|k}$ n'est plus de l'ordre de h^2. Il est donc nécessaire de rajouter une étape supplémentaire afin de conserver la structure donnée par (2.41) et (2.42). Cette étape, nommée *ré-échantillonnage partiel* [28] par leurs auteurs, consiste à rajouter à chaque particule $f_{k+1}(x^i_{k+1|k})$ un bruit gaussien centré de covariance $P^i_{k+1|k} - h^{*2}\Pi_{k+1|k}$ où

$$\Pi_{k+1|k} = \sum_{i=1}^{N} \omega^i_k P^i_{k+1|k} + \sum_{i=1}^{N} \omega^i_k (f_{k+1}(x^i_k) - \overline{f_{k+1}(x_k)})(f_{k+1}(x^i_k) - \overline{f_{k+1}(x_k)})^T$$

est la covariance de la densité prédite. Le paramètre h^* est défini par

$$h^{*2} = \min_i \lambda^i_{min}$$

λ^i_{min}, $i = 1, \ldots, N$ est la plus petite valeur propre de $(C^T_{k+1})^{-1} P^i_{k+1|k} C^{-1}_{k+1}$ et C_{k+1} est la racine carrée de $\Pi_{k+1|k}$ ($C^T_{k+1} C_{k+1} = \Pi_{k+1|k}$).
Lorsque la variance des poids $\{\omega^i_k\}_{i=1}^N$ est élevée, il est préférable d'effectuer un ré-échantillonnage dit total ce qui revient à échantillonner N particules $\left\{x^i_{k+1|k}\right\}_{i=1}^N$ suivant la densité

$$\hat{p}(x_{k+1}|y_{0:k}) = \sum_{i=1}^{N} \omega^i_k \varphi(x_{k+1} - x^i_{k+1|k}|h^{*2}\Pi_{k+1|k}) \tag{2.50}$$

Rappelons que la variance des poids normalisés $\{\omega^i_k\}_{i=1}^N$ peut s'estimer en contrôlant en ligne un critère sur la taille effective de l'échantillon N_{eff} ou bien en évaluant le critère entropique présentée dans la section 2.2.3.

2.5 Comportement du filtre particulaire en présence de multimodalités

La section 2.2.3 a mis en évidence la nécessité de ré-échantillonner le système de particules afin de garantir la stabilité du filtre (mais pas nécessairement la convergence).
Si l'on considère une distribution multimodale approchée par un ensemble fini de N particules pondérées $\{\omega^i_k, x^i_k\}_{i=1}^N$, il n'y a pas de garantie qu'après ré-échantillonnage, ce système de particules représente tous les modes de la distribution conditionnelle. En effet, ré-échantillonner consiste essentiellement à tirer avec remise dans l'ensemble $\{x^i_k\}_{i=1}^N$, la probabilité de tirage de chaque particule étant égale à son poids.

Ce phénomène a été rapporté dans la littérature du filtrage particulaire dans le cadre du pistage multi-cibles [58, 95], du pistage vidéo [59, 118, 95] ou en localisation de robot [85]. Une justification a été fournie par King et Forsyth [67] lorsque l'état x_k prend ses valeurs dans un ensemble

fini discret. Les auteurs considèrent entre autres un modèle à espace d'état où $x_k \in \{-1, 1\}$. Dans cette expérience, l'objet stationnaire est situé en $x = 1$ et on suppose que l'observation est parfaite et se fait dans un plan image par réflection sur deux miroirs (figure 2.2). Il apparaît donc à la fois en $x = 1$ et $x = -1$ sur le plan image. Cela crée une ambiguité sur la position réelle de l'objet, qui peut être modélisée par la fonction de vraisemblance $p(y|x) = 0.5\delta_x(y) + 0.5\delta_{-x}(y)$. Cette modélisation est purement illustrative mais permet de comprendre le phénomène de perte

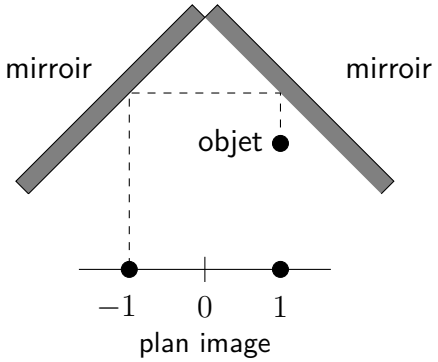


FIGURE 2.2 – "Pistage" d'un objet stationnaire

de mode en filtrage particulaire. Pour cela, King et Forsyth supposent une densité a priori initiale $p_0(x) = 0.5\delta_1(x) + 0.5\delta_{-1}(x)$. Comme l'objet est stationnaire, $p(x_k|x_{k-1}) = \delta_{x_{k-1}}(x_k)$, de sorte que $p(x_k|y_{0:k}) = p_0(x_k)$. La loi a posteriori a donc 2 modes quelque soit k : -1 et 1.
En utilisant un filtre particulaire pour estimer $p(x_k|y_{0:k})$, les échantillons x_k^i, $i = 1, \ldots, N$ prennent les valeurs -1 ou 1 et ont pour poids $\omega_k^i = \frac{1}{N}$. La distribution empirique est donc entièrement caractérisée par le nombre A_k d'échantillons qui sont situé sur le mode 1.

$$A_k = \text{Card}\left\{i \in [\![0, N]\!] \mid x_k^i = 1\right\} \tag{2.51}$$

de sorte que à l'instant k, l'approximation particulaire de $p(x_k|y_{0:k})$ est

$$\hat{p}_k(x) = \frac{A_k}{N}\delta_1(x) + \frac{N - A_k}{N}\delta_{-1}(x) \tag{2.52}$$

Une bonne approximation $\hat{p}_k$ correspond donc à $A_k = \frac{N}{2}$, $\forall k$.
Supposons que le ré-échantillonnage multinomial est appliquée à chaque instant (*bootstrap filter*). Cela correspond à un tirage uniforme avec remise dans la population $\left\{x_k^1, x_k^2, \ldots, x_k^N\right\}$. La v.a. A_k est donc une chaine de Markov homogène à valeur dans $\{0, 1, \ldots, N\}$. La probabilité de transition est une loi binomiale $\mathcal{B}(N, A_k/N)$:

$$\mathbb{P}(A_k = j|A_{k-1} = m) = \binom{N}{j}\left(\frac{m}{N}\right)^j\left(1 - \frac{m}{N}\right)^{N-j} = p_{mj} \tag{2.53}$$

La perte de l'un des modes correspond à $A_k \in \{0, N\}$, c'est-à-dire que la chaîne A_k a atteint un état absorbant. Les états 0 et N sont dits absorbants car une fois que la chaine A_k atteint ces

valeurs, elle y reste. Cette modélisation est analogue aux chaînes de Wright-Fisher en génétique des populations [39]. Les deux modes sont comparables à deux allèles d'une population de taille fixe, asexuée et dont les générations ne se chevauchent pas. On définit le temps d'absorption T comme étant le temps d'atteinte de l'ensemble $\{0, N\}$.

$$T \triangleq \inf \{n \geq 0 \mid A_n \in \{0, N\}\} \tag{2.54}$$

En génétique des populations, cela correspond à la perte définitive d'un allèle. L'espace des états étant fini, le temps d'atteinte T est presque sûrement fini. Si on désigne par u la proportion initiale d'échantillons correspondant au premier mode, le temps moyen d'absorption est asymptotiquement linéaire en N [39].

$$\mathbb{E}[T \mid A_0 = Nu] \underset{N\to\infty}{\sim} -2N\left(u \ln u + (1-u)\ln(1-u)\right) \tag{2.55}$$

Cette approximation est relativement bonne même pour des "petites" valeurs de N. Si l'on considère le problème précédent du pistage d'un objet stationnaire avec $N = 100$ (modes équiprobables), on peut constater que le temps moyen d'absorption empirique, calculé sur 500 réalisations indépendantes, est proche de son équivalent asymptotique (cf. figure 2.3). En particulier pour $u = 1/2$ (modes équiprobables), le temps moyen théorique est de 138 itérations.

Par ailleurs, nous avons calculé la variance du temps moyen d'absorption vaut asymptotiquement :

$$\begin{aligned}\mathrm{Var}(T \mid A_0 = Nu) = &-2N\left[u \ln u + (1-u)\ln(1-u)\right] + 8N^2 u \,\mathrm{dilog}\, u + 8N^2 u \ln u \\ &- 8N^2 \,\mathrm{dilog}\, u - 8N^2 u \,\mathrm{dilog}(1-u) + 8N^2 \ln(1-u) - 8N^2 u \ln(1-u) \\ &- 4N^2[u \ln u + (1-u)\ln(1-u)]^2 + \frac{4}{3}\pi^2 N^2\end{aligned} \tag{2.56}$$

où dilog est la fonction dilogarithme définie par

$$\mathrm{dilog}(u) = \int_1^u \frac{\ln t}{1-t}\, dt \tag{2.57}$$

Des éléments de preuve sont fournis en annexe (B).

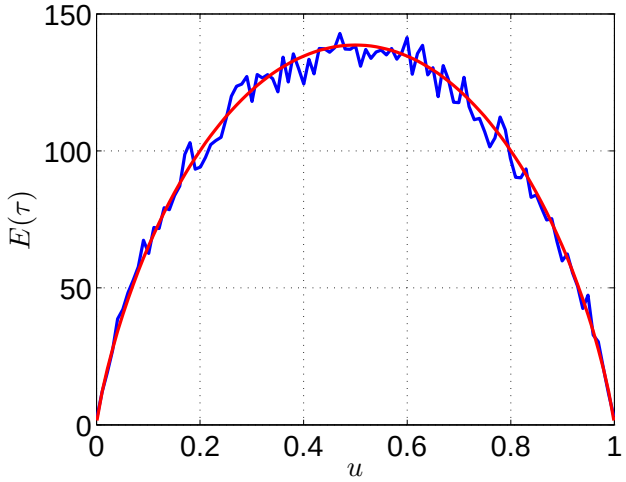


Figure 2.3 – Temps moyen d'absorption de mode. Rouge : approximation asymptotique, Bleu : moyenne empirique

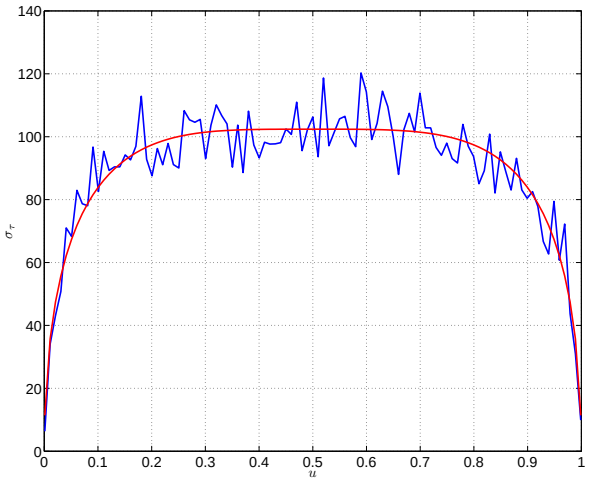


Figure 2.4 – Ecart-type du temps d'absorption de mode. Rouge : approximation asymptotique, Bleu : écart-type empirique

Lorsque la répartition des particules entre les deux modes est identique, la variance asymptotique se réduit à :

$$\mathrm{Var}(T \mid A_0 = N/2) \underset{N\to\infty}{\sim} (\frac{2}{2}\pi^2 - 8\log 2)N^2 + 2N\log 2$$

Lorsque N est grand, l'écart-type du temps de perte de mode est proportionnel à N. Ainsi, le temps moyen d'absorption augmente avec N à la même vitesse que l'écart-type associé. On observe en pratique que même pour des N grands la probabilité de perdre rapidement un mode n'est pas nulle.

2.5.1 Cas où le nombre de modes est supérieur ou égal à 3

Lorsqu'on considère un modèle discret comportant plus de 2 modes dans la loi a posteriori $p(x|y)$, on observe également que le ré-échantillonnage répété provoque une perte progressive des modes jusqu'à n'en avoir plus qu'un seul [67]. S'il y a initialement b modes, contenant initialement $X_1, X_2, \ldots, X_b$ échantillons, le temps moyen de réduction à un mode est asymptotiquement égal à :

$$-2N \sum_{i=1}^{b} \frac{N - X_i}{N} \ln\left(\frac{N - X_i}{N}\right)$$

2.5.2 Cas où l'état est continu

L'expérience décrite par King et Forsyth met en jeu un système à espace d'état sans bruit de dynamique ni bruit de mesure ce qui permet d'obtenir une approximation asymptotique sur les temps moyen de perte de mode.
Dans un cadre plus général avec un bruit de dynamique et/ou d'observation, on observe le même phénomène de perte de modes [67, 118]. Considérons le modèle suivant :

$$\begin{cases} x_k = & x_{k-1} + w_k \\ y_k = & x_k^2 + v_k \end{cases} \tag{2.58}$$

où $x_k\,, k = 1, \ldots, 200$ est une variable scalaire.
La loi a priori est supposée symétrique bimodale $\frac{1}{2}\mathcal{U}(-6.01, -5.99) + \frac{1}{2}\mathcal{U}(5.99, 6.01)$ et les bruits d'état et de mesure sont des bruits blancs gaussiens de variances respectives $\sigma_w^2 = 0.01$ et $\sigma_v^2 = 0.01$. La nature de l'équation de mesure assure que la distribution $p(x_k|y_{0:k})$ est bimodale avec des modes de signes opposés. Un moyen simple d'associer une particule à un mode est de lui associer le mode de même signe. L'écart-type du bruit de dynamique assure qu'une particule x^j qui sera initialement associé au i-ème mode le restera étant donné la faible probabilité pour que la somme cumulée des bruits de dynamique provoque un changement du signe de x^j.
Nous avons étudié dans [88] le temps moyen de perte de mode d'un filtre particulaire standard sur le modèle (2.58). Dans cet article, nous avons considéré un filtre particulaire avec ré-échantillonnage à chaque itération (*bootstrap filter*) et utilisant $N = 100$ particules. Le filtre est initialisé avec $N/2$ particules dans chaque mode. Le temps moyen de perte de mode calculé empiriquement sur la base de 1000 trajectoires indépendantes est de 11 itérations. En comparant (à nombre de particules égal) avec le modèle statique illustré par la figure 2.2 où le temps moyen est de l'ordre de 138 itérations, on observe que la présence de bruits de dynamique et d'observation est de nature à accélérer la perte de l'un des modes. Cela est du au fait que le ré-échantillonnage se fait en tirant les particules proportionnellement à leurs poids normalisés ω_k^i qui sont différents de $\frac{1}{N}$, contrairement au cas stationnaire.
Nous verrons ci-après une implémentation, proposée par Vermaak et al. [118], qui permet de traiter le problème du ré-échantillonnage au niveau de chaque mode de la densité empirique dans le soucis de préserver autant que possible son caractère multimodal.

2.6 Modèles de mélange pour le filtrage particulaire

Les modèles de mélange finis consistent à représenter une densité de probabilité f comme une somme pondérée de M densités de probabilités f_j .

$$f(x) = \sum_{j=1}^{M} \alpha_j f_j(x) \text{ avec } \sum_{j=1}^{M} \alpha_j = 1 \tag{2.59}$$

Les réels α_j sont les coefficients ou poids de mélange.

Les modèles de mélange sont utilisés en classification de données [83], en segmentation d'image [16] ou extraction d'image [101]. Ils sont également utiles pour représenter une loi de probabilité multimodale. Ils se prêtent bien à la problématique du filtrage dans un cadre ou la fonction de mesure est ambiguë. Ils ont notamment été utilisés dans un cadre particulaire en localisation de robot [85, 77], en pistage vidéo [118], en pistage multi-objet[113] ou encore en pistage multi-cibles [47]. Nous présentons ci-après le cadre théorique du filtrage pour le modèle de mélange et l'algorithme particulaire correspondant.

2.6.1 Filtrage pour les modèles de mélange

Nous rappelons le modèle à espace d'état (2.6).

$$\begin{cases} x_k &= f_k\left(x_{k-1}, w_k\right) \\ y_k &= h_k(x_k, v_k) \end{cases}$$

Formulons la loi conditionnelle $p_{k|k} = p(x_k|y_{0:k})$ soit comme un mélange de lois conditionnelles :

$$p_{k|k} = \sum_{j=1}^{M} \alpha_{j,k} p_j(x_k|y_{0:k}) \tag{2.60}$$

où $\sum_{j=1}^{M} \alpha_{j,k} = 1$. Les composantes p_j sont des lois conditionnelles non-paramétriques. Cette décomposition n'étant pas unique, nous allons supposer que chaque p_j est monomodale, de manière à capter la multimodalité de la loi a posteriori $p(x_k|y_{0:k})$.
La mise à jour du filtre suppose la mise à jour des composantes $\{p_j, j = 1 \cdots M\}$ et des poids de mélange $\alpha_{j,k}$.
Supposons que $p_{k-1|k-1} = p(x_{k-1}|y_{0:k-1})$ soit connu. Le calcul de $p_{k|k}$ se fait selon les équations de Chapman-Kolmogorov :

– *Prédiction*

$$p_{k|k-1} = p(x_k|y_{0:k-1}) = \int p(x_k|x_{k-1}) p(x_{k-1}|y_{0:k-1}) \, \mathrm{d}x_{k-1} \tag{2.61}$$

– *Correction*

$$p_{k|k} = p(x_k|y_{0:k}) = \frac{p(y_k|x_k) p(x_k|y_{0:k-1})}{\int p(y_k|x_k) p(x_k|y_{0:k-1}) \, \mathrm{d}x_k} \tag{2.62}$$

En prenant en compte la formulation sous forme de mélange de loi conditionnelle donnée par (2.60), on a

– *Prédiction*

$$p_{k|k-1} = \sum_{j=1}^{M} \alpha_{j,k-1} p_j(x_k|y_{0:k-1}) \text{ avec } p_j(x_k|y_{0:k-1}) = \int p(x_k|x_{k-1}) p_j(x_{k-1}|y_{0:k-1}) \mathrm{d}x_{k-1} \tag{2.63}$$

– *Correction*

$$\begin{aligned} p_{k|k} &= \frac{\sum_{j=1}^{M} \alpha_{j,k-1} p(y_k|x_k) p_j(x_k|y_{0:k-1})}{\sum_{l=1}^{M} \alpha_{l,k-1} \int p(y_k|x_k) p_l(x_k|y_{0:k-1}) \, \mathrm{d}x_k} \\ &= \sum_{j=1}^{M} \frac{\alpha_{j,k-1} \int p(y_k|x_k) p_j(x_k|y_{0:k-1}) \, \mathrm{d}x_k}{\sum_{l=1}^{M} \alpha_{l,k-1} \int p(y_k|x_k) p_l(x_k|y_{0:k-1}) \, \mathrm{d}x_k} \underbrace{\left(\frac{p(y_k|x_k) p_j(x_k|y_{0:k-1})}{\int p(y_k|x_k) p_j(x_k|y_{0:k-1}) \, \mathrm{d}x_k} \right)}_{p_j(x_k|y_{0:k})} \end{aligned} \tag{2.64}$$

Etant donné que $\int p(y_k|x_k) p_j(x_k|y_{0:k-1}) \, \mathrm{d}x_k = p_j(y_k|y_{0:k-1})$, on en déduit donc l'équation de mise à jour des poids de mélange :

$$\alpha_{j,k} = \frac{\alpha_{j,k-1} p_j(y_k|y_{0:k-1})}{\sum_{l=1}^{M} \alpha_{l,k-1} p_l(y_k|y_{0:k-1})} \tag{2.65}$$

2.6.2 Approximation particulaire pour les modèles de mélange et algorithme MPF

La précédente section a mis en évidence la procédure de filtrage pour les modèles de mélange. Voyons maintenant comment obtenir l' équivalent particulaire des filtres de mélange. Dans la littérature du filtrage particulaire, l'algorithme a été introduit par Vermaak, Doucet et Perez sous le nom de *mixture particle filter (MPF)* [118]. Ce même algorithme est également connu comme le *clustered particle filter* [126]. L'idée sous-jacente est d'affecter un filtre particulaire à chaque composante p_j du mélange. A l'étape de prédiction, les filtres évoluent indépendamment et à l'étape de correction, ils interagissent lors de la mise à jour des poids de mélange $\{\alpha_{j,k}\}_{j=1}^{M}$via l'équation (2.65).
Conformément aux notations précédentes, on notera $\{x_k^i\}_{i=1}^{N}$ l'ensemble des particules, $\{\omega_k^i\}_{i=1}^{N}$ les poids associés et $I_j \subseteq \{1, \ldots, N\}$ l'ensemble des indices des particules appartenant à la composante d'indice j.
La loi a posteriori est donc approchée selon

$$p(x_k|y_{0:k}) \approx \sum_{j=1}^{M} \alpha_{j,k} \sum_{i \in I_j} \omega_k^i \delta_{x_k^i} \tag{2.66}$$

avec

$$\sum_{i \in I_j} \omega_k^i = 1$$

et

$$\sum_{j=1}^{M} \alpha_{j,k} = 1$$

L'étape de prédiction se fait en échantillonnant N particules $x_k^i \sim q(x|x_{k-1}^i, y_k)$ où q est une densité d'importance. D'après (2.64), l'équation de mise à jour des poids d'importance $\{\omega_k^i\}_{i=1}^N$ se fait selon :

$$\tilde{\omega}_k^i = \omega_{k-1}^i \frac{p(y_k|x_k^i)p(x_k^i|x_{k-1}^i)}{q(x_k^i|x_{k-1}^i, y_k)} \tag{2.67}$$

$$\forall i \in I_j, \quad \omega_k^i = \frac{\tilde{\omega}_k^i}{\sum_{l \in I_j} \tilde{\omega}_k^l} \tag{2.68}$$

Afin d'obtenir les poids de mélange à l'instant k, il faut estimer la vraisemblance prédite $p_j(y_k|y_{0:k-1}) = \int p(y_k|x_k)p_j(x_k|y_{0:k-1})\,\mathrm{d}x_k$. Appelons $\hat{p}_j(x_k|y_{0:k-1})$ une approximation particulaire de la loi prédite de la composante d'indice j. Alors,

$$p_j(y_k|y_{0:k-1}) = \int p(y_k|x_k)p_j(x_k|y_{0:k-1})\,\mathrm{d}x_k \tag{2.69}$$

$$= \iint p(y_k|x_k)p(x_k|x_{k-1})p_j(x_{k-1}|y_{0:k-1})\,\mathrm{d}x_k\,\mathrm{d}x_{k-1} \tag{2.70}$$

$$\approx \sum_{i \in I_j} \omega_{k-1}^i \frac{p(y_k|x_k^i)p(x_k^i|x_{k-1}^i)}{q(x_k^i|x_{k-1}^i, y_k)} \tag{2.71}$$

$$\approx \sum_{i \in I_j} \tilde{\omega}_k^i \tag{2.72}$$

On en déduit l'estimateur des coefficients de mélange suivant :

$$\alpha_{j,k} \approx \frac{\alpha_{j,k-1} w_{j,k}^+}{\sum_{p=1}^{M} \alpha_{p,k-1} w_{p,k}^+} \tag{2.73}$$

où $w_{j,k}^+$ est défini par :

$$w_{j,k}^+ = \sum_{i \in I_j} \tilde{\omega}_k^i \tag{2.74}$$

Jusqu'ici, on a supposé que le nombre de composantes de la loi a posteriori restait fixe. Dans l'hypothèse où une composante est affectée à un mode de la loi conditionnelle $p_{k|k}$, cela signifie que le nombre de modes est constant. Cependant pour des applications telles que le suivi d'objets dans les séquences vidéo, une composante p_j correspond idéalement à la probabilité d'un objet particulier. Le nombre d'objets pouvant évoluer au cours du temps, il parait naturel d'introduire une dépendance du nombre de composantes M par rapport au temps. En recalage de navigation inertielle par mesures radio-altimétriques, la modélisation de la loi conditionnelle sous-forme de

mélange fini correspond à une ambiguïté sur la position réelle de l'aéronef qui se matérialise par plusieurs positions possibles dans le plan horizontal. En fonction des similarités de terrain et de l'incertitude de mesure du radio-altimètre, le nombre de positions candidates peut évoluer.
Afin de prendre en compte ces remarques, Vermaak et al. [118] formulent la loi a posteriori à l'instant k comme un mélange de M_k composantes :

$$p_{k|k} = \sum_{j=1}^{M_k} \alpha_{j,k} p_j(x_k|y_{0:k}) \tag{2.75}$$

Soit $\hat{p}_{k|k}$ l'approximation particulaire de $p_{k|k}$. Alors,

$$\hat{p}_{k|k} = \sum_{j=1}^{M_k} \alpha_{j,k} \sum_{i \in I_{j,k}} \omega_k^i \delta_{x_k^i} \tag{2.76}$$

où $I_{j,k} \subseteq \{1, \ldots, N\}$ est l'ensemble des indices des particules approximant la composante p_j à l'instant k.
A l'instant $k+1$ on a donc :

$$p_{k+1|k+1} \approx \hat{p}_{k+1|k+1} = \sum_{j=1}^{M_k} \alpha_{j,k+1} \sum_{i \in I_{j,k}} \omega_{k+1}^i \delta_{x_{k+1}^i} \tag{2.77}$$

$$= \sum_{i=1}^{N} \alpha_{c_1(i),k+1} \omega_{k+1}^i \delta_{x_{k+1}^i} \tag{2.78}$$

où $c_1(i) = j$ si i appartient à $I_{j,k}$. Dans la formulation (2.77), les lois $\Pi_j = \sum_{i \in I_{j,k}} \omega_{k+1}^i \delta_{x_{k+1}^i}$ peuvent contenir plusieurs modes, auquel cas il est judicieux de séparer les particules en autant de modes que nécessaire. Inversement, deux lois $\hat{\Pi}_l$ et Π_m peuvent avoir des modes très rapprochés ce qui incite à regrouper les particules correspondantes.
Cela se traduit par la formulation de $\hat{p}_{k+1|k+1}$ comme somme de M_{k+1} composantes $\hat{\Pi}'_l$, $l = 1, \ldots, M_{k+1}$ obtenues après re-clustering des composantes $\hat{\Pi}_l$, $l = 1, \ldots, M_k$.

$$\hat{p}_{k+1|k+1} = \sum_{l=1}^{M_{k+1}} \beta_{l,k+1} \hat{\Pi}'_l$$

$$\hat{p}_{k+1|k+1} = \sum_{l=1}^{M_{k+1}} \beta_{l,k+1} \sum_{i \in I_{l,k+1}} \nu_{k+1}^i \delta_{x_{k+1}^i} \tag{2.79}$$

avec $\sum_{l=1}^{M_{k+1}} \beta_{l,k+1} = 1$, $\sum_{i \in I_{j,k+1}} \nu_{k+1}^i = 1$ et $\hat{\Pi}'_l = \sum_{i \in I_{l,k+1}} \nu_{k+1}^i \delta_{x_{k+1}^i}$.
$I_{l,k+1} \subseteq \{1, \ldots, N\}$ est l'ensemble des indices des particules représentant la loi $\hat{\Pi}'_l$. On peut réécrire l'égalité (2.79) sous la forme :

$$\hat{p}_{k+1|k+1} = \sum_{i=1}^{N} \beta_{c_2(j),k+1} \nu_{k+1}^j \delta_{x_{k+1}^j} \tag{2.80}$$

où $c_2(i) = l$ si $i \in I_{l,k+1}$.
On peut également transformer l'équation (2.78) sous la forme :

$$\hat{p}_{k+1|k+1} = \sum_{i=1}^{N} \alpha_{c_1(i),k+1} \omega_{k+1}^i \delta_{x_{k+1}^i} \tag{2.81}$$

$$= \sum_{l=1}^{M_{k+1}} \sum_{i \in I_{l,k+1}} \alpha_{c_1(i),k+1} \omega_{k+1}^i \delta_{x_{k+1}^i} \tag{2.82}$$

$$= \sum_{l=1}^{M_{k+1}} \sum_{i' \in I_{j,k+1}} \alpha_{c_1(i'),k+1} \omega_{k+1}^{i'} \sum_{i \in I_{l,k+1}} \frac{\alpha_{c_1(i),k+1} \omega_{k+1}^i}{\sum_{i' \in I_{l,k+1}} \alpha_{c_1(i'),k+1} \omega_{k+1}^{i'}} \delta_{x_{k+1}^i} \tag{2.83}$$

En identifiant l'égalité (2.83) avec l'équation (2.79), il vient :

$$\beta_{l,k} = \sum_{i \in I_{l,k+1}} \alpha_{c_1(i),k} \omega_k^i \tag{2.84}$$

et

$$\nu_k^i = \frac{\alpha_{c_1(i),k} \omega_k^i}{\beta_{c_2(i),k}} \tag{2.85}$$

Conclusion du Chapitre 2

Nous avons présenté les briques de base du filtre particulaire dans le cadre des modèles à espace d'état. L'algorithme SIS assure théoriquement une convergence vers la loi de filtrage lorsque le nombre de particule est infiniment grand. Cependant, l'utilisation d'un nombre fini de particules peut occasionner la divergence des algorithmes particulaires. En particulier, lorsque le bruit d'observation est faible ou lorsque la dynamique est faiblement bruitée, l'échantillonnage successif des particules selon la loi a priori ne garantit pas la convergence du filtre. Des variantes telles que le filtre particulaire régularisé (RPF) ou le KPKF permettent de rendre le filtre plus robuste lorsque ces situations surviennent mais n'éliminent pas totalement le risque de divergence. Le filtre Rao-Blackwellisé (RBPF) exploite la structure des modèles conditionnellement linéaires gaussiens afin d'effectuer une approximation particulaire sur une partie réduite de l'espace de l'état et réduire ainsi le nombre d'échantillons nécessaire, le reste de l'état étant estimée par un banc de filtres de Kalman. Nous avons également exposé une autre limitation des algorithmes particulaires, qui est le risque de perte de mode lorsque la loi conditionnelle de l'état est multimodale. Nous avons illustré l'impact du ré-échantillonnage sur la disparition des modes grâce à une modélisation issue de la génétique des populations. Ceci a des implications pratiques car la loi de filtrage est multimodale dans les scénarios de recalage de la navigation par mesures altimétriques. Enfin, nous avons exposé les modèles de mélange dans le cadre du filtrage particulaire comme une solution adaptée au distributions multimodales. Le chapitre 4 proposera une mise en oeuvre de ces modèles en proposant une approche basée sur les techniques de clustering pour identifier automatiquement les modes de la loi conditionnelle. Nous proposerons également une méthode de régularisation adaptée aux densités multimodales, appelée régularisation locale.

Chapitre 3

Systèmes de navigation inertielle et recalage radio-altimétrique

3.1 Introduction à la navigation inertielle

La navigation consiste d'une manière générale à connaitre la position, la vitesse et l'attitude d'un mobile dans un référentiel donné. Dans cette thèse nous nous intéressons plus particulièrement à la navigation terrestre, c'est-à-dire la détermination des caractéristiques cinématiques d'un objet par rapport au globe terrestre. Nous chercherons donc à déterminer la position, la vitesse et éventuellement les angles d'attitude de l'objet en déplacement. Les angles d'attitude donnent accès à l'orientation angulaire du solide considéré. L'objet d'intérêt est un aéronef muni d'une centrale inertielle qui comporte des accéléromètres et des gyromètres et d'un calculateur permettant de déduire la solution de navigation par intégration successive des mesures accélérométriques et gyrométriques. Nous présentons ci-après les principaux repères utilisés en navigation terrestre ainsi que les types de centrales inertielles.

3.1.1 Repères de navigation

Repère inertiel (*inertial frame*) Le repère terrestre $(O, \boldsymbol{x_i}, \boldsymbol{y_i}, \boldsymbol{z_i})$ a pour origine le centre de la Terre O et ses axes ont une direction fixe par rapport aux étoiles.

Repère terrestre Le repère terrestre $(O, \boldsymbol{x_T}, \boldsymbol{y_T}, \boldsymbol{z_T})$, également appelé ECEF (Earth-Centered, Earth-Fixed) , est lié à la Terre et est en rotation avec celle-ci. Il est centré sur la Terre et les 3 vecteurs du trièdre $\boldsymbol{x_T}$, $\boldsymbol{y_T}$ et $\boldsymbol{z_T}$ sont tels que :
- $\boldsymbol{x_T}$ est contenu dans le plan équatorial et passe par le méridien de Greenwich,
- $\boldsymbol{z_T}$ est dirigé vers le pôle nord,
- $\boldsymbol{y_T}$ complète le trièdre direct.

Trièdre géographique local (*navigation frame*) Le trièdre géographique local (TGL) est le repère de navigation. Il a pour origine la projection P du centre de masse du mobile sur la surface de la terre et ses axes sont dirigés respectivement vers le Nord, l'Est et selon la verticale descendante. Dans toute la suite on notera $(\boldsymbol{n}, \boldsymbol{e}, \boldsymbol{d})$ ce repère.

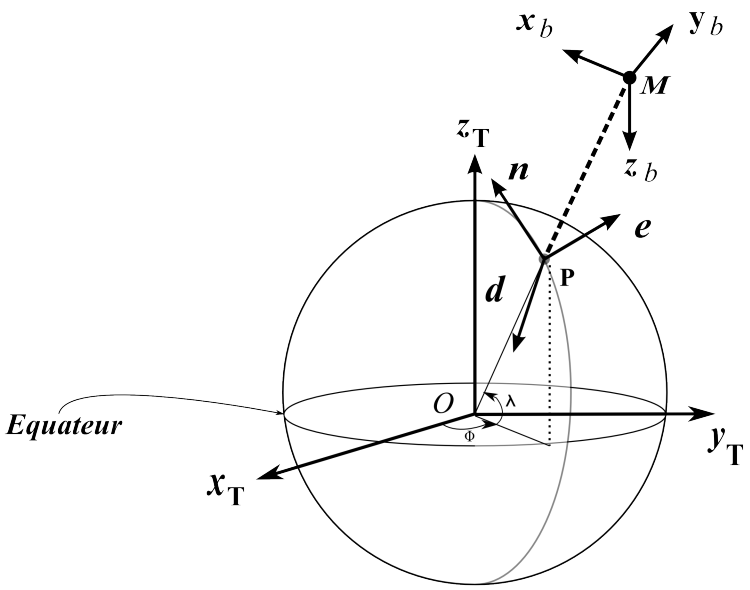


Figure 3.1 – Repères de référence (φ et λ désignent ici respectivement la latitude et la longitude géographique)

Trièdre lié au mobile (*body frame*) Il s'agit du repère $(M, \boldsymbol{x_b}, \boldsymbol{y_b}, \boldsymbol{z_b})$ lié à l'engin. Son origine M est un point quelconque du véhicule et n'est pas nécessairement confondu avec son centre de gravité G. L'orientation des axes est conventionnelle :
- l'axe $\boldsymbol{x_b}$ est dirigé selon l'axe de symétrie longitudinal de l'engin et est orienté vers l'avant,
- l'axe $\boldsymbol{z_b}$ est perpendiculaire au plan de la voilure et dirigé vers l'intrados du corps,
- l'axe $\boldsymbol{y_b}$ complète le trièdre direct.

Les angles d'attitudes, notés φ, θ et ψ, définissent l'orientation du mobile par rapport au TGL. φ désigne le roulis (*roll*), θ le tangage (*pitch*) et ψ le lacet (*yaw*).

3.1.2 Capteurs inertiels

Il existe deux types de centrales inertielles qui se distinguent par la façon dont la plate-forme de capteurs est orienté par rapport au boitier : les centrales à plate-forme stabilisée et les centrales à composants liés. Les 2 types de centrales comportent 3 accéléromètres et 3 gyromètres. Les accéléromètres fournissent l'accélération non-gravitationnelle du véhicule par rapport au repère inertiel, c'est-à-dire l'accélération absolue moins la pesanteur. Le vecteur résultant est appelé force spécifique $\boldsymbol{f}_a$. Les gyromètres mesurent la vitesse angulaire de l'engin par rapport au repère inertiel.

Centrales à plate-forme stabilisé (gyrostabilized platforms) Sur ce type de centrales, les accéléromètres sont installés sur une plate-forme mécanique isolée des mouvements angu-

laires du boitier. Les gyromètres permettent d'asservir la plate-forme en position horizontale de telle sorte que les accéléromètres ont leurs axes parallèles au TGL [106]. Les centrales à plate-forme stabilisée sont relativement précises mais la présence d'un système mécanique d'asservissement les rend plutôt encombrantes.

Centrales à composants liées Dans ce type de centrale, également appelée centrale inertielle *strap-down*, les capteurs inertiels sont directement fixés à la structure rigide du boitier de la centrale. Les 3 gyromètres et les 3 accéléromètres fournissent donc des mesures dans le repère engin $(G, \boldsymbol{x_b}, \boldsymbol{y_b}, \boldsymbol{z_b})$ lié au mobile. Les accélérations sont projetées dans le TGL par intégration des vitesses angulaires issues des gyromètres. Les centrales à composants liés sont plus simples mécaniquement que les centrales à plate-forme stabilisée, ce qui les rend peu encombrantes et exige moins de maintenance mais nécessite un calculateur plus puissant. Aujourd'hui, la puissance des calculateurs embarqués fait qu'elles sont plus répandues que les centrales à plate-forme stabilisée.

Dans la suite de l'exposé, nous considérons un véhicule muni d'une centrale à composants liés. Le rôle du calculateur est de projeter les mesures des capteurs inertiels dans le trièdre géographique local afin d'obtenir par intégration les paramètres cinématiques du véhicule.

3.2 Les équations de la navigation inertielle

Les paramètres cinématiques de l'aéronef sont :

- la position géographique (λ, ϕ, z) où λ est la latitude géographique, ϕ la longitude et z l'altitude. Ces 3 composantes sont exprimées dans le trièdre de navigation.
- la vitesse de déplacement de son centre de masse, exprimée dans le TGL et notée $\boldsymbol{v} = (v_n, v_e, v_d)$
- l'attitude, représentée par les angles d'Euler φ, θ et ψ.

Dans une centrale inertielle à composants liés, les mesures de forces spécifiques et de vitesses de rotation se font dans le trièdre engin $(M, \boldsymbol{x_b}, \boldsymbol{y_b}, \boldsymbol{z_b})$, M étant l'origine de la centrale inertielle. La vitesse du centre de masse $\boldsymbol{v_G}$ est reliée à la vitesse du point M par la relation :

$$\boldsymbol{v_G} = \boldsymbol{v_M} - (\boldsymbol{\omega_m} - \boldsymbol{\omega_{ie}}) \wedge \boldsymbol{GM} \tag{3.1}$$

où $\boldsymbol{\omega_{ie}}$ est le vecteur de la rotation de la terre et $\boldsymbol{\omega_m}$ est le vecteur de vitesse angulaire mesuré par les gyromètres (vitesse de rotation absolue).
Afin de déterminer l'ensemble des paramètres cinématiques de l'aéronef, on représente le globe terrestre comme un ellipsoïde de révolution dont l'axe de symétrie est confondu avec l'axe de révolution autour des pôles. Cet ellipsoïde est défini par son demi grand axe a et son demi petit axe b. Son excentricité e et son aplatissement f_{ap} sont définis par :

$$e = \sqrt{1 - \frac{b^2}{a^2}} \text{ et } f_{ap} = 1 - \frac{b}{a}$$

Les paramètres de l'ellipsoïde sont fournies par le modèle WGS84 (*World Geodetic System 1984*). Les valeurs numériques sont :

$$a = 6378137\,m \tag{3.2}$$
$$f_{ap} = 1/298.257223563 \tag{3.3}$$
$$b = 6356752.3\,m \tag{3.4}$$

3.2.1 Equation d'attitude

L'orientation du mobile par rapport au TGL est définie par les angles d'attitudes φ, θ et ψ (respectivement roulis, tangage et lacet). Ces angles permettent d'obtenir la matrice des cosinus directeurs R_{n2b} qui correspond à la matrice de rotation qui fait passer du TGL $(\boldsymbol{n}, \boldsymbol{e}, \boldsymbol{d})$ au repère engin $(\boldsymbol{x_b}, \boldsymbol{y_b}, \boldsymbol{z_b})$.

$$R_{n2b} = \begin{pmatrix} \cos\theta\cos\psi & \cos\theta\sin\psi & -\sin\theta \\ -\cos\varphi\sin\psi + \sin\varphi\sin\theta\cos\psi & \cos\varphi\cos\psi + \sin\varphi\sin\theta\sin\psi & \sin\varphi\cos\theta \\ \sin\varphi\sin\psi + \cos\varphi\sin\theta\cos\psi & -\sin\varphi\cos\psi + \cos\varphi\sin\theta\sin\psi & \cos\varphi\cos\theta \end{pmatrix} \tag{3.5}$$

On définit également la matrice de rotation R_{b2n} qui fait passer du repère engin au repère de navigation. On a donc :

$$R_{b2n} = R_{n2b}^T = R_{n2b}^{-1}$$

Soit $\boldsymbol{\omega}$ la vitesse angulaire de rotation du mobile par rapport au repère TGL et soit p, q, r ses composantes exprimées dans le repère lié à l'avion. Les angles d'attitude sont reliés aux vitesses angulaires p, q, r selon l'équation différentielle suivante :

$$\begin{pmatrix} \dot{\psi} \\ \dot{\theta} \\ \dot{\varphi} \end{pmatrix} = \begin{pmatrix} 0 & \frac{\sin\varphi}{cos\theta} & \frac{\cos\varphi}{cos\theta} \\ 0 & \cos\varphi & -\sin\varphi \\ 1 & \tan\theta\sin\varphi & \tan\theta\cos\varphi \end{pmatrix} \begin{pmatrix} p \\ q \\ r \end{pmatrix} \tag{3.6}$$

La vitesse angulaire de l'aéronef par rapport au TGL, est égale à

$$\boldsymbol{\omega} = \boldsymbol{\omega_m} - \boldsymbol{\rho} - \boldsymbol{\omega_{ie}}$$

$\boldsymbol{\omega_m}$ est la vitesse angulaire mesurée par les gyromètres dans le repère engin tandis que $\boldsymbol{\rho}$ représente la vitesse angulaire de rotation du TGL par rapport à la Terre. Les composantes de $\boldsymbol{\rho}$ et $\boldsymbol{\omega_{ie}}$ sont définies dans le repère TGL selon :

$$\begin{cases} \rho_n &= \frac{v_e}{R_\phi + z} \\ \rho_e &= -\frac{v_n}{R_\lambda + z} \\ \rho_d &= -\frac{v_e}{R_\phi + z}\tan(\lambda) \end{cases} \tag{3.7}$$

et

$$\begin{cases} \omega_{ie}^n &= \omega_0\cos(\lambda) \\ \omega_{ie}^e &= 0 \\ \omega_{ie}^d &= -\omega_0\sin(\lambda) \end{cases} \tag{3.8}$$

avec $\omega_0 = 7.29 \times 10^5\, rd.s^{-1}$ est la vitesse de rotation propre du globe terrestre.
R_λ est le rayon de courbure de la Terre dans le plan méridien (rayon nord).

$$R_\lambda = a\frac{1-e^2}{(1-e^2\sin^2(\lambda))^{3/2}} \tag{3.9}$$

R_ϕ est la grande normale de l'ellipsoïde (rayon est).

$$R_\phi = \frac{a}{(1-e^2\sin^2(\lambda))^{1/2}} \tag{3.10}$$

L'équation du mouvement angulaire d'un corps rigide en rotation permet d'obtenir l'équation différentielle de la matrice des cosinus directeurs :

$$\dot{R}_{n2b} = -[\boldsymbol{\omega_m}\times]R_{n2b} + R_{n2b}[(\boldsymbol{\rho}+\boldsymbol{\omega_{ie}})\times] \tag{3.11}$$

où $\omega_m = \begin{pmatrix}\omega_x & \omega_y & \omega_z\end{pmatrix}^T$ est le vecteur des vitesses angulaires mesurées par les gyromètres et $[\boldsymbol{\omega_m}\times]$ est la matrice anti-symétrique définie par :

$$[\boldsymbol{\omega_m}\times] = \begin{pmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & -\omega_x & 0 \end{pmatrix} \tag{3.12}$$

3.2.2 Equation de vitesse du mobile

L'équation différentielle régissant la vitesse du mobile et projetée dans le trièdre de navigation est

$$\dot{\boldsymbol{v}} = R_{b2n}\boldsymbol{\gamma_m} + \boldsymbol{g} - (2\boldsymbol{\omega_{ie}} + \boldsymbol{\rho}) \wedge \boldsymbol{v} \tag{3.13}$$

$\boldsymbol{\gamma_m}$ est le vecteur d'accélération spécifique mesurée par les accéléromètres exprimé dans le repère engin. $\boldsymbol{g}$ est la pesanteur terrestre. Il est égal à la gravité $\boldsymbol{\Phi}_T$ de la Terre diminué de l'accélération d'entraînement due à la rotation de celle-ci.

$$\boldsymbol{g} = \boldsymbol{\Phi_T} - \boldsymbol{\omega_{ie}} \wedge (\boldsymbol{\omega_{ie}} \wedge \boldsymbol{OM}) \tag{3.14}$$

Dans le modèle WGS84, la norme de $\boldsymbol{g}$ à la surface de la Terre ne dépend que de la latitude λ et de l'altitude z via la formule suivante :

$$g \approx 9.7803 + 0.0519\sin^2(\lambda) - 3.08\times 10^{-6} z \tag{3.15}$$

où z est exprimé en mètres.

En projetant l'équation (3.13) sur le TGL, et compte tenu des égalités (3.7) et (3.8), il vient :

$$\begin{pmatrix}\dot{v}_n \\ \dot{v}_e \\ \dot{v}_d\end{pmatrix} = R_{b2n}\begin{pmatrix}\gamma_x \\ \gamma_y \\ \gamma_z\end{pmatrix} + \begin{pmatrix}0 \\ 0 \\ g\end{pmatrix} + \begin{pmatrix} -\frac{v_e^2}{R_\phi+z}\tan\lambda - 2\omega_0 v_e\sin\lambda + \frac{v_n v_d}{R_\lambda+z} \\ \frac{1}{R_\phi+z}(v_e v_n\tan\lambda + v_e v_d) + 2\omega_0(v_n\sin\lambda + v_d\cos\lambda) \\ -\frac{v_n^2}{R_\phi+z} - \frac{v_e^2}{R_\lambda+z} - 2\omega_0 v_e\cos\lambda \end{pmatrix} \tag{3.16}$$

où $\begin{pmatrix}\gamma_x & \gamma_y & \gamma_z\end{pmatrix}^T$ est le vecteur des accélérations mesurées par les accéléromètres.

3.2.3 Equation de position

L'équation d'évolution de la position du mobile s'obtient par intégration de l'équation de vitesse (3.16) :

$$\begin{pmatrix} \dot{\lambda} \\ \dot{\phi} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} \frac{1}{R_\lambda + z} & 0 & 0 \\ 0 & \frac{1}{(R_\phi + z)\cos(\lambda)} & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} v_n \\ v_e \\ v_d \end{pmatrix} \tag{3.17}$$

3.3 Modélisation de l'erreur de navigation

La centrale inertielle calcule la solution de navigation par intégration successives des données accélérométriques et gyrométriques. Des erreurs peuvent apparaître lorsque l'initialisation (alignement) de la centrale est erronée ou inévitablement parce que les capteurs ont une précision limitée.
Les erreurs de navigation sont dues à la juxtaposition d'erreurs de projection et d'erreurs d'intégration. En effet, une mesure erronée des gyromètres donnera une matrice d'attitude R_{n2b} fausse. La projection des mesures des accéléromètres se faisant grâce à cette matrice d'attitude, il en résultera une accumulation de l'erreur due à une mauvaise projection et de celle due à l'imperfection des accéléromètres. L'intégration de cette projection entraînera une dérive des composantes de vitesse et donc des coordonnées de position. Par ailleurs, le modèle de pesanteur que nous avons considéré dépend, en ce qui concerne la direction, de la position horizontale du véhicule, et en ce qui concerne l'intensité, de son altitude. Toute erreur sur la position du véhicule entraîne donc une mauvaise estimation de la pesanteur et se répercute sur les futures estimations de vitesses et de position. Ce couplage, appelé phénomène de Schüler, se traduit par une oscillation des erreurs de position horizontale et une divergence de la position verticale. Nous rappelons ci-après les équations d'évolution des erreurs de position, vitesse et attitude.

3.3.1 Variables d'erreurs

Dans la suite, les variables entachées d'erreurs seront notées par un tilde. Ces variables représentent par exemple la position inertielle ou les mesures des accéléromètres. On notera δu le vecteur d'erreur correspondant à la différence entre la valeur vraie ou idéale de u et la valeur erronée $\tilde{u}$.

$$\delta u = u - \tilde{u}$$

Erreur de position On définit l'erreur de position $\delta \boldsymbol{x}$ comme l'écart métrique entre la position inertielle et la position vraie dans le repère TGL. Ses coordonnées sont :

$$\begin{cases} \delta x_n &= (R_\lambda + z)\delta\lambda \\ \delta x_e &= (R_\phi + z)\cos(\lambda)\delta\phi \\ \delta x_d &= -\delta z \end{cases} \tag{3.18}$$

où $\delta\lambda$, $\delta\phi$ et δz sont les erreurs de position géographiques dans le trièdre géographique local.

Erreur de vitesse On définit $\delta \boldsymbol{v} = \boldsymbol{v} - \tilde{\boldsymbol{v}}$ comme l'erreur de vitesse dans le repère TGL.

Erreur d'attitude Soit $\delta\phi$, $\delta\theta$ et $\delta\psi$ les erreurs commises sur les angles d'attitude. Si on note $\tilde{R}_{n2b}$ la matrice des cosinus directeurs dont dispose la centrale inertielle (R_{n2b} étant la matrice exacte), on a la relation suivante [5], valable dans l'approximation des petits angles :

$$R_{n2b} = \tilde{R}_{n2b} \begin{pmatrix} 1 & -\delta\psi & \delta\theta \\ \delta\psi & 1 & -\delta\varphi \\ -\delta\theta & \delta\varphi & 1 \end{pmatrix} \tag{3.19}$$

On définit la matrice antisymétrique $[\boldsymbol{\Phi}\times]$ comme

$$[\boldsymbol{\Phi}\times] = \begin{pmatrix} 0 & -\delta\psi & \delta\theta \\ \delta\psi & 0 & -\delta\varphi \\ -\delta\theta & \delta\varphi & 0 \end{pmatrix}$$

L'équation (3.19) est donc équivalente à :

$$R_{n2b} = \tilde{R}_{n2b}(I_3 + [\boldsymbol{\Phi}\times]) \tag{3.20}$$

Erreurs de mesure des capteurs inertiels On définit l'erreur de mesure des accéléromètres comme

$$\delta\boldsymbol{\gamma} = \boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}} \tag{3.21}$$

où $\tilde{\boldsymbol{\gamma}}$ est la mesure accélérométrique fournie par la centrale inertielle. De la même façon l'erreur de mesure des gyromètres est :

$$\delta\boldsymbol{\omega} = \boldsymbol{\omega} - \tilde{\boldsymbol{\omega}} \tag{3.22}$$

Erreur d'orientation du TGL Soit B la matrice de rotation qui fait passer du repère terrestre au repère de navigation. B dépend de la latitude et longitude du véhicule de la façon suivante :

$$B = \begin{pmatrix} -\cos\phi\sin\lambda & -\sin\phi\sin\lambda & \cos\lambda \\ -\sin\phi & \cos\phi & 0 \\ -\cos\phi\cos\lambda & -\sin\phi\cos\lambda & -\sin\lambda \end{pmatrix} \tag{3.23}$$

Ainsi, une erreur sur les coordonnées géographiques a une répercussion sur la matrice B de sorte que la matrice $\tilde{B}$ dont dispose la centrale s'exprime sous la forme.

$$B = (I - [\delta\boldsymbol{\Theta}\times])\tilde{B} \tag{3.24}$$

avec

$$[\delta\boldsymbol{\Theta}\times] = \begin{pmatrix} 0 & -\delta\Theta_n & \delta\Theta_e \\ \delta\Theta_d & 0 & -\delta\Theta_n \\ -\delta\Theta_e & \delta\Theta_n & 0 \end{pmatrix} \tag{3.25}$$

$\delta\Theta_n$, $\delta\Theta_e$ et $\delta\Theta_d$ sont trois petits angles décrivant l'erreur d'orientation du TGL en fonction de l'erreur de position :

$$\begin{cases} \delta\Theta_n &= \frac{\delta x_e}{R_\phi + z} \\ \delta\Theta_e &= -\frac{\delta x_n}{R_\lambda + z} \\ \delta\Theta_d &= -\frac{\tan(\lambda)\delta x_e}{R_\phi + z} \end{cases}$$

Nous donnons ci-après les équations de propagation des erreurs de navigation. Elles sont obtenues par différentiation au premier ordre des équations de la navigation. Dans un premier temps, ces équations correspondent à un modèle dit *angle-phi*, pour lequel les variables d'erreurs $\delta\boldsymbol{x}$, $\delta\boldsymbol{v}$ sont projetées dans le repère lié à la Terre [5]. Ensuite, nous présentons la modélisation dite *angle-psi* [5], qui permet d'éliminer certaines variables intermédiaires du problème.

3.3.2 Équation d'évolution de l'erreur d'attitude

Rappelons que $\boldsymbol{\Phi} = \begin{pmatrix} \delta\phi & \delta\theta & \delta\psi \end{pmatrix}^T$ est le vecteur des erreurs d'attitude en repère TGL. Notons $\boldsymbol{\epsilon_g} = R_{b2n}\delta\boldsymbol{\omega_m}$ la projection des erreurs des gyromètres dans le repère TGL. On introduit également $\delta\boldsymbol{\rho}$ et $\delta\boldsymbol{\omega_{ie}}$ les erreurs d'estimation de vitesse angulaire de $\boldsymbol{\rho}$ et $\boldsymbol{\omega_{ie}}$. L'équation d'évolution du vecteur des erreurs d'attitude est alors

$$\dot{\boldsymbol{\Phi}} = -(\boldsymbol{\rho} + \boldsymbol{\omega_{ie}}) \wedge \boldsymbol{\Phi} - \boldsymbol{\epsilon_g} + \delta\boldsymbol{\rho} + \delta\boldsymbol{\omega_{ie}} \tag{3.26}$$

3.3.3 Équation d'évolution de l'erreur de vitesse

Soit $\boldsymbol{f} = R_{b2n}\boldsymbol{\gamma_m}$ et $\boldsymbol{\epsilon_a} = R_{b2n}\delta\boldsymbol{\gamma_m}$ les projections respectives de la mesure accélérométrique et de l'erreur correspondante dans le TGL. L'équation d'évolution du vecteur des erreurs de vitesse projetée dans le TGL est

$$\delta\dot{\boldsymbol{v}} = -\boldsymbol{\Phi} \wedge \boldsymbol{f} + \boldsymbol{\epsilon_a} + \delta\boldsymbol{g} - (\boldsymbol{\rho} + 2\boldsymbol{\omega_{ie}}) \wedge \delta\boldsymbol{v} - (\delta\boldsymbol{\rho} + 2\delta\boldsymbol{\omega_{ie}}) \wedge \boldsymbol{v} \tag{3.27}$$

où $\delta\boldsymbol{g} \approx \begin{pmatrix} 0 & 0 & -\frac{2g}{a}\delta z \end{pmatrix}^T$ est le vecteur d'erreur de la pesanteur qui a une contribution nulle dans le plan horizontal. De plus, la composante sur l'axe vertical du TGL, g_D, a une erreur δg_D qui s'exprime comme

$$\delta g_D = \frac{\partial g}{\partial z}\delta z + \frac{\partial g}{R_\phi \partial \lambda} R_\phi \delta\lambda$$

Une analyse numérique montre que le terme $\frac{\partial g}{R_\phi \partial \lambda} R_\phi \delta\lambda$ peut être négligé, ce qui réduit δg_D à $\frac{\partial g}{\partial z}\delta z$.

3.3.4 Équation d'évolution de l'erreur de position

L'équation d'évolution de l'erreur de position fait intervenir un terme dû à l'erreur d'estimation de la vitesse et un terme dû à une mauvaise estimation du TGL.

$$\delta\dot{\boldsymbol{x}} = \delta\boldsymbol{v} - \boldsymbol{\rho} \wedge \delta\boldsymbol{x} + \delta\boldsymbol{\Theta} \wedge \boldsymbol{v} \tag{3.28}$$

3.3.5 Modélisation *angle-psi* pour les équation d'erreurs de navigation

Une analyse des équations (3.26), (3.27) et (3.28) montre que les variables $\delta\boldsymbol{\rho}$, $\delta\boldsymbol{\omega}_{ie}$ et $\delta\boldsymbol{\Theta}$ doivent être exprimées en fonctions des paramètres cinématiques $\delta\boldsymbol{x}$, $\delta\boldsymbol{v}$ et $\delta\boldsymbol{\Phi}$ afin d'être intégrées, ce qui est possible, mais résulte en un système d'équations différentielles couplées peu commode à manipuler. La modélisation dite *angle-psi* permet d'obtenir un ensemble d'équations plus simple d'utilisation en introduisant les variables suivantes :

$$\begin{array}{rcl} \delta\boldsymbol{v'} & = & \delta\boldsymbol{v} + \delta\boldsymbol{\Theta} \wedge \boldsymbol{v} \\ \boldsymbol{\Psi} & = & \boldsymbol{\Phi} - \delta\boldsymbol{\Theta} \end{array} \tag{3.29}$$

On aboutit alors au système d'équations suivant [28] :

$$\left\{ \begin{array}{rcl} \dot{\boldsymbol{\Psi}} & = & -(\boldsymbol{\rho} + \boldsymbol{\omega}_{ie}) \wedge \boldsymbol{\Psi} - \boldsymbol{\epsilon_g} \\ \delta\dot{\boldsymbol{v'}} & = & -\boldsymbol{\Psi} \wedge \boldsymbol{f} + \boldsymbol{\epsilon_a} + \delta\boldsymbol{g'} - (\boldsymbol{\rho} + 2\boldsymbol{\omega}_{ie}) \wedge \delta\boldsymbol{v'} \\ \delta\dot{\boldsymbol{x}} & = & \delta\boldsymbol{v'} - \boldsymbol{\rho} \wedge \delta\boldsymbol{x} \end{array} \right. \tag{3.30}$$

3.3.5.1 Modélisation des erreurs des capteurs inertiels

L'erreur accélérométrique $\delta\boldsymbol{\gamma}_m$ est modélisée de la manière suivante :

$$\delta\boldsymbol{\gamma}_m = \boldsymbol{b_a} + K_a\boldsymbol{\gamma}_m + \boldsymbol{\xi}_a \tag{3.31}$$

où

- $\boldsymbol{b_a}$ est un terme de biais accélérométrique,
- K_a est un facteur d'échelle,
- $\boldsymbol{\xi}_a$ est un processus de Wiener [38].

L'erreur gyrométrique est modélisée de la même façon :

$$\delta\boldsymbol{\omega}_m = \boldsymbol{b_g} + K_g\boldsymbol{\omega}_m + \boldsymbol{\xi}_g \tag{3.32}$$

- $\boldsymbol{b_g}$ est un terme de biais gyrométrique,
- K_g est un facteur d'échelle propre au gyromètre,
- $\boldsymbol{\xi}_g$ est un processus de Wiener.

Les biais $\boldsymbol{b_a}$ et $\boldsymbol{b_g}$ sont communément modélisés par un processus de Markov du premier ordre noté b :

$$\dot{\boldsymbol{b}} = \frac{-1}{\tau}\boldsymbol{b} + \boldsymbol{\nu} \tag{3.33}$$

où

- τ est la période de corrélation du processus $\boldsymbol{b}$,
- $\boldsymbol{\nu}$ est un processus de Wiener.

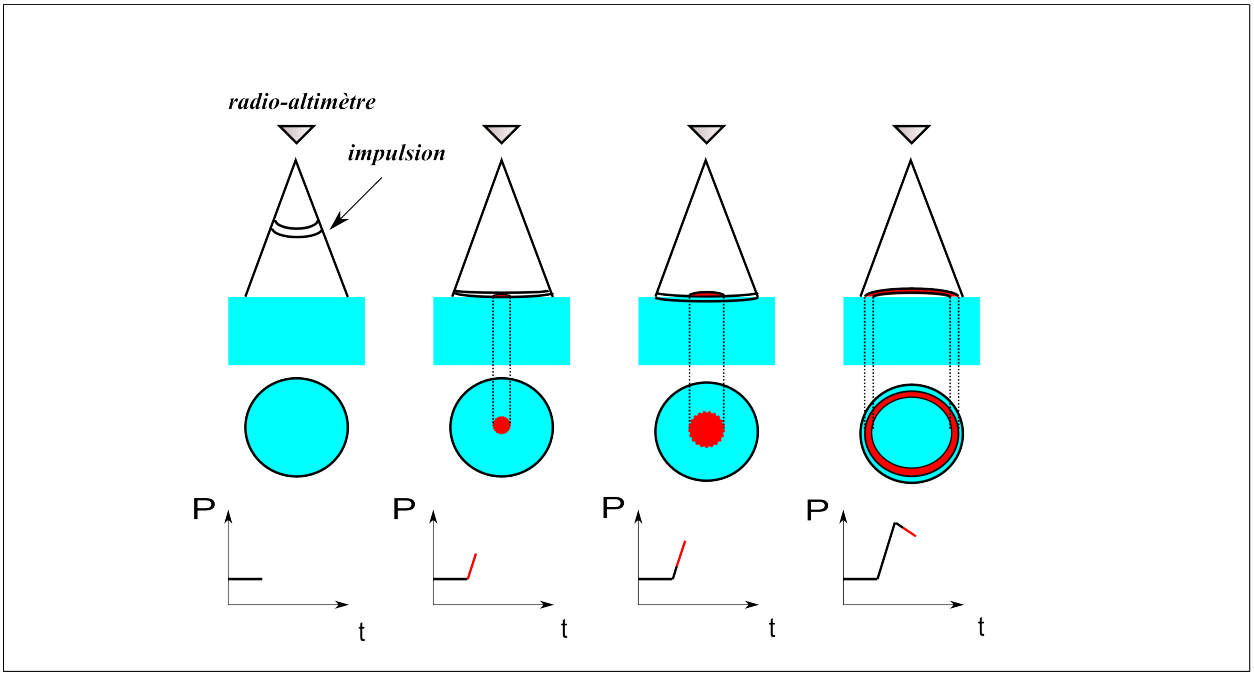


Figure 3.2 – Mesure altimétrique sur une surface plane : évolution de la puissance P reçue par le radar

3.4 Radio-altimétrie et architecture de la navigation hybridée

3.4.1 Généralités sur les radio-altimètres

Un radio-altimètre permet de mesurer la distance entre un mobile donné et le sol, appelée hauteur-sol. L'altimètre émet une onde électromagnétique et reçoit le signal renvoyé par les cibles situées sur le trajet de l'onde (surfaces terrestres) avec un décalage temporel. La distance h entre l'objet en vol (porteur de l'instrument) et le sol est déduite du temps de propagation aller-retour (t_0) d'une onde électromagnétique (de célérité c) entre l'objet et la surface du sol par la relation :

$$h = \frac{ct_0}{2} \tag{3.34}$$

Il existe deux classes de fonctionnement pour les radio-altimètres, qui se distinguent par la largeur de la zone éclairée au sol par son faisceau d'antenne :

- les altimètres pour lesquels la surface éclairée au sol est limitée par l'ouverture θ_0 du faisceau de l'antenne (*beam-limiter altimeters*),
- les altimètres dont la durée d'impulsion courte limite la surface de la zone de sol reçue dans une cellule de résolution (*pulse-limited altimeter*).

La figure 3.2 représente l'évolution d'une impulsion électro-magnétique et la courbe d'évolution de la puissance reçue par le radar, pour une mesure altimétrique effectuée sur une surface

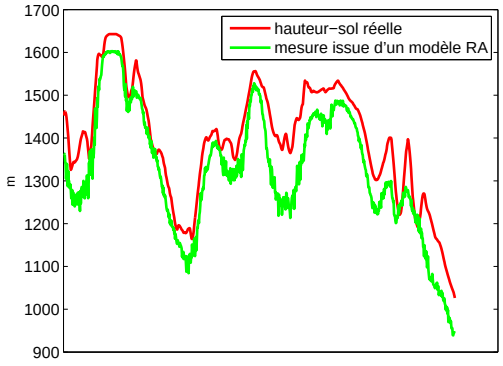


FIGURE 3.3 – Comparaison de la hauteur-sol réelle et d'une mesure simulée à partir d'un modèle d'antenne RA réaliste. Le profil considéré est une portion de terrain de 20 km.

plane (un océan,...) dans le cas d'un radio-altimètre à durée d'impulsion courte. La surface éclairée par l'onde est l'intersection de la coquille sphérique, qui représente l'impulsion radar, par la surface observée. L'impulsion émise ayant une durée finie, cette surface passe d'un point à un disque, puis à un anneau de surface constante. La puissance renvoyée par la surface est plus faible à des angles d'incidences plus grands, d'où la décroissance observée de la puissance reçue.

Le radio-altimètre a tendance à indiquer une distance qui diffère de la hauteur-sol d'autant plus que le lobe de l'antenne radar est large, même s'il est orienté verticalement vers le sol. Une illustration de cette observation est donnée par la figure 3.3 où est représentée la hauteur-sol théorique le long d'une trajectoire donnée et la mesure issue d'un modèle radio-altimétrique réaliste. La distance indiquée est plutôt la distance la plus courte entre le mobile et le relief ce qui introduit un biais. Idéalement ceci doit être pris en compte dans le développement de filtres de recalage réalistes en modifiant le modèle d'observation décrivant la mesure altimétrique. Cependant dans cette thèse, la modélisation retenue ne tient pas compte de cette erreur systématique.

3.4.2 Principe de la navigation hybridée

La navigation hybridée consiste à coupler un capteur indépendant des senseurs inertiels avec la centrale inertielle afin d'augmenter la précision et la fiabilité de la solution de navigation. Une architecture d'hybridation permet donc d'effectuer un recalage de la navigation inertielle. Parmi les capteurs couramment utilisées pour le recalage d'aéronef, on peut citer :

- le GPS (Global Positioning System), système de positionnement par satellites fondé sur la triangulation. Le système GPS est géré par les autorités américaines et assure une précision dans le pire des cas de 7,8 mètres à 95% de certitude (GPS civil) [1].
- le radar à synthèse d'ouverture (*SAR, Synthetic Aperture Radar*) qui permet d'obtenir une cartographie de la zone de relief survolé et d'effectuer le recalage par comparaison avec une carte de référence embarquée [81].
- le radio-altimètre, présenté précédemment, offre la possibilité de se recaler en comparant les mesures successives de hauteur-sol avec un *Modèle Numérique de Terrain (MNT)*

embarqué.

Les avantages principaux du radio-altimètre par rapport au GPS sont essentiellement son fonctionnement autonome et la difficulté de brouillage des signaux issue de son antenne-radar.

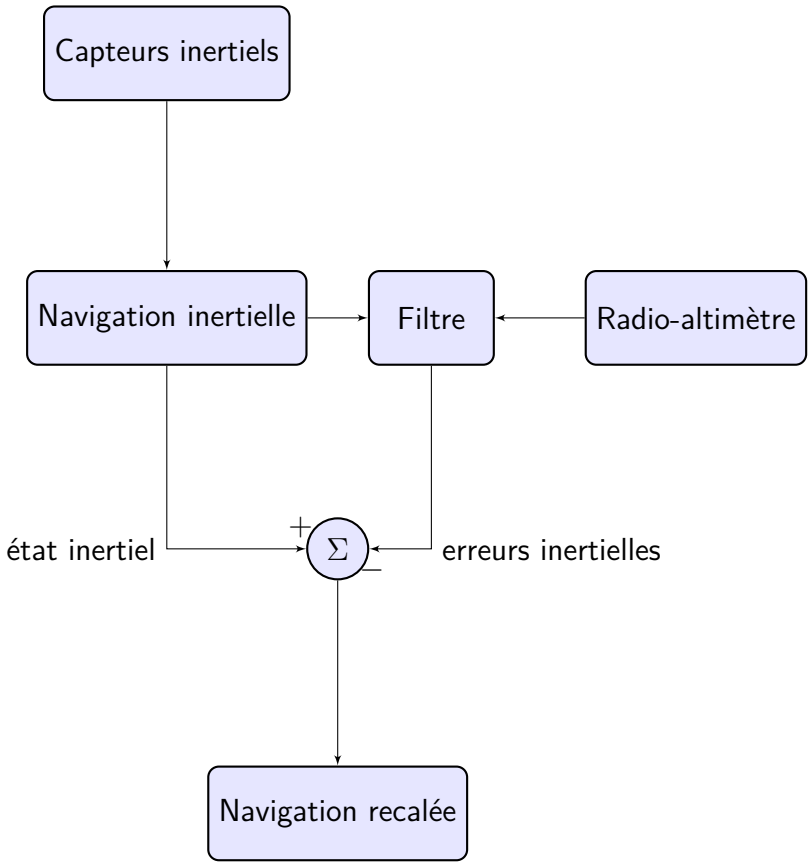


Figure 3.4 – Architecture de navigation hybride

3.5 Approches algorithmiques pour le recalage altimétrique

3.5.1 Recalage altimétrique par bloc

Le recalage altimétrique par bloc consiste à acquérir une suite de mesures altimétriques afin d'inférer la position courante par comparaison de profils d'altitude. Parmi les approches de recalage par bloc, on peut citer le procédé TERCOM (Terrain Contour Matching) fondé sur une mesure de corrélation entre un vecteur de mesures de hauteurs-sol et un profil de terrain de même longueur. Si $h = (h_1, \ldots, h_p)$ est une suite de p mesures RA consécutives et $\boldsymbol{H}$ désigne la matrice représentant le modèle numérique de terrain, une mesure de corrélation possible est [6] :

$$c_{k,m} = \frac{1}{p} \sum_{n=1}^{p} |h_n - \boldsymbol{H}_{m,n+k}| \tag{3.35}$$

La position estimée est celle maximisant la corrélation $c_{k,m}$:

$$(\hat{k}, \hat{m}) = \underset{(k,m)}{\arg\max}\, c_{k,m} \tag{3.36}$$

Le système TERCOM est applicable aux véhicules dont les manoeuvres sont rares et limitées ce qui contraint son utilisation. De plus, il suppose une connaissance précise des vitesses de l'aéronef.

3.5.2 Recalage par banc de filtres de Kalman

Le système SITAN (Sandia Inertial Terrain Aided Navigation), développé à la fin des années 70 par Sandia Laboratories, est le premier à adopter une approche récursive au problème de navigation par corrélation de terrain, contrairement à TERCOM qui procède par "batch". La version d'origine utilise un filtre de Kalman étendu modifié, dans lequel est introduit une linéarisation stochastique adaptative censée atténuer les effets des fortes non-linéarités du relief [57]. Par la suite, des modifications ont été apportées à l'algorithme afin d'éviter les divergences du filtre : dans la version de Hollowel et al. [56], un banc de filtre de Kalman remplace l'unique EKF. L'idée est que les filtres initialisés au voisinage de la position vraie ont moins de chance de diverger.

3.5.3 Du Point-Mass Filter aux filtres particulaires

Le Point-Mass Filter (PMF) [11, 3], est un filtre bayésien capable de traiter les ambiguïtés de terrain inévitable en début de recalage lorsque l'incertitude est grande. A notre connaissance, la première application du PMF au recalage altimétrique date des travaux de Bergman sur l'estimation bayésienne pour la navigation et le pistage [7]. Le PMF suppose que l'espace d'état soit discrétisé selon une grille déterministe $x^{(1)}, \ldots, x^{(N)}$. Les équations du filtrage linéaire sont alors exprimées sur cette grille en introduisant des poids $\omega_k^{(1)}, \ldots, \omega_k^{(N)}$ qui représente les probabilités $p(x_k = x^{(i)}|y_{0:k})$, $i = 1, \ldots, N$. En pratique, le PMF peut s'implémenter aisément pour des vecteurs de petite dimension, c'est-à-dire inférieure ou égale à 3 : au delà, le maillage régulier de l'espace devient inefficace. Ainsi, dans [7], les auteurs s'intéressent au recalage de la position horizontale, ayant supposé que l'altitude est fournie par un baro-altimètre. En outre, le calcul récursif des poids devient extrêmement couteux puisqu'il revient à une convolution discrète en dimension d [51]. Concrètement cela signifie que le PMF est envisageable pour le recalage de véhicules dont les trois composantes de vitesse sont connues avec une précision suffisante. Une autre problématique du PMF est celle de l'adaptation de la résolution du maillage par rapport au support de la densité a posteriori $p(x_k|y_{0:k})$: en début de recalage, l'incertitude sur x est grande donc une maille de faible résolution suffit tandis que dès que $p(x_k|y_{0:k})$ se résume à quelques modes, il vaut mieux augmenter la résolution de la maille.
Bergman a proposé dans ses travaux de thèse [6] la première étude du filtre particulaire standard (SIR) pour le recalage radio-altimétrique. Plus récemment, on peut citer les travaux de :

- K.Dahia [28] qui a développé et appliqué le KPKF (Kernel Particle Kalman Filter) présenté au chapitre 2 au recalage d'aéronefs. Le KPKF combine la philosophie des méthodes particulaires classiques avec le filtre de Kalman, mais contrairement au RBPF

(Rao-Blackwellized Particle Filter), il ne suppose pas de structure partiellement linéaire de l'équation de mesure. L'avantage algorithmique du KPKF est sa robustesse aux grandes incertitudes initales.
- M. Flament [41] a proposé une étude comparative des filtres particulaires gaussiens avec le filtre de Kalman étendu et le Point Mass Filter.
- La thèse de Nordlund [93] propose une étude des performances du RBPF pour le recalage altimétrique.

3.5.4 Recalage altimétrique continu

Le recalage altimétrique continu (RAC) est une technique de recalage utilisée lorsque l'erreur de position initiale a été réduite par un algorithme de type TERCOM, SITAN ou particulaire. Le RAC s'appuie sur un filtre de Kalman étendu pour estimer les variables d'état du véhicule : les paramètres cinématiques (3 erreurs de position , 3 erreurs de vitesse et 3 erreurs d'angles d'attitude) et éventuellement les défauts des capteurs (bias gyrométriques et accélérométriques).

Conclusion du Chapitre 3

Dans ce chapitre, la problématique de navigation inertielle est introduite. Après avoir rappelé les principales équations régissant l'évolution des paramètres cinématiques, nous avons exposé les équations d'évolution des erreurs de la navigation inertielle. La navigation hybridée permet de fusionner les informations fournies par la centrale inertielle avec le filtre de recalage afin de fournir un estimation recalée des variables d'états du véhicule. Le chapitre suivant est consacré au développement d'un nouveau filtre particulaire de recalage permettant de lutter contre les difficultés soulevées par les ambiguïtés de terrain.

Chapitre 4

Développement de filtres particulaires adaptés aux multimodalités

Introduction

Nous avons vu dans le chapitre 2 que l'une des limitations du filtre particulaire concernait la gestion des lois conditionnelles multimodales. En effet, la phase de ré-échantillonnage peut provoquer la perte de l'un des modes de la loi conditionnelle (voir 2.5). Dans l'application du recalage par mesures altimétriques, la présence de plusieurs modes dans la loi a posteriori $p_{k|k} = p(x_k|y_{0:k})$ est due à l'ambiguïté du terrain survolé. La perte du mode correspondant à la vraie position provoquant quasi-systématiquement la divergence du filtre particulaire, il est souhaitable de mettre en oeuvre un algorithme permettant de limiter cette occurrence.
Dans cette thèse, nous proposons d'adopter la représentation sous-forme d'un modèle de mélange de lois conditionnelles, c'est-à-dire que le filtre $p_{k|k}$ est formulé sous la forme suivante :

$$p_{k|k} = \sum_{j=1}^{M_k} \alpha_{j,k} p_j(x_k|y_{0:k}) \tag{4.1}$$

Cette approche a été introduite par Vermaak et. al dans leurs travaux sur le suivi de joueurs de football dans une séquence vidéo [118]. S'inspirant de [118], Okuma a utilisé cette modélisation dans le cadre du pistage vidéo de joueurs de hockey [95]. La particularité de [95] est de détecter les joueurs par un classifieur à partir de données d'apprentissage composées de joueurs de hockey. Après la phase de détection, une composante $p_j(x_k|y_{0:k})$ est affectée à chaque joueur détecté. Cela permet de gérer l'apparition ou la disparition de joueurs d'une image à l'autre. Dans l'approche de Vermaak[118], un mode peut correspondre à plusieurs joueurs proches. Pour regrouper les particules correspondant au même mode, Vermaak et ses co-auteurs utilisent l'algorithme de clustering k-means [80], qui nécessite une connaissance a priori du nombre de modes.
Nous reprenons dans cette thèse la formulation (4.1) afin de conserver au mieux les modes de la loi a posteriori. Les contributions du chapitre sont les suivantes :

- l'utilisation de l'algorithme mean-shift [19] pour la détermination automatique des modes et des clusters de particules,

- la gestion automatique du nuage de particules par suppression des clusters peu vraisemblable et ré-allocation des particules supprimées aux clusters restants,
- l'introduction d'une méthode de régularisation locale,
- l'échantillonnage de particules selon une série de densités d'importances centrées sur les modes de la densité a posteriori

4.1 Détermination des modes de la loi a posteriori par clustering

Afin d'utiliser la modélisation (4.1), il nous faut identifier à chaque instant l'ensemble des particules associé à la même composante p_j de manière à écrire l'approximation suivante :

$$p_{k|k} \approx \hat{p}_{k|k} = \sum_{j=1}^{M_k} \alpha_{j,k} \sum_{i \in I_{j,k}} \omega_k^i \delta_{x_k^i} \tag{4.2}$$

C'est-à-dire que $p_j \approx \sum_{i \in I_{j,k}} \omega_k^i \delta_{x_k^i}$ où, idéalement, chaque p_j est unimodale. Le regroupement des particules rattachées au même mode se fait par une méthode de clustering.

De manière générale, le clustering consiste à partitionner un ensemble de points $\mathcal{X} = \{x^i,\, i = 1, \ldots, N\}$ en M sous-ensembles $C_1, C_2, \ldots, C_M$, ou clusters, selon un critère de similarité de sorte que les éléments appartenant au même cluster soient plus semblables (en un certain sens) que les éléments d'autres clusters. Il existe une grande variété de types d'algorithmes de clustering (mesure de distance, décomposition hiérarchique, réseau de neurones,...) sur lesquels nous ne nous attarderons pas : le lecteur intéressé pourra par exemple consulter un état de l'art détaillé dans [60] et [124]. A titre d'exemple, considérons l'algorithme k-means [80]. La mesure de similarité entre deux points x^i et x^j $(i \neq j)$ de $\mathbb{R}^2$ est la distance euclidienne $\|x^i - x^j\|_2$. Les clusters sont alors déterminées en minimisant la somme des écarts quadratiques intra-classes :

$$\underset{C_1,\ldots,C_M}{\arg\min} \sum_{i=1}^{M} \sum_{x^j \in C_i} \left\| x^j - \mu(C_i) \right\|^2$$

où $\mu(C_i)$ est la moyenne empirique des points appartenant à C_i et les C_i forment une partition de $\mathcal{X}$.
L'utilisation de l'algorithme k-means n'est pas recommandée dans l'application de recalage car nous ignorons le nombre M_k de modes de $\hat{p}_{k|k}$. Par ailleurs, cet algorithme définit des clusters convexes, bien qu'a priori les clusters recherchés aient une forme quelconque. En revanche, l'algorithme mean-shift, basé sur l'estimation de densité par noyau, permet de déterminer les clusters de points correspondant au même mode, sans connaître M_k. C'est cette méthode que nous retenons afin d'identifier les ensembles $I_{j,k}$, $j = 1, \ldots, M$ ci-dessus. A chaque ensemble $I_{j,k}$ sera associé le cluster $C_{j,k}$ constitué de l'ensemble des particules d'indices appartenant à $I_{j,k}$, i.e. :

$$C_{j,k} = \left\{ x_k^i,\, i \in I_{j,k} \right\} \tag{4.3}$$

Dans toute la suite, on notera c_k la fonction qui associe l'indice i d'une particule à l'indice j du cluster auquel il appartient, k désignant la variable de temps courant. Ainsi,

$$\forall\, i \in [\![0, N]\!],\, c_k(i) = j \Leftrightarrow x_k^i \in C_{j,k} \tag{4.4}$$

Enfin on désignera par $\mathcal{X}_k$ l'ensemble des particules à l'instant k.

$$\mathcal{X}_k = \left\{x_k^i,\, i = 1, \cdots, N\right\} \tag{4.5}$$

4.1.1 L'algorithme mean-shift clustering

L'algorithme mean-shift est une méthode populaire de clustering utilisée notamment en segmentation d'image, pistage vidéo ou débruitage d'image. Les fondements de cette méthode se trouvent dans les travaux de Fukunaga et Hostetler [45] dans le cadre de l'estimation du gradient d'une fonction. Ils ont été ensuite repris dans [19] puis par Cominiciu et Meer [21] pour la segmentation d'image. La popularité du mean-shift est due au fait qu'il ne nécessite pas de connaître le nombre de clusters a priori. De plus, il permet de trouver des clusters de forme quelconque. Le mean-shift s'appuie sur l'identification des modes de la densité lissée d'un nuage de point. La localisation de ces modes est effectuée par une recherche itérative des zéros du gradient de la densité lissée.

4.1.1.1 Estimation de densité par noyau

Etant donné un ensemble de N points $x^i, i = 1, \ldots, N$, dans $\mathbb{R}^d$, on rappelle que l'estimateur de densité par noyaux est défini par :

$$\hat{f}(x) = \frac{1}{N} \sum_{i=1}^{N} K_H(x - x^i) \tag{4.6}$$

où

$$K_H(x) = \det(H)^{1/2} K(H^{-1/2} x)$$

et H est une matrice $d \times d$ symétrique définie positive, K une fonction (noyau) à support compact telle que :

$$\int K(x)\,\mathrm{d}x = 1 \qquad \int xK(x)\,\mathrm{d}x = 0 \qquad \int \|x\|^2 K(x)\,\mathrm{d}x = c_K I_d \qquad \lim_{\|x\| \to +\infty} \|x\|^d K(x) = 0 \tag{4.7}$$

Une classe importante de noyaux est celle des noyaux radialement symétriques [21] i.e. tels que :

$$K(x) = c_{k,d} k(\|x\|^2) \tag{4.8}$$

où k est une fonction définie sur $\mathbb{R}_+$ et $c_{k,d} > 0$ est telle que l'intégrale de K sur $\mathbb{R}^d$ vaut 1.
Dans toute la suite, on considère que K est radialement symétrique.
En pratique, la matrice H est choisie sous la forme $H = h^2 I_d$, où $h > 0$ est le *facteur de lissage*

global, paramètre que nous avons introduit dans la section 2.3 dans le cadre du filtre particulaire régularisé. L'estimateur de densité à noyaux devient donc :

$$\hat{f}(x) = \frac{1}{Nh^d} \sum_{i=1}^{N} K\left(\frac{x - x^i}{h}\right) \tag{4.9}$$

Notons qu'il est également possible d'avoir un estimateur à noyau où le facteur de lissage dépend des données x^i [9] : on parle alors de *facteur de lissage local*. C'est une stratégie préférable lorsque le nuage de points comporte plusieurs clusters de tailles différentes. Dans ce cas, l'expression de l'estimateur de densité est :

$$\hat{f}(x) = \sum_{i=1}^{N} \frac{1}{Nh_i^d} K\left(\frac{x - x^i}{h_i}\right) \tag{4.10}$$

avec $h_i = h(x^i)$. L'un des intérêts d'une fenêtre dépendant des données est que le biais de l'estimateur est inférieur à celui de l'estimateur à fenêtre constante [22] pour un certain choix des $h_i = h_i^*$. Cependant, la détermination des h_i^* dans le cas multi-dimensionnel est un problème non encore résolu [12].

Rappelons que la qualité de l'estimateur $\hat{f}$ est classiquement mesurée par l'erreur quadratique intégrée moyenne, ou MISE *(Mean Integrated Squared Error)* définie par :

$$\mathrm{MISE} = \mathbb{E}\left(\int \left(f(x) - \hat{f}(x)\right)^2 \mathrm{d}x\right) \tag{4.11}$$

Souvent, seule une approximation asymptotique, dénomée AMISE *(Asymptotic Mean Integrated Squared Error)*, est disponible. Lorsque N tend vers $+\infty$ et h tend vers 0 moins vite que $\frac{1}{N}$, l'AMISE est minimisée par le noyau d'Epanechnikov (cf. eq. (2.27)) [120, p.104].

La première étape de l'algorithme mean-shift consiste à déterminer les modes de la densité sous-jacente f. Pour cela, on recherche les zéros du gradient ∇f à l'aide de l'estimateur $\hat{f}(x)$ défini par (4.9).

4.1.1.2 Estimation du gradient de la densité

L'estimateur donnée par l'equation (4.9) peut être réécrit comme suit :

$$\hat{f} = \frac{c_{k,d}}{Nh^d} \sum_{i=1}^{N} k\left(\left\|\frac{x - x^i}{h}\right\|^2\right) \tag{4.12}$$

Un estimateur du gradient de la densité f est obtenu en calculant le gradient de $\hat{f}$:

$$\hat{\nabla} f(x) = \nabla \hat{f}(x) = \frac{2c_{k,d}}{Nh^{d+2}} \sum_{i=1}^{N} \left(x - x^i\right) k'\left(\left\|\frac{x - x^i}{h}\right\|^2\right) \tag{4.13}$$

En posant $g(x) = -k'(x)$, on peut réécrire (4.13) sous la forme :

$$\hat{\nabla} f(x) = \nabla \hat{f}(x) = \frac{2c_{k,d}}{Nh^{d+2}} \sum_{i=1}^{N} \left(x^i - x\right) g\left(\left\|\frac{x - x^i}{h}\right\|^2\right) \tag{4.14}$$

$$= \frac{2c_{k,d}}{Nh^{d+2}} \left\{\sum_{i=1}^{N} g\left(\left\|\frac{x - x^i}{h}\right\|^2\right)\right\} \left\{\frac{\sum_{i=1}^{N} x^i g\left(\left\|\frac{x-x^i}{h}\right\|^2\right)}{\sum_{i=1}^{N} g\left(\left\|\frac{x-x^i}{h}\right\|^2\right)} - x\right\} \tag{4.15}$$

En pratique pour les noyaux usuels (Epanechnikov, gaussien), $\sum_{i=1}^{N} g\left(\left\|\frac{x-x^i}{h}\right\|^2\right) > 0$. Le terme $\frac{2c_{k,d}}{Nh^{d+2}} \sum_{i=1}^{N} g\left(\left\|\frac{x-x^i}{h}\right\|^2\right)$ est proportionnel à l'estimateur de densité en x calculée avec un noyau G - appelé *kernel shadow* - tel que $G(x) = c_{g,d} g(\|x\|^2)$ où $c_{g,d}$ est une constante de normalisation :

$$\hat{f}_g = \frac{c_{g,d}}{Nh^d} \sum_{i=1}^{N} g\left(\left\|\frac{x - x^i}{h}\right\|^2\right) \tag{4.16}$$

Le deuxième terme est appelé vecteur *mean-shift* :

$$m_{h,G}(x) = \frac{\sum_{i=1}^{N} x^i g\left(\left\|\frac{x-x^i}{h}\right\|^2\right)}{\sum_{i=1}^{N} g\left(\left\|\frac{x-x^i}{h}\right\|^2\right)} - x \tag{4.17}$$

En combinant les expressions (4.15) et (4.16), le vecteur mean-shift s'exprime comme

$$m_{h,G}(x) = \frac{1}{2} h^2 \frac{c_{g,d}}{c_{k,d}} \frac{\nabla \hat{f}_{h,K}(x)}{\hat{f}_{h,G}(x)} \tag{4.18}$$

Lorsque l'estimateur à noyau est défini grâce à une matrice de lissage H et que le noyau K est gaussien, le vecteur mean-shift s'écrit sous la forme

$$m_{H,G}(x) = H \frac{\nabla \hat{f}_{h,K}(x)}{\hat{f}_{h,G}(x)} \tag{4.19}$$

L'équation (4.19) suggère que le vecteur mean-shift pointe vers la direction de plus fort gradient de densité et peut donc être utilisé dans un schéma de recherche des modes de la densité f.
En particulier, le gradient $\hat{\nabla} f(x)$ va s'annuler lorsque ce vecteur sera nul. Sous certaines conditions sur le noyau K, la suite $\{y_j^i, j \geq 1\}$ définie par :

$$\begin{cases} y_1^i & = x^i \\ y_{j+1}^i & = y_j^i + m_{h,G}(y_j^i) \quad \text{si } j \geq 1 \end{cases} \tag{4.20}$$

converge vers une limite y_∞^i correspondant à un maximum local de la densité $\hat{f}$ [21].
Cette suite est définie pour chaque donnée x^i, $i = 1, \ldots, N$ et initialisée par x^i. Dans la suite on

appellera *procédure mean-shift* le calcul de cette suite pour $1 \leq i \leq N$. En pratique, lorsque le facteur de lissage h est adéquatement choisi, la procédure mean-shift peut converger en quelques itérations (typiquement entre 5 et 20) [43]. La recherche de la limite peut alors se faire en calculant les termes successifs de la suite $\{y_j^i, j \geq 1\}$ jusqu'à ce que $\|y_{j+1}^i - y_j^i\| < \varepsilon^{MS}$ où ε^{MS} est un seuil de convergence défini par l'utilisateur. Lorsque K est le noyau d'Epanechnikov, G correspond au noyau uniforme et le nombre d'itérations nécessaires pour la convergence de la suite est fini [21].

On remarque que le coeur d'une procédure mean-shift est le calcul du vecteur $m_{h,G}(x)$ pour un x donné. Ceci nécessite d'évaluer les valeurs $g(\left\|\frac{x-x^i}{h}\right\|^2)$ pour tout $i = 1, \ldots, N$. Or, on peut considérer que pour les noyaux K vérifiant les hypothèses de l'équation (4.7), $g(x) \approx 0$ lorsque $\|x\| \geq h$ (pour les noyaux uniformes ou triangulaires, l'approximation devient une vraie égalité). Ceci entraîne que $g(\left\|\frac{x-x^i}{h}\right\|^2) \approx 0$ lorsque $\|x - x^i\| \geq h$. L'expression du vecteur mean-shift (4.17) peut donc être approchée par :

$$m_{h,G}(x) \approx \frac{\sum\limits_{x^i \in \mathcal{B}_h(x)} x^i g\left(\left\|\frac{x-x^i}{h}\right\|^2\right)}{\sum\limits_{x^i \in \mathcal{B}_h(x)} g\left(\left\|\frac{x-x^i}{h}\right\|^2\right)} - x \tag{4.21}$$

où $\mathcal{B}_h(x)$ désigne la boule de rayon h centrée en x. En particulier, si K est le noyau d'Epanechnikov, $g(x) = \mathbb{1}_{\|x\| \leq h}$ et

$$m_{h,G}(x) \approx \frac{1}{n_x} \sum_{x^i \in \mathcal{B}_h(x)} x^i - x \tag{4.22}$$

où n_x désigne le nombre d'éléments x^i de $\mathcal{X}$ qui sont à une distance inférieure ou égale à h.

La figure 4.1 illustre le calcul du vecteur mean-shift dans le plan, avec un noyau gaussien pour une valeur de $h = 1.8$. Les données en rouge représentent les points qui sont dans le cercle de rayon h centré en y_j, le point où s'effectue le calcul de vecteur mean-shift.

Le vecteur mean-shift peut être également généralisé à un échantillon $\{x^i, \omega^i\}_{i=1}^N$ pondéré avec $\sum_{i=1}^N \omega^i = 1$. Dans ce cas,

$$\hat{f}(x) = \frac{1}{h^d} \sum_{i=1}^N \omega^i K\left(\frac{x - x^i}{h}\right) \tag{4.23}$$

et

$$m_{h,G}(x) = \frac{\sum_{i=1}^N \omega^i x^i g\left(\left\|\frac{x-x^i}{h}\right\|^2\right)}{\sum_{i=1}^N \omega^i g\left(\left\|\frac{x-x^i}{h}\right\|^2\right)} - x \tag{4.24}$$

C'est cette forme qui sera utilisée pour l'identification des clusters.

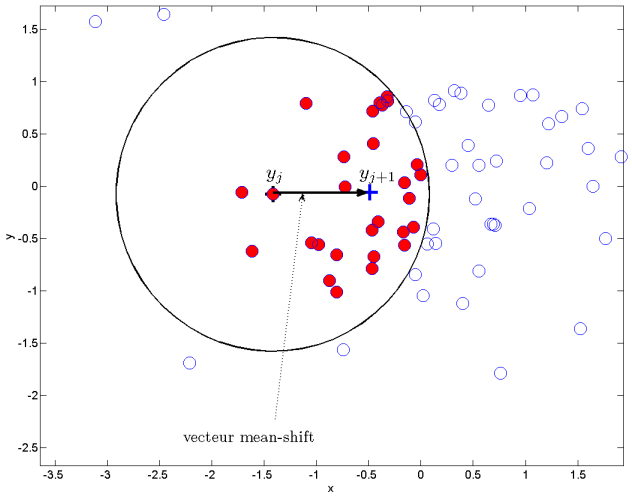


FIGURE 4.1 – Calcul du vecteur mean shift

4.1.1.3 Détermination de clusters par procédures mean-shift

La section précédente a montré que l'on pouvait déterminer les modes de la densité lissée d'un nuage de points en identifiant les points où le vecteur mean-shift s'annule. Trouver les clusters correspondants consiste essentiellement à regrouper les points dont les procédures mean-shift ont convergé vers le même mode ou vers deux modes très proches. En principe, cela suppose de calculer les limites y_∞^i de (4.20) pour chaque x^i, $i = 1, \ldots, N$. Cependant, il est possible de réduire le nombre de procédures mean-shift en définissant un pavage de l'ensemble des données $\mathcal{X} = \{x^i,\, i = 1, \ldots, N\}$. Un pavage des données est un sous-ensemble $\tilde{\mathcal{X}} = \{x^{i_j},\, j = 1, \ldots, \tilde{N}\}$ de $\mathcal{X}$ tel que :

1. $\tilde{N} < N$
2. Deux points distincts x^{i_j} et x^{i_k} $(j \neq k)$ sont à une distance supérieure à la taille du facteur de lissage h.
3. La réunion des boules de rayon h centrées en les points de $\tilde{\mathcal{X}}$ contient $\mathcal{X}$.

En utilisant ce schéma, la détermination des modes de la densité lissée du nuage de points se fait en effectuant $\tilde{N}$ procédures mean-shift au lieu de N ce qui réduit la complexité algorithmique. Une fois ces modes trouvés, il s'agit d'associer chaque échantillon x^i à un cluster correspondant à l'un des modes. Pour cela, on introduit un paramètre R_m, appelé rayon de fusion, qui a une influence sur le degré de séparation des clusters et le nombre de clusters identifiés. Si deux procédures mean-shift initialisées à partir des points x^{i_j} et x^{i_k} de $\tilde{\mathcal{X}}$ convergent vers deux modes $y_\infty^{i_j}$ et $y_\infty^{i_k}$ tels que $\|y_\infty^{i_j} - y_\infty^{i_k}\| \leq R_m$, les deux modes sont fusionnées en un seul (par exemple en prenant leur moyenne) et les deux points x^{i_j} et x^{i_k} sont associés au même cluster. Dans le cas contraire, ils définissent deux clusters distincts. Dans la figure 4.2, les procédures mean-shift initialisées à partir de deux points distincts notés P_1 convergent vers le même mode P_n ; ils sont donc associés au même cluster.

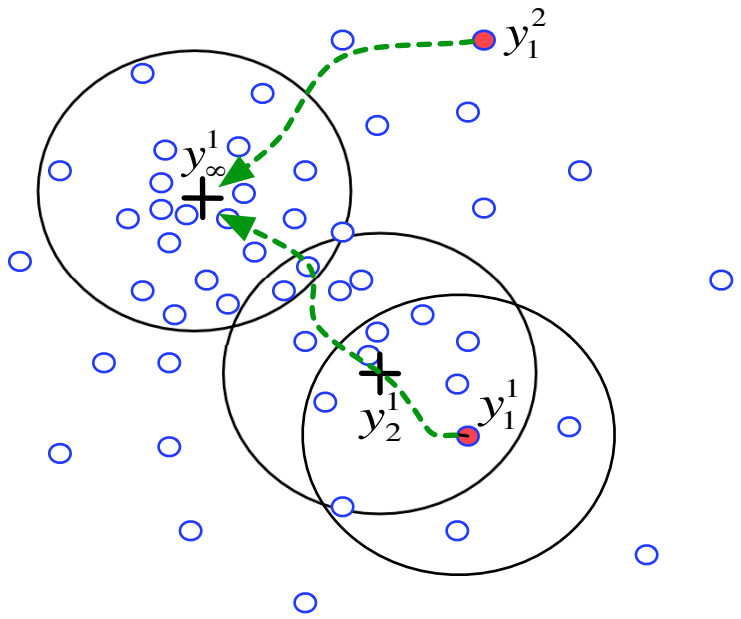


Figure 4.2 – Procédures mean-shift : les cercles matérialisent les positions successives du noyau. y_1^1 et y_1^2 désignent les points initiaux de chaque procédure mean-shift. y_∞^1 est le mode vers lequel converge les 2 procédures.

Enfin, si un échantillon x^i n'appartient pas à l'ensemble $\tilde{\mathcal{X}}$, on lui associe le cluster majoritaire de ses k-plus proches voisins issus de $\tilde{\mathcal{X}}$.

4.1.2 Sélection automatique du facteur de lissage du mean-shift clustering

Contrairement à la méthode du k-means [80], l'algorithme mean-shift ne requiert pas la connaissance a priori du nombre de clusters recherchés. Cependant, il nécessite de spécifier h la taille du facteur de lissage, ce qui a une influence directe sur le nombres de clusters trouvés. Par ailleurs, il faut également spécifier le paramètre de rayon de fusion R_m, qui permet de fusionner deux clusters dont les modes correspondants serait à une distance inférieure à R_m.
Il existent plusieurs règles, heuristiques ou non, qui permettent d'obtenir une valeur de h appropriée. Nous rappelons ci-après quelques techniques génériques.

4.1.2.1 Sélection d'un facteur de lissage global fondée sur un critère statistique

Les critères de sélection fondés sur l'analyse statistique des propriétés de l'estimateur à noyau (4.6) visent à minimiser une mesure de l'erreur entre $f(x)$ et $\hat{f}(x)$ telle que l'erreur quadratique moyenne (MSE), l'erreur quadratique moyenne intégrée (MISE) ou l'erreur quadratique moyenne intégrée asymptotique (AMISE), pour un $x \in \mathbb{R}^d$ donné. Par exemple, l'erreur quadratique

moyenne de l'estimateur $\hat{f}(x)$ s'exprime comme la somme des biais de la variance de $\hat{f}(x)$.

$$\mathrm{MSE}(x) = \mathbb{E}(\hat{f}(x) - f(x))^2 = (b(x))^2 + v(x) \tag{4.25}$$

où $b(x) = \mathbb{E}(\hat{f}(x)) - f(x)$ et $v(x)$ sont respectivement le biais et la variance de l'estimateur $\hat{f}(x)$. Les expressions approchées de $b(x)$ et $v(x)$ sont [120]

$$b(x) \approx \frac{1}{2} h^2 \mu_2(K) \Delta f(x) \tag{4.26}$$

$$v(x) \approx N^{-1} h^{-d} R(K) f(x) \tag{4.27}$$

avec $\mu_2(K) = \int z_1^2 K(z) dz$ et $R(K) = \int K(z) dz$, z_1 étant la première composante du vecteur z. Au vu de (4.26) et (4.27), le biais augmente lorsque h augmente tandis que la variance diminue avec l'accroissement de h. Dans le cadre d'un estimateur à noyau à fenêtre constante, le choix d'une valeur de h optimale implique un compromis entre le biais $b(x)$ et la variance $v(x)$ pour tout x de $\mathbb{R}^d$. Cela revient à la minimisation de l'erreur quadratique moyenne intégrée $\mathrm{MISE} = \mathbb{E} \int (f(x) - \hat{f}(x))^2 dx$. Malheureusement, la valeur optimale h_{MISE} est peu exploitable car elle dépend du Laplacien de la densité inconnue f [112]. Néanmoins, il existe des techniques qui permettent d'approcher h_{MISE} décrites dans la section 2.3.

4.1.2.2 Sélection d'un facteur de lissage global basée sur un indice de validité des clusters

Cette approche consiste à trouver le paramètre h qui maximise une fonction objectif représentant la qualité du clustering. De manière générale, cette fonction évalue le caractère compact des clusters délimités et le degré de séparation de chaque cluster. Une possibilité est d'évaluer l'isolement et la connectivité des clusters pour plusieurs valeurs de h et de choisir la valeur de h qui optimise les deux critères [99]. L'isolement se réfère au fait que pour un point x^i donné appartenant au cluster d'indice $c(i)$, ses plus proches voisins appartiennent au même cluster. Cela correspond au fait qu'idéalement, un algorithme de clustering satisfaisant isole chaque cluster du reste des données. Une mesure d'isolement d'un cluster est la moyenne de la norme des k plus proches voisins des points du cluster. La norme $\nu_k(x)$ des k plus proches voisins d'un point x est définie comme la fraction des k plus proches voisins appartenant au même cluster que x. Autrement dit, N étant le nombre total de points, l'isolement Iso_k se mesure comme :

$$\mathrm{Iso}_k = \frac{1}{N} \sum_{i=1}^{N} \nu_k(x^i)$$

Cette mesure seule ne suffit cependant pas à caractériser la qualité de clustering car elle ne peut détecter des clusters fusionnés à tort. En effet, dans le cas extrême où l'on associerait indûment tous les points au même cluster, la mesure d'isolement est optimale. L'utilisation d'une mesure de connectivité permet d'éviter ceci en pénalisant les partitions qui regroupent des points appartenant en réalité à des clusters distincts. La connectivité d'un cluster signifie que pour n'importe quelle paire de point (x^i, x^j), $i \neq j$, il existe un chemin entre x^i et x^j sur lequel la densité de points

reste élevée. Pour une estimation d'un indice de connectivité Con_k du cluster, on pourra se référer à l'article de Pauwels et al. [99]. A partir des indices Iso_k et Con_k il est possible de déduire, pour une partition donnée, un indice composite Z prenant en compte les deux indicateurs. La meilleure partition est alors celle qui maximise ce critère. Afin de déterminer une valeur optimale de h, on peut donc procéder de la manière suivante :

1. Définir b facteurs de lissage $h^{(1)}$, $h^{(2)}$,..., $h^{(b)}$
2. Pour $j = 1, \ldots, b$, déterminer la partition $\mathcal{C}^{(j)}$ correspondant au facteur $h^{(j)}$
3. Pour chaque partition $\mathcal{C}^{(j)}$, évaluer l'indice composite d'isolement et de connectivité $Z^{(j)}$ (cf. [99])
4. La valeur optimale $h^{(j)}$ est celle qui maximise $Z^{(j)}$, $j = 1, \ldots, b$

4.1.2.3 Sélection d'un facteur de lissage local basée sur la stabilité de la partition

Une méthode de recherche systématique d'une fenêtre de lissage consiste à démarrer d'une petite valeur $h = h_{min}$ et à effectuer plusieurs instances de clustering en augmentant peu à peu la valeur de h [44]. Typiquement, pour $h = h_{min}$ le nombre de clusters est élevé puis baisse progressivement pour ensuite se stabiliser sur un intervalle $[h_a, h_b]$, avant de baisser à nouveau jusqu'à atteindre la valeur 1. On choisit alors la valeur moyenne $h = \frac{h_a + h_b}{2}$. Cette méthode est malheureusement gourmande car elle nécessite d'appliquer l'algorithme de clustering plusieurs fois.
Intuitivement, la valeur de h est reliée à l'échelle des clusters du nuage de points. Comaniciu propose d'utiliser une approche semi-paramétrique pour sélectionner un facteur de lissage local h_i. Il se base sur le résultat suivant, plus général puisqu'il suppose l'utilisation d'une matrice de lissage H. Ce résultat est valide lorsque N est grand.

Théorème 3 (Comaniciu [20])**.** *Supposons que la vraie densité f soit une gaussienne de moyenne m et de matrice de covariance $\Sigma = \sigma^2 I$. Soit $m_{h,G}(x)$ le vecteur mean-shift calculé selon (4.19) où G est un noyau gaussien. Alors le vecteur $\|h^{-1/2} \operatorname{plim} m_{h,G}(x)\|$ atteint son maximum lorsque $h = \sigma$, la notation plim signifiant la limite en probabilité.*

Lorsque la distribution f comporte M modes, Comaniciu émet l'hypothèse qu'au voisinage de chaque mode, la distribution est approximativement gaussienne. Cela permet de déduire une série de matrices de covariance locales $\Sigma^{(1)}, \Sigma^{(2)}, \ldots, \Sigma^{(M)}$.
Afin d'en déduire M facteurs de lissage optimaux h_i, Comaniciu observe que si les M clusters peuvent effectivement être paramétrés par un ensemble de M paires moyenne-covariance $(\mu^{(i)}, \Sigma^{(i)})$, $i = 1, \cdots, M$, alors la meilleure partition est celle pour laquelle l'estimation des $(\mu^{(i)}, \Sigma^{(i)})$ est la plus stable, c'est-à-dire celle pour laquelle une petite variation des facteurs de lissage a le moins d'effet sur l'estimation des paramètres des lois normales [20].

L'algorithme de sélection d'un facteur de lissage locale de Comaniciu se déroule en deux étapes. On suppose que l'on dispose d'une série de b facteurs de lissage $h^{(1)}$, $h^{(2)}$,..., $h^{(b)}$ définis à l'avance, sur une échelle logarithmique par exemple.

1. Pour chaque facteur $h^{(j)}$, $j = 1, \ldots, b$

(a) Déterminer une partition de taille M_j représentée par un ensemble de M_j clusters. Noter $c_j(i) = u$ si x^i appartenant au cluster d'indice u, obtenu avec le facteur de lissage $h^{(j)}$.

(b) Pour le u-ème mode de la partition ($1 \leq u \leq M_j$), estimer la moyenne $\mu^{(u,j)}$ et la covariance $\Sigma^{(u,j)}$ de l'approximation gaussienne correspondant à l'ensemble des particules appartenant au cluster d'indice u.

(c) Associer à chaque donnée x^i appartenant au cluster d'indice u, la moyenne $\mu^{(u,j)}$ et la covariance $\Sigma^{(u,j)}$.

2. Pour chaque donnée x^i, identifier l'indice u du cluster qui la contient, ceci pour chaque facteur $h^{(j)}$, i.e trouver u tel que $c_j(i) = u$, $j = 1, \ldots, b$

(a) Soit G_1, G_2,...,G_b les lois des gaussiennes de paramètres $(\mu^{(u,j)}, \Sigma^{(u,j)})$, $j = 1, \ldots, b$.

(b) Déterminer la valeur $h^{(j_0)}, j_0 = 1, \ldots, b$ pour laquelle l'estimation de $(\mu^{(u,j)}, \Sigma^{(u,j)})$ est la plus stable. Pour cela, minimiser la divergence de Jensen-Shannon entre G_{j-1}, G_j et G_{j+1} pour $r = 2, \cdots, b-1$ (voir ci-après).

(c) Associer à x^i le facteur local $h_i = h^{(j_0)}$

La divergence de Jensen-Shannon $JS(\varphi_1, \cdots, \varphi_r)$ entre r distributions de densités $\varphi_1, \cdots, \varphi_r$ est une mesure de dissimilarité entre ces lois. Elle permet notamment de traiter des problèmes de classification multi-classes.
Etant donné r coefficients positifs $\pi_1, \cdots, \pi_r$, de somme 1, $JS(\varphi_1, \cdots, \varphi_r)$ est définie par [75] :

$$JS(\varphi_1, \cdots, \varphi_r) = \mathcal{H}(\sum_{i=1}^{r} \pi_i \varphi_i) - \sum_{i=1}^{r} \pi_i \mathcal{H}(\varphi_i)$$

où $\mathcal{H}$ désigne la fonction d'entropie de Shannon définie pour toute densité de probabilité q par

$$\mathcal{H}(q) = \int q(x) \log q(x) dx \tag{4.28}$$

Lorsque les densités $\varphi_1, \varphi_2, \cdots, \varphi_r$ sont gaussiennes, de moyennes et covariances respectives μ_j et Σ_j, $j = 1, \cdots, r$, la divergence de Jensen-Shannon se calcule selon :

$$JS(\varphi_1, \cdots, \varphi_r) = \frac{1}{2} \log \frac{|\sum_{j=1}^{r} \Sigma_j|}{(\prod_{j=1}^{r} |\Sigma_j|)^{1/r}} + \frac{1}{2} \sum_{j=1}^{r} (\mu_j - \mu)^T \left(\sum_{j=1}^{r} \Sigma_j \right)^{-1} (\mu_j - \mu)$$

où $\mu = \frac{1}{r} \sum_{j=1}^{r} \mu_j$.

La méthode décrite de sélection ci-dessus permet d'obtenir une série de facteurs de lissage locaux $h_i = h(x^i)$ permettant d'effectuer un clustering adapté à l'échelle locale du nuage de particule. Cependant, par souci de simplicité, nous avons cherché une méthode qui effectue le clustering des particules grâce à un unique facteur d'échelle global. Une autre limitation des techniques de sélection de facteur de lissage h basées sur la stabilité de la partition est qu'elles requièrent d'appliquer plusieurs fois l'algorithme de clustering sur le même nuage de points afin d'identifier la valeur optimale de h. Dans le cadre de notre thèse, nous souhaitons une méthode de clustering dont le coût de calcul reste raisonnable et donc nous laisserons de côté les approches fondées sur la stabilité de la partition.

4.1.2.4 Sélection d'un facteur de lissage global adapté à l'échelle du nuage de particules

Le problème posé est le suivant : étant donné l'algorithme de mélange de filtre particulaire exposé dans la sous-section 2.6.2, il s'agit à chaque instant k de déterminer un ensemble de M_k clusters $C_{j,k}, j = 1, \ldots, M_k$ de particules grâce à l'algorithme mean-shift. Pour cela, il faut régler un facteur de lissage h qui dépend essentiellement de l'échelle locale du nuage de particules à l'instant considéré. Comme le nuage de particule change de taille au cours du temps, h doit être modifié en conséquence. Nous proposons un algorithme qui met à jour, à chaque instant, la valeur du facteur de lissage global. On notera $h = h^{(k)}$ pour rappeler la dépendance en temps.

Pour cela, on suppose que l'on dispose d'une partition à l'instant k de l'ensemble des particules $\mathcal{X}_k$ sous forme de clusters $C_{j,k}$ $(j = 1, \ldots, M_k)$. En s'inspirant de l'approche de Comaniciu, on effectue une approximation gaussienne de chaque cluster afin de déterminer son échelle locale. On note $\hat{\Sigma}_k^{(j)}$, $j = 1, \ldots, M_k$ les matrices de covariance empiriques correspondantes et $m_k^{(j)}$, $j = 1, \ldots, M_k$ les modes associés aux clusters. L'échelle locale d'un cluster est alors donnée par le demi-grand axe et le demi-petit axe de l'ellipsoïde de confiance de la covariance du cluster pour un niveau de confiance donnée (par exemple 95%). Cela permet d'avoir une matrice de lissage bi-dimensionnelle ; cependant, comme nous souhaitons un unique paramètre, nous résumons l'échelle du cluster au demi petit-axe. Afin d'obtenir un facteur de lissage *global*, nous proposons de prendre la moyenne $\bar{b}$ des échelles locales (i.e. les demis-petit axes). Le facteur de lissage global obtenu à l'instant k sert d'entrée à l'algorithme de clustering à l'itération suivante, ce qui est motivé par le fait que le nuage de particules évolue lentement d'un instant à l'autre.

1. Pour $j = 1, \ldots, M_k$, déterminer l'ellipsoïde de confiance à 95% associé $\hat{\Sigma}_k^{(j)}$. Calculer son demi-petit axe b_j.
2. Poser $\bar{b} = \frac{1}{M_k} \sum_{i=1}^{M_k} b_j$
3. Poser $h^{(k+1)} = \bar{b}$

4.1.3 Stratégies de réduction du temps de calcul de l'algorithme de mean-shift clustering

La complexité algorithmique de l'algorithme mean-shift standard est en $\mathcal{O}(JdN^2)$[46] où J désigne le nombre d'itérations de la procédure mean-shift, d la dimension et N le nombre de particules à classifier. En adoptant la stratégie consistant à trouver un pavage de l'ensemble des particules $\mathcal{X}_k$ (cf. section 4.1.1.3), c'est-à-dire un sous-ensemble de taille $\tilde{N} < N$, la complexité est de l'ordre de $\mathcal{O}\left(Jd\tilde{N}^2\right)$. Les algorithmes particulaires standards ont une complexité de l'ordre de $\mathcal{O}(N)$ ce qui indique qu'il faut envisager des stratégies permettant de réduire la complexité de l'algorithme de clustering.
L'étape la plus couteuse est le calcul de la suite $\{y_j, j \geq 1\}$ qui converge vers les modes du nuage

	mean-shift avec arbres k-d	mean-shift standard
N = 5000	0.1516	0.2279
N = 2000	0.1050	0.1404
N = 1000	0.0509	0.0513

TABLE 4.1 – Temps de calcul (en secondes) de l'algorithme mean-shift sur un N-échantillon issu du mélange gaussien F ; moyenne sur 50 réalisations

de points, dont on rappelle l'expression ci-dessous :

$$y_{j+1} = \frac{\sum_{x^i \in \mathcal{B}_h(y_j)} x^i g\left(\left\|\frac{y_j - x^i}{h}\right\|^2\right)}{\sum_{x^i \in \mathcal{B}_h(y_j)} g\left(\left\|\frac{y_j - x^i}{h}\right\|^2\right)} \tag{4.29}$$

$$\text{où } \mathcal{B}_h(y_j) = \{x^i, i = 1, \ldots, N : \|y_j - x^i\| \leq h\} \tag{4.30}$$

Plus précisément, les goulots d'étranglement sont :

(i) La détermination de la boule $\mathcal{B}_h(y_j)$: c'est un problème de recherche des plus proches voisins dont la distance à y_j est inférieure à h. Des structures telles que les arbres k-dimensionnels (*k-d trees*) [123] permettent de déterminer le voisinage $\mathcal{B}_h(y_j)$ efficacement.

(ii) L'évaluation du quotient (4.30) qui consiste en $2Nd$ multiplications (Nd au numérateur et au dénominateur), sans compter le coût de l'évaluation de $g(x)$ pour x qui dépend du noyau utilisé.

Dans le cadre de la thèse, nous avons retenu l'utilisation d'un noyau G uniforme pour le calcul de la suite $\{y_j\}$ car le nombre d'itérations nécessaire pour la convergence de la suite est fini [21]. De plus, l'évaluation de y_j pour un noyau uniforme est moins coûteuse que celle utilisant un noyau gaussien.
En ce qui concerne la recherche des points qui tombent dans le support du noyau [point (ii)], nous avons constaté que l'utilisation d'une structure d'arbre k-d permettait de réduire le temps de calcul de manière appréciable lorsque $N \geq 2000$ (voir 4.1). Nous avons toutefois noté que pour des valeurs moindres de N, l'utilisation de telles structures de données n'était pas justifié. La comparaison a été menée sur des données simulées selon le mélange gaussien F défini par :

$$F = \frac{1}{5}\sum_{i=1}^{5} \mathcal{N}(\mu_i, \Sigma_i) \tag{4.31}$$

avec $\mu_1 = [3,\, 3]^T$, $\mu_2 = [-6,\, -6]^T$, $\mu_3 = [-5,\, 2]^T$, $\mu_4 = [5,\, -3]^T$, $\mu_5 = [0,\, 0]^T$ et

$$\Sigma_1 = \begin{pmatrix} 1/2 & 1/4 \\ 1/2 & 1/4 \end{pmatrix}, \quad \Sigma_2 = \tfrac{1}{2} I_2, \quad \Sigma_3 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad \Sigma_4 = \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}, \quad \Sigma_5 = \begin{pmatrix} 0.6 & 0.15 \\ 0.15 & 0.375 \end{pmatrix}$$

Nous exposons ci-après quelques méthodes permettant de réduire le coût algorithmique du clustering.

Représentation compacte de densité (RCD) Cette méthode introduite par Freedman [43] vise à accélérer l'étape de clustering en réduisant le nombre N de points $x_1, \dots, x_N$ sur lequel la procédure mean-shift est appliquée. Pour cela, on s'appuie sur une représentation compacte de la densité $\hat{f}$.

Soit $\hat{\hat{f}}(x) = \frac{1}{L}\sum_{j=1}^{L} K_H(x - \hat{x}_j)$, avec $L << N$. Pour un choix approprié des $\hat{x}_j$ et de H, l'erreur $\mathbb{E}\left[\int \left(\hat{f}(x) - \hat{\hat{f}}(x)\right)^2 dx\right]$ tend vers 0 lorsque $L \to \infty$.

L'algorithme de clustering se résume alors à la succession des étapes suivantes :

1. Tirer L échantillons $\hat{x}_1, \dots, \hat{x}_L \sim \hat{f}$ afin de construire $\hat{\hat{f}}$.
2. Effectuer une procédure mean-shift sur chacun des échantillons $\hat{x}_j$, $j = 1, \cdots, L$, i.e. évaluer le mode limite y^j_∞.
3. Pour chaque x^i, trouver l'échantillon $\hat{x}_{j*}$ le plus proche et associer le mode $\hat{\hat{m}}_{j*}$ à x_i.

La seconde étape permet un gain considérable puisque on utilise L points pour le calcul de la suite des vecteurs mean shift au lieu de N. Selon [43], une implémentation de cet algorithme permet de diviser le temps de calcul par un facteur $\frac{N}{L}$.

Nous proposons une simplification qui permet de se passer de la génération des L échantillons selon $\hat{f}$. Pour cela, les échantillons $\{\hat{x}_j\}$ peuvent être tirés au hasard et sans remise parmi les échantillons de départ $x^1, \dots, x^N$. Cette méthode ne garantit plus que l'approximation $\hat{\hat{f}}$ soit proche de $\hat{f}$ mais elle s'est avérée efficace expérimentalement.

Clustering périodique Nous faisons également l'observation que la forme du nuage de particules n'évolue que légèrement d'un instant à l'autre lorsqu'il n'y a pas d'étape de ré-échantillonnage. On peut raisonnablement s'attendre à ce que deux particules appartenant au même cluster le restent à l'itération suivante. Ainsi, on peut attendre quelques cycles de prédiction/correction (typiquement 3 à 10) avant d'appliquer l'algorithme de clustering.

4.2 Ré-échantillonnage et régularisation locale

Le filtre particulaire régularisé (RPF) [37, p. 247-271], introduit au chapitre 2, permet d'augmenter la robustesse du filtre dans le cas où le bruit de dynamique et/ou d'observation est faible. Il s'appuie sur une régularisation de la loi discrète obtenue après la phase de correction (cf. 2.3) : concrètement, la régularisation consiste à ajouter un bruit contrôlé aux particules ré-échantillonnées de façon à améliorer l'exploration de l'espace d'état. Soit $\hat{p}_{k|k} = \sum_{i=1}^{N} \omega_k^i \delta_{x_k^i}$, l'approximation de la loi conditionnelle à l'instant k.

Dans sa version originale, la phase de ré-échantillonnage du RPF se fait en échantillonnant un estimateur à noyau de $\hat{p}_{k|k}$. Cet estimateur est noté $K_h * \hat{p}_{k|k}$ et est défini par

$$K_h * \hat{p}_{k|k} = \frac{1}{h^d}\sum_{i=1}^{N} \omega_k^i K\left(\frac{x - x_k^i}{h}\right)$$

où K est le noyau de régularisation considéré. Une implémentation du RPF nécessite de préciser le coefficient de dilatation h, à ne pas confondre avec le facteur de lissage utilisé dans l'algorithme de clustering mean-shift. Le choix optimal de h se fait en minimisant l'erreur quadratique moyenne intégrée entre $K_h * \hat{p}_{k|k}$ et la loi a posteriori inconnue $p_{k|k} = p(x_k|y_{0:k})$ (cf. 2.3). Dans sa formulation générique, le RPF approxime la loi de filtrage $p_{k|k}$ par une gaussienne dont les moments sont la moyenne empirique m_k et la covariance empirique S_k du système de particules pondérés $\{\omega_k^i, x_k^i\}_{i=1}^N$.

$$m_k = \sum_{i=1}^{N} \omega_k^i x_k^i$$

et

$$S_k = \sum_{i=1}^{N} \omega_k^i (x_k^i - m_k)(x_k^i - m_k)^T$$

Afin de sa ramener au cas connu d'une gaussienne standard, les particules sont blanchies, c'est-à-dire que l'on considère les particules obtenue par la transformation $S_k^{-\frac{1}{2}}(x_k^i - m_k)$. En notant $\{\tilde{x}_k^i\}_{i=1}^N$ les particules obtenues après ré-échantillonnage multinomial, le système de particules régularisé est obtenu en posant :

$$x_k^i = \tilde{x}_k^i + h_{opt} S_k^{\frac{1}{2}} \varepsilon^i$$

où les $\varepsilon^i, i = 1, \ldots, N$ sont des tirages i.i.d. du noyau K.
Lorsque $p_{k|k}$ est multimodale, Silverman [112] conseille d'utiliser le facteur $h = ch_{opt}$ où $0 < c \leq 0.5$ est un paramètre de réglage, h_{opt} étant calculé d'après 2.28 ou 2.29 selon le noyau K utilisé. Nous avons néanmoins constaté que cette méthode n'est pas satisfaisante, notamment lorsque les modes de la loi conditionnelle sont éloignés. A titre d'exemple, considérons un échantillon $x_1, x_2, \ldots, x_N$ de taille $N = 100$ issu de la distribution bi-modale

$$\nu = \frac{1}{2}\mathcal{N}(m_1, \Sigma_1) + \frac{1}{2}\mathcal{N}(m_2, \Sigma_2) \tag{4.32}$$

avec

$$m_1 = \begin{pmatrix} 5 & 5 \end{pmatrix}^T \quad m_2 = \begin{pmatrix} -50 & -60 \end{pmatrix}^T$$
$$\Sigma_1 = \begin{pmatrix} 7 & 2 \\ 2 & 1 \end{pmatrix} \quad \Sigma_2 = \frac{\sqrt{2}}{2}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

Considérons à présent la régularisation de l'échantillon $x_1, x_2, \ldots, x_N$ avec la méthode classique décrite précédemment ($\mu = 0.4$). On suppose ici l'utilisation d'un noyau de régularisation K gaussien, ce qui a peu d'influence sur le résultat (cf. 2.3). L'échantillon après régularisation est représenté sur la figure 4.3 : on constate que la distribution initiale est déformée après régularisation. En particulier, les ellipsoïdes des covariances des deux composantes ne sont plus orientées de la même façon. Cela est due à l'étape de blanchiment où le bruit de régularisation issu du noyau K est mis à l'échelle par la matrice de covariance empirique de $x_1, x_2, \ldots, x_N$.

Nous proposons d'utiliser la représentation de la loi conditionnelle $p_{k|k} = p(x_k|y_{0:k})$ sous forme de mélange de lois afin de proposer un algorithme de régularisation adapté aux distributions

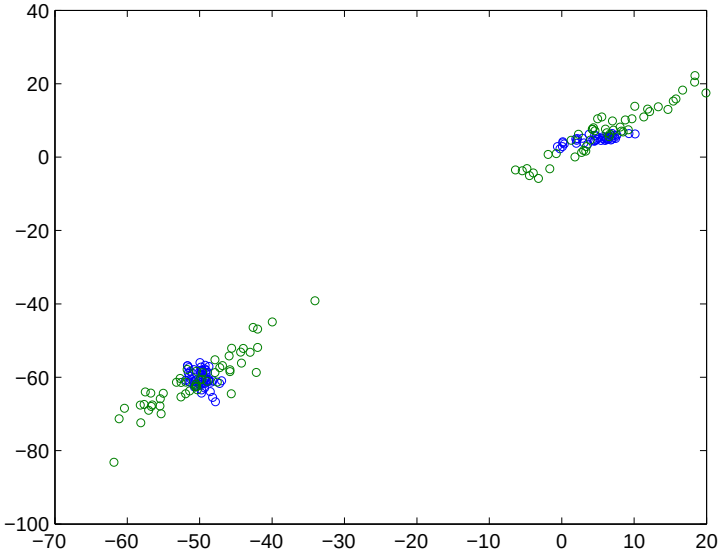


Figure 4.3 – Echantillon avant (bleu) et après régularisation globale (vert)

multimodales. Rappelons l'approximation particulaire pour les modèles de mélange, obtenue après l'étape de correction à l'instant k.

$$p_{k|k} \approx \hat{p}_{k|k} = \sum_{j=1}^{M_k} \alpha_{j,k} \hat{p}_j$$

avec

$$\hat{p}_j = \sum_{i \in I_{j,k}} \omega_k^i \delta_{x_k^i} \text{ et } \qquad \sum_{i \in I_{j,k}} \omega_k^i = 1$$

Dans l'algorithme du MPF, les étapes de prédiction/correction sont essentiellement celles du SIR à la différence près qu'une étape supplémentaire est ajoutée pour mettre à jour les coefficients de mélange $\alpha_{j,k}$. Outre l'introduction de ces coefficients, la différence avec le SIR est la possibilité de ré-échantillonner le système de particules localement. Concrètement, pour chaque composante $\hat{p}_j$, on peut définir une mesure de dégénérescence du système de particule comme la taille effective (locale) du cluster de particules correspondant à $\hat{p}_j$, N_{eff}^j, où l'entropie. Ces deux indicateurs sont définis au chapitre 2, section 2.2.3. Par exemple,

$$N_{eff}^j \approx \frac{1}{\sum_{i \in I_{j,k}} (\omega_k^i)^2} \tag{4.33}$$

Lorsque $N_{eff}^j / N_j < N_{th}/N$, les particules du cluster d'indice j subissent un ré-échantillonnage multinomial selon les poids d'importance de la composante $\hat{p}_j$, $\{\omega_k^i\}_{i \in I_{j,k}}$. N_{th} désigne le seuil de ré-échantillonnage.
Tout comme il est possible de ré-échantillonner localement grâce à la formulation (4.1), nous

proposons un algorithme de régularisation locale fondé sur le blanchiment local et de l'utilisation d'un paramètre de lissage propre à chaque cluster.
Soit $S_{j,k} = \sum_{i \in I_{j,k}} \omega_k^i (x_k^i - m_{k,j})(x_k^i - m_{k,j})^T$ la matrice de covariance empirique du cluster d'indice j et $m_{k,j} = \sum_{i \in I_{j,k}} \omega_k^i x_k^i$ sa moyenne empirique. Par souci de simplicité, nous supposerons que le noyau de régularisation K est gaussien. Soit le paramètre de lissage local au cluster $C_{j,k}$, h_{opt}^j défini par :

$$h_{opt}^j = D(K) N_j^{\frac{-1}{d+4}} \tag{4.34}$$

Soit $\tilde{x}_k^i$, $i \in I_{j,k}$, les particules après ré-échantillonnage multinomial local dans le cluster $C_{j,k}$. Alors les particules après régularisation locale sont obtenues selon l'équation :

$$\forall i \in I_{j,k}, \quad x_k^i = \tilde{x}_k^i + h_{opt}^j S_{j,k}^{\frac{1}{2}} \varepsilon^i \tag{4.35}$$

où $\varepsilon^i \underset{i.i.d.}{\sim} K$.
Dans la suite de l'exposé, nous appellerons MRPF (Mixture Regularized Particle Filter) la combinaison du MPF et de la routine de ré-échantillonnage/régularisation locale.

Afin d'illustrer l'impact de la régularisation locale, revenons à la distribution bi-modale ν définie par (4.32). La figure 4.4 représente le nuage de point avant et après régularisation locale. On peut constater sur cet exemple que la méthode développée permet de conserver la forme initiale de la distribution multimodale, contrairement à la régularisation globale.

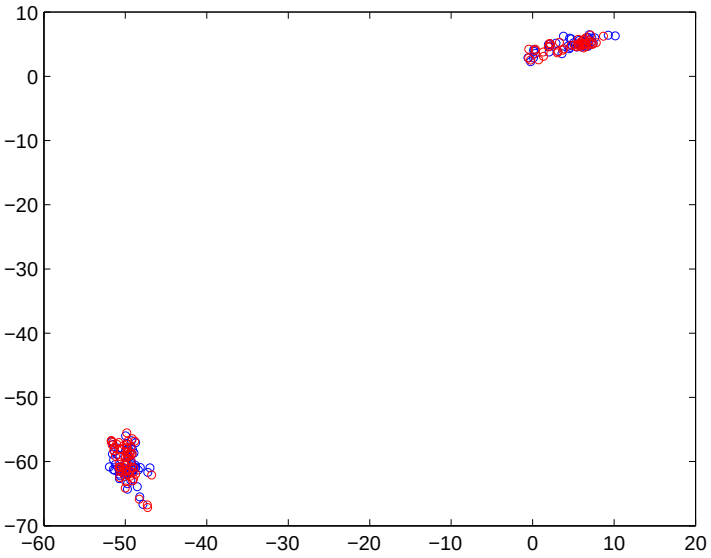


FIGURE 4.4 – Echantillon avant (bleu) et après régularisation locale (rouge)

Remarquons qu'il est possible d'effectuer une étape de régularisation locale dans un algorithme particulaire standard avec ré-échantillonnage global tel que le RPF. Dans ce cas, la succession des étapes de ré-échantillonnage et de régularisation se fait selon le schéma suivant.

Algorithme 5 : Ré-échantillonnage global et régularisation locale

- Calculer N_{eff} selon (2.19)
- Si $N_{eff} < N_{th}$ (N_{th} : seuil fixé par l'utilisateur)
  1. Identifier une partition des particules sous-forme de M_k clusters $C_{j,k}$, $j = 1, \ldots, M_k$ (avec l'algorithme de clustering mean-shift par exemple)
  2. Effectuer un ré-échantillonnage multinomial global selon les poids $\{\omega_k^i\}$. Noter I les indices des particules ré-échantillonnées.
  3. Pour $j = 1, \ldots, M_k$
     (a) Calculer la covariance $S_{j,k}$ et la moyenne $m_{j,k}$ des particules avant ré-échantillonnage
     (b) Calculer h_{opt}^j selon (4.34)
     (c) Identifier l'ensemble des indices J_I des particules ré-échantillonnées qui appartiennent au cluster $C_{j,k}$ i.e. $J_I = \{i \in I \mid x_k^i \in C_{j,k}\}$
     (d) Poser pour tout $i \in J_I$, $x_k^i = x_k^{I_i} + h_{opt}^j S_{j,k}^{\frac{1}{2}} \varepsilon^i$ avec $\varepsilon^i \underset{i.i.d.}{\sim} K$

4.3 Gestion dynamique des clusters du filtre

Dans la formulation originale du MPF [118], il peut arriver qu'une partie des ressources soit utilisée en pure perte pour mettre à jour les poids de mélange $\{\alpha_{j,k}\}$ et les poids intra-cluster ω_k^i associés à des clusters devenus très peu vraisemblables pendant plusieurs itérations. Il faut donc un schéma permettant de supprimer les clusters correspondant, et ré-allouer un nombre équivalent de particules aux clusters restant. Cependant, la littérature sur le sujet [118, 95, 126] ne détaille pas comment un tel mécanisme peut être mis en oeuvre.

Concrètement, nous suggérons de vérifier à chaque pas de temps k, la contribution de chaque cluster à l'approximation de la loi a posteriori $p(x_k|y_{0:k})$.
Sachant que

$$p(x_k|y_{0:k}) \approx \sum_{j=1}^{M_k} \alpha_{j,k} \sum_{i \in I_{j,k}} \omega_k^i \delta_{x_k^i}$$

si pour $j_0 \in [\![1, M_k]\!]$, $\alpha_{j,k} < \alpha_{min}$, on supprime les N_{j_0} particules du cluster $C_{j_0,k}$. Le paramètre α_{min} est un seuil fixé au préalable, typiquement pris inférieur à 10^{-5}. Il reste à échantillonner N_{j_0} nouvelles particules afin de conserver un système de particules de taille constante N. Cela peut être fait selon le schéma suivant :

1. Soit $\sum_{\substack{j=1 \\ j \neq j_0}}^{M_k} \pi_{j,k} \hat{p}_j$ la distribution empirique sans la composante $\hat{p}_{j_0}$, renormalisée avec $\forall j =$

$1, \ldots, N$, $j \neq j_0$

$$\pi_{j,k} = \frac{\alpha_{j,k}}{\sum\limits_{\substack{l=1 \\ l \neq j_0}} \alpha_{l,k}} \tag{4.36}$$

et

$$\hat{p}_j = \sum_{i \in I_{j,k}} \omega_k^i \delta_{x_k^i}$$

2. Faire un tirage multinomial de N_{j_0} indices $e_1, e_2, \ldots, e_{N_{j_0}}$ selon les poids $\pi_{j,k}$, $j \in \{1, \ldots, M_k\} \backslash \{j_0\}$
3. Pour $j = 1, \ldots, M_k$, $j \neq j_0$,
   (a) Soit $D_j = \mathrm{Card}\{l = 1, \ldots, N_{j_0} \mid e_l = j\}$ le nombre d'indices égaux à j
   (b) – Tirer D_j particules $\xi^1, \xi^2, \ldots, \xi^{D_j}$ selon $\hat{p}_j = \sum_{i \in I_{j,k}} \omega_k^i \delta_{x_k^i}$, en posant $\xi^i = x_k^q$ avec probabilité ω_k^q, $q \in I_{j,k}$.
   – Associer le poids $\gamma^i = \omega_k^q$ à ξ^i.
   – Rajouter les ξ^i au cluster $C_{j,k}$ et poser

$$\hat{p}_j = \frac{1}{\sum_{i \in I_{j,k}} \omega_k^i + \sum_{i=1}^{D_j} \gamma^i} \left(\sum_{i \in I_{j,k}} \omega_k^i \delta_{x_k^i} + \sum_{i=1}^{D_j} \gamma^i \xi^i \right)$$

4.4 Résumé de l'algorithme MRPF

Nous donnons ici une version de l'algorithme MRPF (Mixture Regularized Particle Filter) qui est une extension du MPF. La nouveauté est l'introduction d'une étape de régularisation locale et l'utilisation de l'algorithme mean-shift pour le clustering du nuage de particule. Par souci de simplicité, on suppose ici que la loi d'importance utilisée dans l'étape de prédiction est la transition $p(x_k|x_{k-1})$.

Algorithme 6 : MRPF

1. *[Initialisation]*
   – Tirer N particules selon la loi initiale $x_0^i \sim \mu_0$
   – poser $\omega_0^i \propto p(y_0|x_0^i)$
   – Normaliser les ω_0^i
2. Pour $k \geq 1$
   (a) *[Prédiction]* Tirer N particules $x_k^i \sim p(x_k|x_{k-1}^i)$
   (b) *[Correction]*
      i. Pour $j = 1, \ldots, M_{k-1}$
      – $\forall i \in I_{j,k-1}$, calculer

$$\tilde{\omega}_k^i = \omega_{k-1}^i p(y_k|x_k^i) \tag{4.37}$$

– Calculer les poids normalisés de la j-ème composante : $\forall i \in I_{j,k-1}$,

$$\omega_k^i = \frac{\tilde{\omega}_k^i}{\sum_{l \in I_{j,k-1}} \tilde{\omega}_k^l} \tag{4.38}$$

ii. Pour $j = 1, \ldots, M_{k-1}$, calculer $\alpha_{j,k}$ en fonction des $\{\alpha_{j,k-1}\}_{1 \leq j \leq M_{k-1}}$ selon (2.73) et (2.74)

(c) *[Clustering]* Décomposer le nuage de particules en M_k clusters $C_{j,k}$.
$[C'_{j,k}, I'_{j,k}] = \texttt{meanshiftClustering}(\{x_k^i\}_{i=1}^N, \{\omega_k^i\}_{i=1}^N, K, h^{MS}, R_m, n_{itermax}, \varepsilon^{MS})$

(d) Mettre à jour les coefficients de mélange $\alpha_{j,k}$ et les poids des particules intra-cluster ω_k^i pour la nouvelle partition $C'_{j,k}$ en utilisant les équations (2.84) et (2.85).

3. Poser $j = 1$. Tant que $j \leq M_k$
   – Si $\alpha_{j,k} < \alpha_{min}$ supprimer le cluster $C_{j,k}$ et échantillonner N_j nouvelles particules selon la procédure décrite dans la section 4.3
   – Sinon
     – Calculer N_{eff}^j selon (4.33)
     – Si $\frac{N_{eff}^j}{N_j}$ est inférieur au seuil de ré-échantillonnage, ré-échantillonner et régulariser les particules du cluster $C_{j,k}$ selon l'algorithme décrit dans la section 4.2
     – $j = j + 1$
4. Mettre à jour la fenêtre de lissage h^{MS} de l'algorithme de clustering mean-shift

---

On rappelle que les paramètres R_m, $n_{itermax}$ et ε^{MS} utilisés dans l'algorithme de clustering mean-shift représentent respectivement :

– le rayon de fusion, qui contrôle le degré de séparation des clusters
– le nombre d'itérations maximum autorisés par procédure mean-shift
– le seuil de convergence d'une procédure mean-shift

La routine de clustering peut être effectuée avant l'étape de correction.

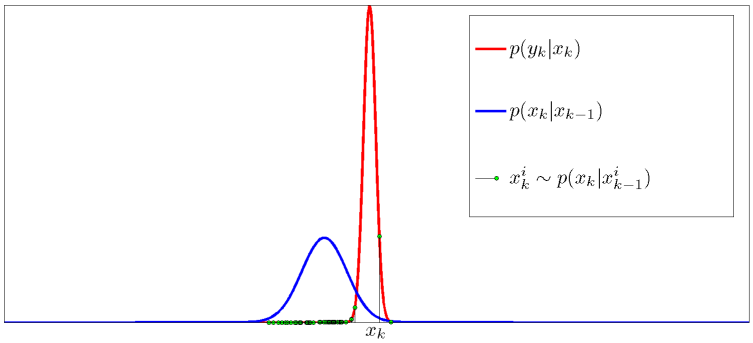
Par ailleurs, puisque les modèles de mélange du filtrage particulaire ne sont pas spécifiques à un filtre particulier, nous utiliserons également dans la suite de la thèse une extension naturelle du MRPF, le MRBPF, obtenu par Rao-Blackwellisation (cf. 2.4.1), en conservant l'étape de régularisation locale. Le MRPF et le MRBPF seront tous les deux combinés avec une procédure d'échantillonnage d'importance fondée sur les maxima locaux de la densité cible, ce qui est l'objet de la section suivante.

4.5 Échantillonnage d'une loi d'importance basée sur les maxima locaux de la densité a posteriori

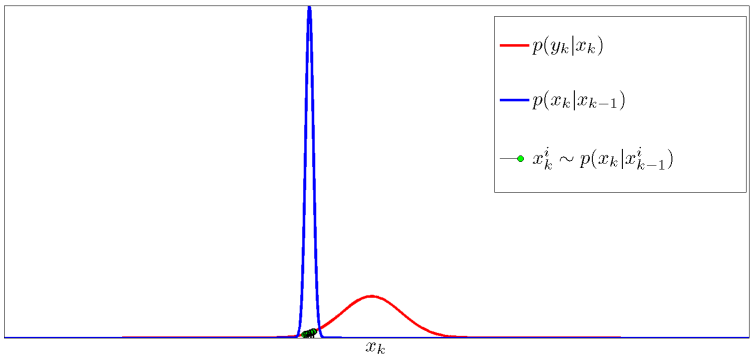
4.5.1 Motivation

Dans le chapitre précédent (sous-section 2.2.4), nous avons évoqué la nécessité de sélectionner une loi d'importance prenant en compte la mesure courante. En effet, lorsque la loi d'importance

dans l'étape de prédiction est le noyau de transition $p(x_k|x_{k-1})$, le risque est que les particules échantillonnées tombent dans la queue de distribution de la fonction de vraisemblance. Ceci arrive d'autant plus souvent que la fonction de vraisemblance décroit rapidement autour de son maximum, par exemple dans le cas d'un bruit additif v gaussien de variance petite, ou quand la fonction d'observation du modèle présente un fort gradient (cf. Figure 4.5 (a)). De même, lorsque le bruit de dynamique est additif gaussien de petite matrice de covariance Q, il est possible que, malgré l'étalement de la vraisemblance, $p(x_k|x_{k-1}^i)$ et $p(y_k|x_k)$ aient un faible recouvrement (cf. Figure 4.5 (b)).



(a) Cas d'une vraisemblance « pointue »



(b) Cas d'un a priori « pointu »

FIGURE 4.5 – Incohérence entre la fonction de vraisemblance et le noyau de transition

Parmi les approches algorithmiques permettant d'échantillonner les particules dans des régions de forte vraisemblance, on peut citer :

– *L'échantillonnage de la fonction de vraisemblance* [18] : si la fonction de vraisemblance $p(y_k|x_k)$ est intégrable en x_k [42], elle peut servir de loi d'importance lors de l'étape de prédiction et dans ce cas l'équation de mise à jour des poids est $\omega_k^i \propto \omega_{k-1}^i p(x_k^i|x_{k-1}^i)$. Cette technique est difficile à implémenter lorsque la dimension de l'espace d'observation est inférieure ou égale à celle de l'espace d'état ($m \leq d$) : en effet, dans ce cas on ne peut pas inverser la fonction de vraisemblance $x_k \mapsto p(y_k|x_k)$. En particulier, en recalage par

mesure radio-altimétriques, les mesures sont scalaires et l'état est de dimension supérieure ou égal à 3 donc la méthode n'est pas envisageable.

- *La correction progressive* [97] : cet algorithme permet de simuler un échantillon approximativement issu de $p(x_k|y_{0:k})$ en introduisant, à l'étape de correction, une série de fonctions de vraisemblance intermédiaires $g_{\lambda_m}(x_k) = p(y_k|x_k)^{1/\lambda_m}$, $m = 1, \ldots, n$ de moins en moins diffuses, où $\lambda_1, \ldots, \lambda_n$ est une suite de coefficients supérieurs ou égaux à 1 avec $\lambda_n = 1$, de sorte que $g_{\lambda_n}(x_k) = p(y_k|x_k)$. L'étape de correction est divisée en étapes intermédiaires. L'algorithme est initialisé avec les N particules $\{\xi_1^i, i = 1, \ldots, N\}$ obtenues lors de la phase de prédiction. Ensuite, pour $m = 2, \ldots, n$, les particules obtenues à l'étape $m-1$ sont pondérées avec la fonction de vraisemblance intermédiaire g_{λ_m}, ce qui donne des poids intermédiaires $\omega_{\lambda_m,k}^i$. Le nouveau nuage de particule est alors obtenu en tirant N échantillons $\{\xi_m^i, i = 1, \ldots, N\}$ selon la distribution régularisée $\sum_{i=1}^N \omega_{\lambda_m,k}^i K(\frac{x-\xi_{m-1}^i}{h})$, où K est un noyau de régularisation.
  Dans le cas d'un bruit gaussien additif, cette méthode revient à considérer une suite de densités gaussiennes de covariances décroissantes et supérieures à la vraie covariance. L'utilisation de vraisemblances intermédiaires de moins en moins étalées permet de ramener progressivement les particules dans les zones où la (vraie) fonction vraisemblance prend des valeurs significatives.
  Cette méthode a été testée sans succès.
- *Le filtre particulaire auxiliaire* [105] : cet algorithme permet de propager prioritairement les particules à l'instant $k-1$ qui seront compatibles avec la mesure y_k. L'étape de prédiction à l'instant k consiste à tirer N indices $i_1, \ldots, i_N$ selon une loi multinomiale dont les poids sont égaux à $\omega_{k-1}^i \hat{p}(y_k|x_{k-1}^i)$, où $\hat{p}(y_k|x_{k-1}^i)$ est une approximation de la vraisemblance prédite $p(y_k|x_{k-1}^i)$. Les particules prédites à l'instant k sont alors obtenues en échantillonnant $x_k^j \sim p(x_k|x_{k-1}^{i_j})$, $j = 1, \ldots, N$. Cela nécessite néanmoins d'avoir une approximation $\hat{p}(y_k|x_{k-1})$ qui soit plus diffuse que $p(y_k|x_{k-1}) = \int p(y_k|x_k)p(x_k|x_{k-1})dx_k$. Selon Whiteley et. al [121], les performances du filtre particulaire auxiliaire dépendent fortement du choix de l'approximation de la vraisemblance prédite, ce qui, selon le modèle, peut s'avérer une tâche compliquée en soit.

Nous proposons ci-après, une loi d'importance adaptée au MRPF, qui consiste à définir au niveau de chaque composante unimodale du filtre, une loi d'importance basée sur une approximation du maximum a posteriori (MAP) de la loi conditionnelle locale.
Pour cela, nous présentons d'abord plusieurs estimateurs d'importance utilisant une loi d'importance centrée sur le MAP dans un cadre non-dynamique. En particulier, nous présentons une méthode d'évaluation du MAP lorsque l'observation est linéaire par rapport à une partie de l'espace d'état et lorsque la loi a priori est gaussienne. Ensuite, nous discutons du choix de la matrice de covariance de la loi de proposition lorsque cette dernière est gaussienne. Enfin nous illustrons comment cette loi peut être mise en pratique dans un cadre particulaire (dynamique) pour les distributions multimodales.

4.5.2 Evaluation du maximum a posteriori pour un modèle d'observation partiellement linéaire avec a priori gaussien

Soit $x \in \mathbb{R}^d$ une v.a. gaussienne de moyenne μ et de covariance P, observée via une mesure $y \in \mathbb{R}^m$ corrompue par un bruit additif v.

$$y = H(x) + v \tag{4.39}$$

On suppose que v est un bruit gaussien de moyenne nulle et de covariance R. On souhaite calculer le maximum a posteriori $\hat{x}_{MAP}$ défini par

$$\hat{x}_{MAP} = \arg\max_{x \in \mathbb{R}^d} p(x|y) \tag{4.40}$$

où $p(x|y)$ est la densité conditionnelle de x sachant y. Le calcul du MAP est difficile lorsque la dimension de l'état est grande, ce qui est le cas dans notre application. C'est pourquoi nous proposons une méthode approchée du calcul du MAP en exploitant le fait que seule une petite partie du vecteur d'état participe à la non-linéarité de l'équation d'observation.

Nous considérons une équation de mesure générale avec une partie linéaire :

$$y = Ax_2 - h(x_1) + v \tag{4.41}$$

où x_1 et x_2 sont des sous-vecteurs de x, i.e. $x = (x_1^T, x_2^T)^T$, h une fonction non-linéaire de x_1 et A est une matrice indépendante de x.

La structure de la fonction de mesure (4.41) est caractéristique du recalage altimétrique : x_1 représente le vecteur des composantes horizontales, x_2 regroupe l'altitude et les composantes de vitesse et h correspond au modèle numérique de terrain. En posant $A = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$ on vérifie que $Ax_2 - h(x_1)$ correspond à la hauteur relative d'un mobile situé en x. Notons que le signe $(-)$ dans (4.41) est introduit par pure commodité puisque aucune hypothèse n'est faite sur la fonction h.
Dans la suite, on note q la densité de x et on notera $q(x) = q(x_1, x_2)$ pour faire apparaître la décomposition du vecteur $x = (x_1^T, x_2^T)^T$. De même, on note $\mu = (\mu_1^T, \mu_2^T)^T$, la moyenne de x avec $\mu_1 = \mathbb{E}(x_1)$ et $\mu_2 = \mathbb{E}(x_2)$. La covariance de x se décompose comme

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$$

Remarquons d'abord que

$$\begin{aligned} p(x|y) &= p(x_1, x_2|y) && (4.42)\\ &\propto p(y|x_1, x_2)q(x_1, x_2) && (4.43)\\ &\propto p(y|x_1, x_2)q(x_2|x_1)q(x_1) && (4.44) \end{aligned}$$

de sorte que

$$\begin{aligned} \max_{x_1,x_2} p(x_1, x_2|y) &\propto \max_{x_1} q(x_1) \max_{x_2} \{p(y|x_1, x_2)q(x_2|x_1)\} && (4.45)\\ &= \max_{x_1} q(x_1) \max_{x_2} F_1(x_1, x_2) && (4.46) \end{aligned}$$

La maximisation de la loi conditionnelle $p(x_1, x_2|y)$ se fait donc en maximisant d'abord la fonction $(x_1, x_2) \mapsto F_1(x_1, x_2) = p(y|x_1, x_2)q(x_2|x_1)$ par rapport à x_2. La maximisation de F_1 par rapport à x_2 est équivalente à la minimisation de

$$x_2 \mapsto Q(x_1, x_2) = -\frac{1}{2} \log F_1(x_1, x_2) \tag{4.47}$$

Etant donné que,

- $p(y|x_1, x_2)$ est une densité gaussienne de moyenne $y - (Ax_2 - h(x_1))$ et de covariance R
- $q(x_2|x_1)$ est une densité gaussienne de moyenne

$$\mathbb{E}(x_2|x_1) = \mathbb{E}(x_2) + P_{21}P_{11}^{-1}(x_1 - \mathbb{E}(x_1)) = \mu_2 + P_{21}P_{11}^{-1}(x_1 - \mu_1)$$

et de matrice de covariance

$$\Sigma_{2|1} = P_{22} - P_{21}P_{11}^{-1}P_{12}$$

on a

$$\begin{aligned} Q(x_1, x_2) = &(y + h(x_1) - Ax_2) R^{-1} (y + h(x_1) - Ax_2)^T \\ &+ (x_2 - \mathbb{E}_1(x_2)) \Sigma_{2|1}^{-1} (x_2 - \mathbb{E}_1(x_2))^T \end{aligned} \tag{4.48}$$

où $\mathbb{E}_1(x_2) = \mathbb{E}(x_2|x_1)$.
En regroupant les termes en x_2, on peut réécrire Q sous la forme

$$Q(x_1, x_2) = \Gamma(x_1) + (x_2 - \gamma(x_1)) \Omega^{-1} (x_2 - \gamma(x_1))^T \tag{4.49}$$

avec

$$\begin{aligned} \Gamma(x_1) &= [y + h(x_1) - A\gamma(x_1)] R^{-1} [y + h(x_1) - A\gamma(x_1)]^T \\ &+ [\gamma(x_1) - \mathbb{E}_1(x_2)] \Sigma_{2|1}^{-1} [\gamma(x_1) - \mathbb{E}_1(x_2)]^T \\ \gamma(x_1) &= \Omega \left[A^T R^{-1} (y + h(x_1)) + \Sigma_{2|1}^{-1} \mathbb{E}_1(x_2) \right] \\ \Omega &= \left(A^T R^{-1} A + \Sigma_{2|1}^{-1} \right)^{-1} \end{aligned}$$

L'espérance conditionnelle $\mathbb{E}_1(x_2) = \mathbb{E}(x_2|x_1)$ est bien une fonction de x_1 uniquement puisque $\mathbb{E}_1(x_2) = \mu_2 + P_{21}P_{11}^{-1}(x_1 - \mu_1)$. Les fonctions γ et Γ ne dépendent donc que de x_1. A x_1 fixé, Q atteint donc son minimum $\Gamma(x_1)$ en $\hat{x}_2 = \gamma(x_1)$. On a donc

$$\max_{x_2} F_1(x_1, x_2) = \exp\left(-\tfrac{1}{2} \min_{x_2} Q(x_1, x_2)\right) \tag{4.50}$$

$$= \exp\left(-\tfrac{1}{2}\Gamma(x_1)\right) \tag{4.51}$$

En reportant $\max_{x_2} F_1(x_1, x_2)$ dans (4.46), il reste donc à maximiser $q(x_1)\exp(-\frac{1}{2}\Gamma(x_1))$ par rapport à x_1. Etant donné que $q(x_1) = (2\pi)^{-1/2}\sqrt{\det(P)}\exp(-\frac{1}{2}(x_1 - \mu_1)^T P_{11}^{-1}(x_1 - \mu_1))$, cela revient à minimiser

$$G(x_1) = (x_1 - \mu_1)^T P_{11}^{-1}(x_1 - \mu_1) + \Gamma(x_1) \tag{4.52}$$

par rapport à x_1.

La minimisation de G relève de l'optimisation non-linéaire. Heureusement, dans l'application du recalage par mesures radio-altimétriques, le vecteur x_1 est un vecteur bi-dimensionnel regroupant les composantes de position horizontales. La recherche du minimum de G peut donc se faire facilement en utilisant les algorithmes issus de la méthode de Newton comme l'algorithme BFGS (Broyden-Fletcher-Goldfarb-Shanno) [10]. Une autre possibilité est d'utiliser un algorithme du gradient conjugé [54], qui évite le calcul et l'inversion de la hessienne de la fonction objectif.

En notant $\hat{x}_1 = \arg\min G(x_1)$, il est maintenant possible de calculer explicitement $\hat{x_2} = \gamma(\hat{x}_1)$.

Finalement, le MAP est donc estimé par $\hat{x}_{MAP} = (\hat{x}_1, \hat{x_2})$.

4.5.3 Loi de proposition centrée sur le MAP pour l'échantillonnage d'importance

4.5.3.1 Cadre statique

En échantillonnage d'importance, il est souhaitable de choisir une densité de proposition "proche" en un certain sens de la densité a posteriori $p(x|y)$. Si l'on suppose que $p(x|y)$ est unimodale, le choix d'une densité d'importance gaussienne centrée sur le MAP $\hat{x}_{MAP}$ semble judicieux. Ensuite, on peut par exemple prendre comme covariance de la loi de proposition, une approximation de la covariance a posteriori $\mathrm{Cov}(x|y)$.

La méthode de Laplace [66, 115] fournit une approximation au premier ordre de la covariance a posteriori en fonction du MAP et de la matrice d'information de Fisher observée J selon

$$\mathrm{Cov}(x|y) \approx J^{-1}(\hat{x}_{MAP}) \tag{4.53}$$

avec

$$J(x) = -\frac{\partial^2 \log g}{\partial x^2}(x) - \frac{\partial^2 \log q}{\partial x^2}(x) \tag{4.54}$$

On rappelle que $g(x) = p(y|x)$ et q est la densité a priori. Revenons au modèle défini par (4.39). Dans ce cas, q étant gaussien, J se réduit à

$$J(x) = 2\left[\frac{\partial^2 H}{\partial x^2} R^{-1}\left(y - H(x)\right) - R^{-1}\left(\frac{\partial H}{\partial x}\right)\left(\frac{\partial H}{\partial x}\right)^T\right] + P^{-1} \tag{4.55}$$

On voit que l'approximation (4.53) dépend essentiellement du gradient et de la hessienne de H. L'autre possibilité est de prendre une densité d'importance $\tilde{q}$ gaussienne de moyenne $\hat{x}_{MAP}$ et de covariance la covariance a priori P. Nous proposons de comparer ces deux approches dans un cadre statique proche de la problématique de recalage par mesures altimétriques. Pour cela nous considérons le problème d'estimation suivant :

$x = (x_1, x_2) \sim \mathcal{N}(\mu, P)$ est un vecteur statique dans le plan que nous cherchons à localiser. On suppose qu'une observation y est fournie via l'équation de mesure

$$y = H(x) + v$$

avec $H(x) = \frac{1}{2\pi|C|^{1/2}} \exp\left(-\frac{1}{2}(x-\theta)^T C^{-1}(x-\theta)\right)$. H est la densité d'une distribution gaussienne de moyenne θ et de matrice de covariance C avec

$$\theta = (1,1)^T$$

et

$$C = \begin{pmatrix} 0.0075 & 0.0025 \\ 0.0025 & 0.0075 \end{pmatrix}$$

La fonction d'observation H joue ici le rôle d'un relief artificiel dont le sommet correspond à la moyenne m et dont l'inclinaison autour du sommet est contrôlé par la matrice C. La matrice C permet de modéliser les situations où le terrain est subitement très informatif (présence d'un canyon,...). Le bruit d'observation suit une distribution gaussienne centrée d'écart-type $\sigma_v = 0.01$.

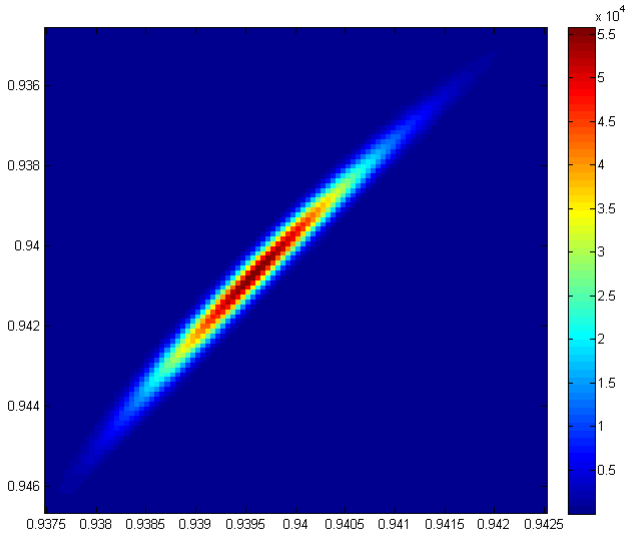


Figure 4.6 – densité a posteriori non normalisée

L'objectif est de calculer la moyenne a posteriori $\hat{x} = \mathbb{E}(x|y)$ par échantillonnage pondéré en utilisant trois lois de propositions différentes dont les densités sont notées $\tilde{q}_0$, $\tilde{q}_1$ et $\tilde{q}_2$. Dans la suite de cette section, $\varphi(.|m,\kappa)$ désignera la densité gaussienne de moyenne m et de covariance κ.

- $\tilde{q}_0 = q$ est la densité a priori pour laquelle on obtient l'estimateur Monte Carlo standard
- $\tilde{q}_1(x) = \varphi(x|\hat{x}_{MAP}, P)$
- $\tilde{q}_2(x) = \varphi(x|\hat{x}_{MAP}, J^{-1}(\hat{x}_{MAP}))$

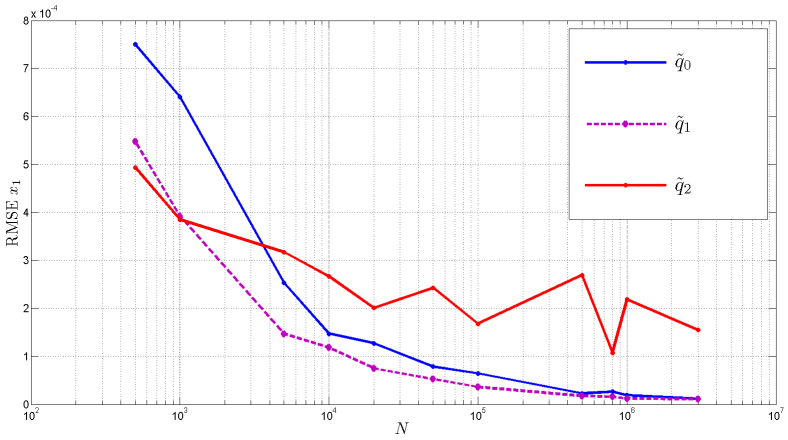
L'estimateur par échantillonnage pondéré de $\hat{x}$ en utilisant une densité d'importance $\tilde{q}$ est obtenu en tirant N échantillons indépendants $x^i \sim \tilde{q}$ et en calculant

$$\hat{x}^N = \sum_{i=1}^{N} \omega^i x^i \tag{4.56}$$

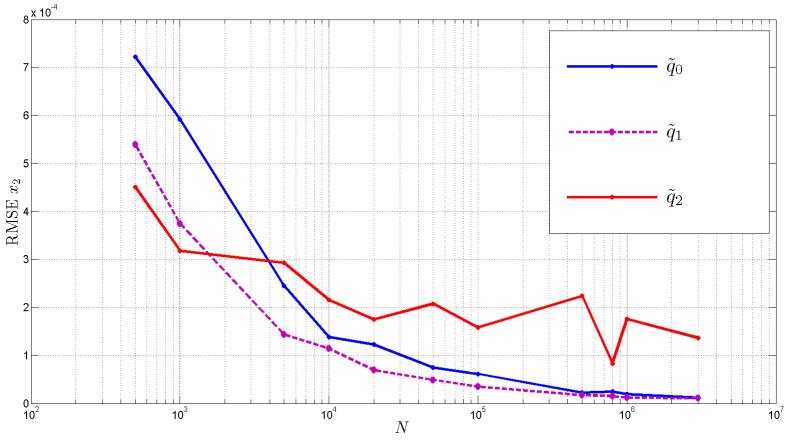
avec

$$\begin{aligned}\tilde{\omega}^i &= \frac{g(x^i)q(x^i)}{\tilde{q}(x^i)} \\ \omega^i &= \frac{\tilde{\omega}^i}{\sum_{j=1}^N \tilde{\omega}^j}\end{aligned} \tag{4.57}$$

Afin de comparer les différentes densités instrumentales $\tilde{q}_0$, $\tilde{q}_1$ et $\tilde{q}_2$, nous calculons $\hat{x}^N$ en fonction du nombre d'échantillons N et pour $\tilde{q} \in \{\tilde{q}_0, \tilde{q}_1, \tilde{q}_2\}$. Nous évaluons ensuite l'erreur quadratique moyenne (RMSE) $\mathbb{E}\left(\left\|\hat{x}^N - \hat{x}\right\|^2\right)^{1/2}$ en fonction de N, en utilisant 100 réalisations indépendantes. La valeur de référence $\hat{x} = \mathbb{E}(x|y)$ est calculée par intégration numérique.

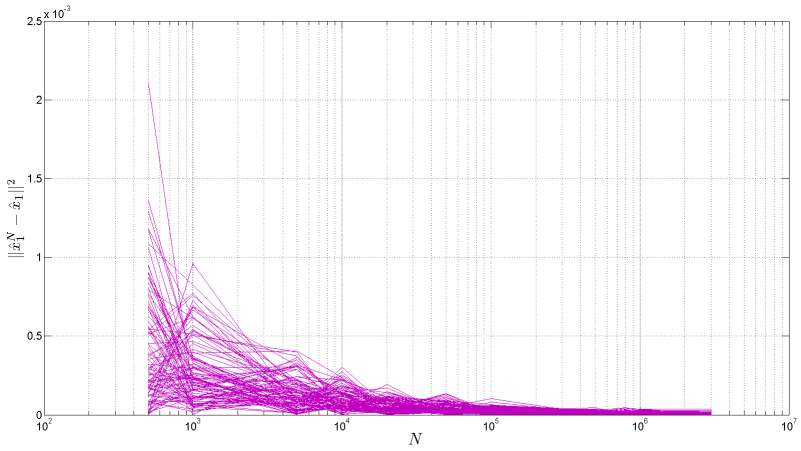


(a) RMSE x_1

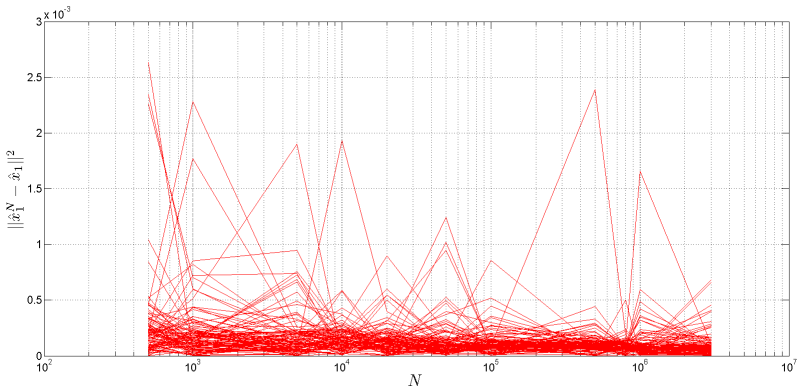


(b) RMSE x_2

FIGURE 4.7 – Comparaison de lois de proposition pour l'estimation de $\hat{x}$ par échantillonnage d'importance



(a) $\tilde{q}_1$



(b) $\tilde{q}_2$

Figure 4.8 – Réalisations de $\hat{x}^N$ en fonction de N

La figure 4.7 montre que l'utilisation d'une loi d'importance $\tilde{q}_1$ obtenu par translation de la loi a priori q sur le MAP réduit l'erreur d'estimation de l'estimateur Monte Carlo standard. Par exemple, il faut tirer au moins 10^4 échantillons selon q pour avoir la même erreur quadratique moyenne qu'un estimateur utilisant 3.10^3 échantillons selon $\tilde{q}_1$.

En revanche, le choix de la densité $\tilde{q}_2$ ne semble pas améliorer l'erreur d'estimation. Pire, l'estimateur correspondant semble ne pas converger. En effet, si l'on observe les réalisations de $\hat{x}^N$ à N fixé, on observe que sur les 100 tirages indépendants, il existe des valeurs aberrantes (figure 4.8). La raison de cette instabilité est que tous les choix de fonctions d'importance ne garantissent pas une variance de l'estimateur finie. La qualité d'un estimateur d'importance est

liée à la variance asymptotique $V_{\tilde{q}}$ des poids normalisés, dont l'expression est [31] :

$$V_{\tilde{q}} = \frac{1}{N}\left(\frac{\int_{\mathbb{R}^d}\frac{g^2(x)q^2(x)}{\tilde{q}(x)}dx}{\left(\int_{\mathbb{R}^d} g(x)q(x)dx\right)^2} - 1\right) \tag{4.58}$$

La fonction d'importance optimale est égale à la densité a posteriori. En effet, en posant $\tilde{q} \propto gq$, on a $V_{\tilde{q}} = 0$. $V_{\tilde{q}}$ est finie si et seulement si $\int \frac{g^2(x)q^2(x)}{\tilde{q}(x)}dx < +\infty$. Une condition suffisante est que le rapport $\frac{q}{\tilde{q}}$ soit borné. En effet, le produit gq est intégrable, la vraisemblance g est bornée donc si $\sup q/\tilde{q} < +\infty$, alors

$$\int_{\mathbb{R}^d}\frac{g^2(x)q^2(x)}{\tilde{q}(x)}dx \leq \sup_{x\in\mathbb{R}^d} g \sup_{x\in\mathbb{R}^d}\frac{q}{\tilde{q}}\int_{\mathbb{R}^d} g(x)q(x)dx < +\infty$$

Dans le cas de l'utilisation de la fonction d'importance $\tilde{q} = \tilde{q}_2$, on voit que le rapport $\frac{q}{\tilde{q}_2}$ peut être très grand lorsque la covariance de $\tilde{q}_2$, $J^{-1}(\hat{x}_{MAP})$, est petite devant la covariance a priori P, de sorte que le numérateur de $V_{\tilde{q}}$ n'est pas fini. Lors du calcul de $\hat{x}^N$ selon (4.56), quelques tirages peuvent avoir un poids ω^i très important et avoir trop d'impact sur l'estimation finale.
Afin de garantir la convergence de l'estimateur, on peut alors chercher une condition suffisante pour assurer que $\frac{q}{\tilde{q}}$ soit majoré lorsque $\tilde{q}$ est une densité gaussienne de moyenne $\hat{x}_{MAP}$ et de covariance à déterminer Σ. On maintient l'hypothèse gaussienne sur la densité a priori q, de moyenne μ et de covariance P.
Le rapport $\frac{q}{\tilde{q}}$ est borné ssi $\ln(\frac{q}{\tilde{q}})$ borné.
Etant donné que $q(x) \propto e^{-\frac{1}{2}(x-\mu)^T P^{-1}(x-\mu)}$ et $\tilde{q}(x) \propto e^{-\frac{1}{2}(x-\hat{x}_{MAP})^T\Sigma^{-1}(x-\hat{x}_{MAP})}$, on a

$$\begin{aligned}
\ln\frac{q(x)}{\tilde{q}(x)} &= c_1 + \frac{1}{2}\left(x-\hat{x}_{MAP}\right)^T\Sigma^{-1}\left(x-\hat{x}_{MAP}\right) - \frac{1}{2}\left(x-\mu\right)^T P^{-1}\left(x-\mu\right)\\
&= c_1 + \frac{1}{2}\left(x-\mu+\mu-\hat{x}_{MAP}\right)^T\Sigma^{-1}\left(x-\mu+\mu-\hat{x}_{MAP}\right) - \frac{1}{2}\left(x-\mu\right)^T P^{-1}\left(x-\mu\right)\\
&= c_1 + \frac{1}{2}\left(x-\mu\right)^T\Sigma^{-1}\left(x-\mu\right) + \frac{1}{2}\left(\mu-\hat{x}_{MAP}\right)^T\Sigma^{-1}\left(x-\mu\right)\\
&+ \underbrace{\frac{1}{2}\left(\mu-\hat{x}_{MAP}\right)^T\Sigma^{-1}\left(\mu-\hat{x}_{MAP}\right)}_{\text{constante}}\\
&+ \frac{1}{2}\left(x-\mu\right)^T\Sigma^{-1}\left(\mu-\hat{x}_{MAP}\right) - \frac{1}{2}\left(x-\mu\right)^T P^{-1}\left(x-\mu\right)\\
&= c_2 + \frac{1}{2}\left(x-\mu\right)^T\left(\Sigma^{-1}-P^{-1}\right)\left(x-\mu\right) + \frac{1}{2}\left(\underbrace{u(x-\mu)}_{\text{scalaire}} + \underbrace{(x-\mu)^T u^T}_{\text{scalaire}}\right)\\
&= \frac{1}{2}\left(x-\mu\right)^T\left(\Sigma^{-1}-P^{-1}\right)\left(x-\mu\right) + u_1(x-\mu) + c_2 \qquad (4.59)\\
&= \frac{1}{2}x^T\left(\Sigma^{-1}-P^{-1}\right)x + u_2 x + c_3 \qquad (4.60)\\
& \qquad (4.61)
\end{aligned}$$

où $u_1 = (\mu - \hat{x}_{MAP})^T \Sigma^{-1}$, u_2, c_1, c_2 et c_3 des constantes. $\ln \frac{q}{\tilde{q}}$ est majoré ssi le terme $K(x) = \frac{1}{2} x^T \left(\Sigma^{-1} - P^{-1}\right) x + u_2 x$ est majoré. Ce dernier terme est majoré ssi sa limite vaut $-\infty$ lorsque $\|x\| \to +\infty$. Une condition suffisante est que $P^{-1} - \Sigma^{-1}$ soit définie positive.
En effet, dans ce cas on a

$$K(x) = -\frac{1}{2} (x-a)^T \left(P^{-1} - \Sigma^{-1}\right) (x-a) + c_4$$

avec

$$a = \mu + \left(\Sigma^{-1} - P^{-1}\right)^{-1}$$
$$c_4 = \frac{1}{2} a^T \left(\Sigma^{-1} - P^{-1}\right) a + u_2 a$$

En posant $x' = (P^{-1} - \Sigma^{-1})^{1/2}(x-a)$ (possible car on suppose $P^{-1} - \Sigma^{-1} > 0$), on a $K(x) = -\frac{1}{2}\|x'\|^2 + c_4$. Comme $\lim_{\|x\| \to +\infty} \|x'\| = +\infty$, on a $\lim_{\|x\| \to +\infty} K(x) = -\infty$.
Il s'en suit que si la matrice $P^{-1} - \Sigma^{-1}$ est définie positive, alors la forme quadratique $\ln(\frac{q}{\tilde{q}})$ est majorée.
Cherchons maintenant une caractérisation de la propriété $P^{-1} - \Sigma^{-1} > 0$.

$$\begin{aligned} \Sigma^{-1} - P^{-1} < 0 &\Leftrightarrow x^T(\Sigma^{-1} - P^{-1})x < 0 \quad \forall x \in \mathbb{R}^d \setminus \{0\} \\ &\Leftrightarrow \frac{x^T \Sigma^{-1} x}{x^T P^{-1} x} < 1 \quad \forall x \in \mathbb{R}^d \setminus \{0\} \\ &\Leftrightarrow \sup_{x \in \mathbb{R}^d \setminus \{0\}} \frac{x^T \Sigma^{-1} x}{x^T P^{-1} x} < 1 \quad \forall x \in \mathbb{R}^d \setminus \{0\} \end{aligned} \tag{4.62}$$

Soit $P^{-1} = DD^T$ la décomposition de Choleski de P^{-1}. On pose $C = D^{-T} \Sigma^{-1} D^{-1}$ avec $D^{-T} = (D^{-1})^T$. Alors,

$$\frac{x^T \Sigma^{-1} x}{x^T P^{-1} x} = \frac{x^T D^T C D x}{x^T D^T D x}$$

En posant $y = Dx$,

$$\frac{x^T \Sigma^{-1} x}{x^T P^{-1} x} = \frac{y^T C y}{y^T y}$$

et

$$\sup_{x \in \mathbb{R}^d \setminus \{0\}} \frac{x^T \Sigma^{-1} x}{x^T P^{-1} x} < 1 \Leftrightarrow \sup_{y \in \mathbb{R}^d \setminus \{0\}} \frac{y^T C y}{y^T y} < 1 \tag{4.63}$$

(En effet, D étant inversible, y décrit $\mathbb{R}^d \setminus \{0\}$ lorsque x décrit $\mathbb{R}^d \setminus \{0\}$). En remarquant que $\frac{y'^T C y'}{y'^T y'} = \frac{y^T C y}{y^T y}$ pour $x' = cx$ avec $c > 0$, (4.63) est équivalent à

$$\sup_{\|y\|=1} y^T C y < 1$$

Il faut donc résoudre le problème suivant : Maximiser $y^T C y$ sous la contrainte $\|y\| = 1$. On introduit le multiplicateur de Lagrange

$$L(y, \lambda) = y^T C y + \lambda (y^T y - 1)$$

$$\begin{cases} \frac{\partial L}{\partial y} & = 2y^T C + 2\lambda y^T \\ \frac{\partial L}{\partial \lambda} & = y^T y - 1 \end{cases}$$

$$\begin{aligned} \frac{\partial L}{\partial y} = 0 &\Leftrightarrow 2y^T C + 2\lambda y^T = 0 \\ &\Leftrightarrow y^T C = -\lambda y^T \\ &\Leftrightarrow Cy = -\lambda y \\ &\Leftrightarrow Cy = \lambda' y \end{aligned}$$

avec $\lambda' = -\lambda$. Les extrema de $y^T C y$ sont donc obtenus pour y vecteur propre de C. Comme par ailleurs $y \mapsto y^T C y$ est continue sur le compact $\{y \in \mathbb{R}^d | \|y\| = 1\}$, le maximum est bien atteint pour un vecteur propre unitaire e de C tel que $Ce = \lambda' e$. Comme $e^T C e = \lambda' e^T e = \lambda'$, le maximum de $y^T C y$ est la valeur propre maximale de C.
Finalement,

Proposition 1. *Si la valeur propre maximale de $D^{-T}\Sigma^{-1}D^{-1}$ est inférieure à 1 alors le rapport $\frac{q}{\tilde{q}}$ est majoré.*

Choix de la covariance de la loi de proposition On a vu sur un exemple synthétique, que l'utilisation directe d'une loi de proposition gaussienne de covariance donnée par l'approximation de Laplace $J^{-1}(\hat{x}_{MAP})$ de la covariance a posteriori, peut donner un estimateur de mauvaise qualité. Néanmoins, afin d'exploiter l'information donnée par la matrice $\hat{J}^{-1} = J^{-1}(\hat{x}_{MAP})$, nous suggérons d'utiliser une loi de proposition gaussienne $\tilde{q}$ dont la matrice de covariance Σ a pour valeurs propres les valeurs propres de P et pour vecteurs propres, les vecteurs propres de $\hat{J}^{-1}$: schématiquement, cela revient à aligner l'ellipsoïde de confiance de la covariance Σ sur celle de $\hat{J}^{-1}$. De cette façon nous nous attendons à générer les échantillons selon les axes principaux de $\hat{J}^{-1} \approx \mathrm{Cov}(x|y)$. Ainsi, ces échantillons sont orientés conformément à la matrice de covariance de la fonction d'importance optimale.
Soit E_J la matrice des vecteurs propres de $\hat{J}^{-1}$. Soit Λ_P la matrice des valeurs propres de P. On construit la matrice P_{rot} définie par :

$$P_{rot} \triangleq E_J \Lambda_P E_J^T \tag{4.64}$$

et on pose $\Sigma = P_{rot}$.

D'après la proposition précédente, ceci ne garantit cependant pas que la variance asymptotique $V_{\tilde{q}}$ des poids d'importance ω^i soit finie, même si en pratique nous avons observé que l'estimateur correspondant était stable. On peut tout de même chercher une matrice de covariance sous la forme $\Sigma = sP_{rot}$ avec $s > 0$ et trouver le plus petit réel s qui garantissent que $\frac{q}{\tilde{q}}$ soit borné. D'après la proposition 1, cela revient à déterminer

$$s^* = \inf\{s > 0 \mid \lambda_{\max}(\frac{1}{s} D^{-T} P_{rot}^{-1} D^{-1}) < 1\}$$

où $\lambda_{\max}(B)$ désigne la plus grande valeur propre de la matrice B lorsqu'elle existe. On en déduit que

$$s^* = \lambda_{\max}(D^{-T} P_{rot}^{-1} D^{-1}) \tag{4.65}$$

Pour tout $s > s^*$, la variance $V_{\tilde{q}}$ sera finie si $\tilde{q}$ est une densité gaussienne de moyenne $\hat{x}_{MAP}$ et de covariance $\Sigma = sP_{rot}$.

Une autre approche permettant d'exploiter l'approximation $\hat{J}^{-1} \approx \mathrm{Cov}(x|y)$ consiste à chercher la matrice de covariance Σ^* la plus proche de $\hat{J}^{-1}$ en norme, telle que $\frac{q}{\tilde{q}}$ soit bornée. Autrement dit,

$$\Sigma^* = \arg\min_{\Sigma} \|\Sigma - \hat{J}^{-1}\|_F^2\} \tag{4.66}$$

sous la contrainte $\Sigma^{-1} - P^{-1} < 0$. La norme matricielle $\|.\|_F$ est définie pour toute matrice réelle $A = (a_{i,j})_{i,j}$ de taille $d \times d$ par

$$\|A\|_F = \left(\sum_{i,j} a_{i,j}^2 \right)^{\frac{1}{2}} \tag{4.67}$$

Remarquons que $\Sigma^{-1} - P^{-1} < 0$ est équivalent à $\Sigma - P > 0$ (cf. annexe C).

Théorème 4 (N. Higham [55]). *Soit A est une matrice réelle symétrique. Le problème suivant :*

$$\textit{minimiser } \|X - A\|_F^2 \textit{ sous la contrainte } X \geq 0$$

admet une unique solution X^* de la forme

$$X^* = \frac{1}{2}(A + H) \tag{4.68}$$

où H est une matrice symétrique positive telle que $A = UH$ avec $U^T U = I$. L'égalité $A = UH$ est appelé décomposition polaire de A.

On peut alors résoudre

$$\min_{\Sigma - P \geq 0} \|\Sigma - \hat{J}^{-1}\|_F^2$$

en remarquant que $\|\Sigma - \hat{J}^{-1}\|_F^2 = \|(\Sigma - P) - (\hat{J}^{-1} - P)\|_F^2$. On a alors,

$$\min_{\Sigma - P \geq 0} \|\Sigma - \hat{J}^{-1}\|_F^2 = \min_{X \geq 0} \|X - (\hat{J}^{-1} - P)\|_F^2$$

En notant $X_F = \arg\min_{X \geq 0} \|X - (\hat{J}^{-1} - P)\|$, d'après le théorème précédent, on a

$$X_F = \frac{1}{2}(\hat{J}^{-1} - P + H)$$

avec $\hat{J}^{-1} - P = UH$ et $U^T U = I$. On en déduit que $\Sigma_F = \arg\min_{\Sigma - P \geq 0} \|\Sigma - \hat{J}^{-1}\|_F^2$ est donné par

$$\Sigma_F = \frac{1}{2}(\hat{J}^{-1} + P + H) \tag{4.69}$$

Cette formule est relativement simple à mettre en oeuvre pour peu que l'on dispose d'une matrice $H = H^T \geq 0$ vérifiant $\hat{J}^{-1} - P = UH$ où U est une matrice unitaire. Heureusement, il existe une alternative à la décomposition polaire permettant de calculer Σ_F.
Si on diagonalise $\hat{J}^{-1} - P$ sous la forme $\hat{J}^{-1} - P = Z\tilde{\Lambda}Z^T$ avec $\tilde{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ matrice diagonale et $ZZ^T = I$. En posant

$$\tilde{\Lambda}_+ = \mathrm{diag}(\max(0, \tilde{\lambda}_1), \ldots, \max(0, \tilde{\lambda}_d))$$

on a [55],

$$\Sigma_F = Z\tilde{\Lambda}_+ Z^T + P \tag{4.70}$$

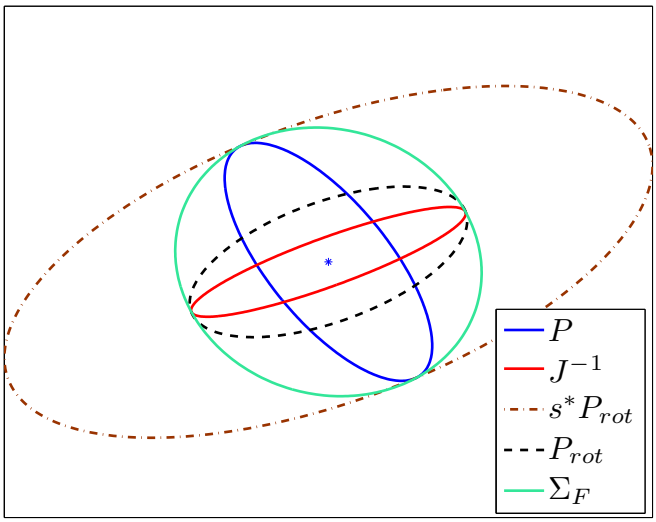


FIGURE 4.9 – Ellipsoïdes de confiance à 99 % des matrices de covariance P, J^{-1}, P_{rot}, s^*P_{rot} et Σ_F

Enfin, pour conserver une variance d'estimation finie, on peut également échantillonner selon une loi de Student multivariée, dont la queue de distribution est plus lourde et garantit automatiquement $\sup \frac{q}{\tilde{q}} < +\infty$. On rappelle qu'une v.a X de $\mathbb{R}^d$ suit une distribution de Student multivariée $T_d(\nu, \mu, \Sigma)$ si elle admet pour fonction de densité

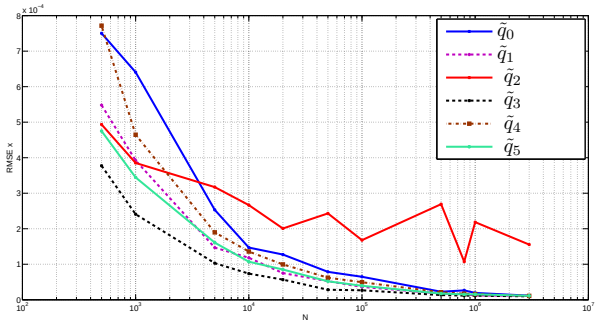
$$t_d(x; \nu, \mu, \Sigma) = \frac{\Gamma(\frac{\nu+d}{2})(\pi d)^{-\frac{d}{2}}|\Sigma|^{-1/2}}{\Gamma(\frac{\nu}{2})\left[1 + (x-\mu)^T\Sigma^{-1}(x-\mu)/\nu\right]^{\frac{\nu+d}{2}}} \tag{4.71}$$

$\nu \in \mathbb{N}_+^*$ désigne le degré de liberté, μ est un paramètre de localisation et Σ une matrice de produit scalaire (symétrique, définie positive) [70]. Ici, Γ désigne la fonction Gamma d'Euler.
X a pour moyenne μ (resp. matrice de covariance) uniquement lorsque $\nu > 1$ (resp. $\nu > 2$) et, lorsqu'elle existe, sa covariance est donné par $\mathrm{Cov}(X) = \frac{\nu}{\nu-2}\Sigma$. Plus ν est petit, plus la queue de la distribution de Student est lourde.

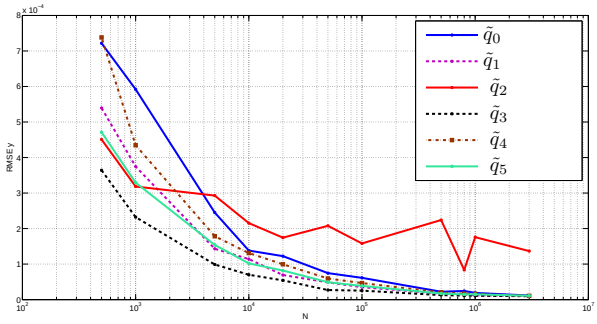
Comparaison des densités d'importances Dans un premier temps, nous avons évalué les 5 densités d'importances gaussiennes centrées sur une approximation du MAP précédemment :

- $\tilde{q}_1(x) = \varphi(x|\hat{x}_{MAP}, P)$
- $\tilde{q}_2(x) = \varphi(x|\hat{x}_{MAP}, \hat{J}^{-1})$
- $\tilde{q}_3(x) = \varphi(x|\hat{x}_{MAP}, P_{rot})$
- $\tilde{q}_4(x) = \varphi(x|\hat{x}_{MAP}, s^* P_{rot})$ (cf. (4.65))
- $\tilde{q}_5(x) = \varphi(x|\hat{x}_{MAP}, \Sigma_F)$

Nous nous proposons de les comparer dans le problème d'estimation statique présenté précédemment. Plus précisément, pour chaque densité $\tilde{q} \in \{\tilde{q}_1, \ldots, \tilde{q}_5\}$ on calcule l'estimateur $\hat{x}^N$ de $\mathbb{E}(x|y)$ en fonction du nombre d'échantillons N. On peut alors en déduire l'erreur quadratique moyenne de chaque estimateur en fonction de N.



(a) RMSE x_1

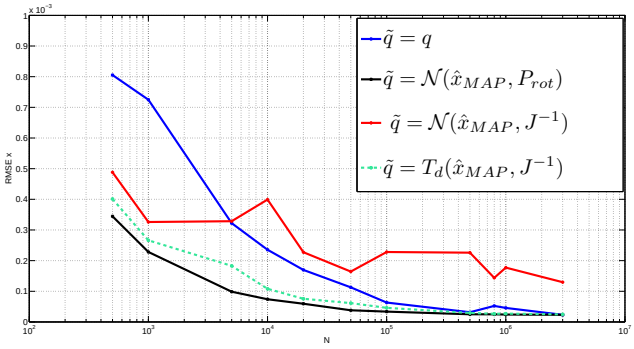


(b) RMSE x_2

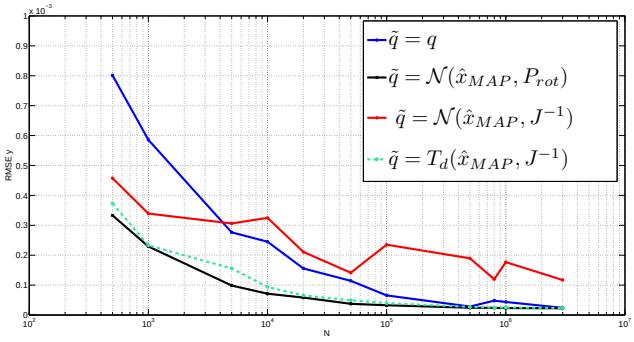
Figure 4.10 – Comparaison de lois de proposition pour l'estimation de $\mathbb{E}(x|y)$ par échantillonnage d'importance : densité instrumentale gaussienne

On constate sur la figure 4.10 que le paramétrage le plus avantageux est donné par $\tilde{q}_3$, c'est-à-dire la densité gaussienne centrée sur le MAP et de covariance P_{rot}. Les densités d'importance $\tilde{q}_5$ et $\tilde{q}_1$ centrées sur le MAP et de covariances respectives P et Σ_F sont relativement proches en termes d'erreur quadratique moyenne ($\tilde{q}_5$ étant légèrement meilleure). En revanche, l'utilisation de la covariance $s^* P_{rot}$ dans la loi instrumentale n'améliore que marginalement le RMSE. $\tilde{q}_4$ garantit une variance d'estimation finie mais s'éloigne de la densité a posteriori comme l'illustre la figure 4.9.

Dans un second temps, nous avons examiné l'utilisation d'une densité d'importance suivant une loi de Student $T_d(\nu, \hat{x}_{MAP}, \Sigma)$. La matrice Σ a été choisie de sorte que $\mathrm{Cov}(X) = J^{-1}(\hat{x}_{MAP})$, i.e. $\Sigma = \frac{\nu-2}{\nu} J^{-1}(\hat{x}_{MAP})$, et le degré de liberté ν a été fixé expérimentalement à 4. Ce choix donne des résultats satisfaisants comme l'atteste la figure 4.11.



(a) RMSE x_1



(b) RMSE x_2

FIGURE 4.11 – Comparaison de lois de proposition pour l'estimation de $\mathbb{E}(x|y)$ par échantillonnage d'importance : densité instrumentale Student

4.5.3.2 Cadre dynamique : illustration

On considère un problème de pistage de cible dans le plan cartésien par mesure de distance et d'azimuth. La cible suit une dynamique linéaire suivante :

$$x_k = F x_{k-1}$$

où

$$F = \begin{pmatrix} I_2 & \Delta I_2 \\ 0_2 & I_2 \end{pmatrix}$$

Δ désigne le pas de temps et $x_k = [x_{k,1}, x_{k,2}, \dot{x}_{k,1}, \dot{x}_{k,2}]^T$ est le vecteur d'état comprenant la position et la vitesse horizontale de la cible.

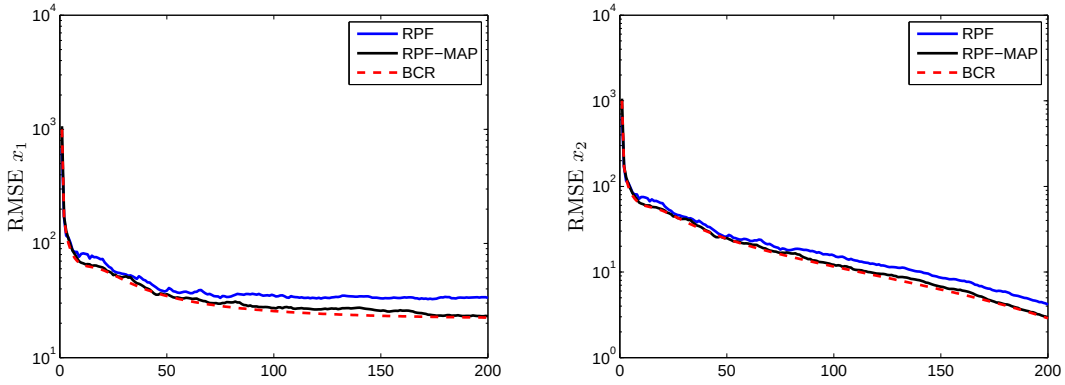
Le capteur radar est supposé fixe à la position $(0, 0)$ et mesure à chaque instant k une distance $r = \sqrt{x_{k,1}^2 + x_{k,2}^2}$ et un azimuth $\theta = \arctan(\frac{x_{k,2}}{x_{k,1}})$ tous deux corrompus par un bruit de mesure additif. On note y_k le vecteur de mesure correspondant dont l'expression est

$$y_k = \begin{bmatrix} \sqrt{x_{k,1}^2 + x_{k,2}^2} \\ \arctan(\frac{x_{k,2}}{x_{k,1}}) \end{bmatrix} + v_k$$

où $v_k \sim \mathcal{N}\left(0_{2\times 1}, \mathrm{diag}(\sigma_r^2, \sigma_\theta^2)\right)$ est le bruit de mesure gaussien. σ_r et σ_θ désignent respectivement l'écart-type du bruit sur la mesure de distance et sur la mesure d'angle.

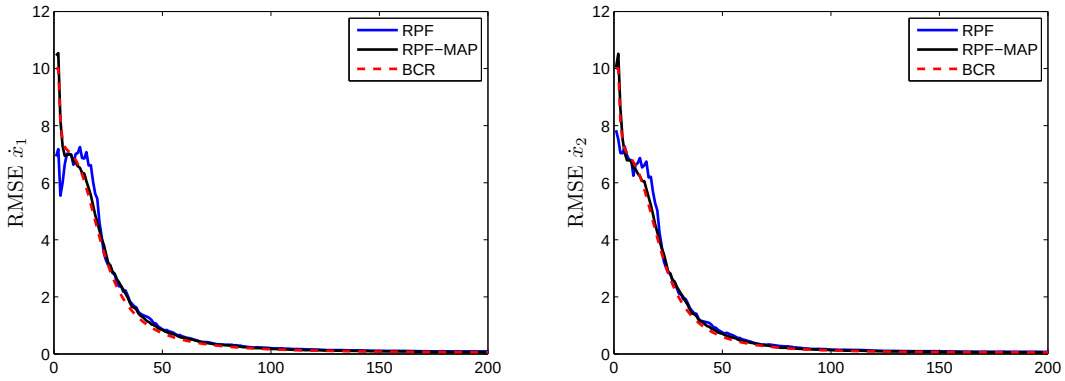
L'objectif est d'estimer la position et la vitesse de la cible sachant les mesures passées en calculant l'espérance conditionnelle $\mathbb{E}(x_k|y_{0:k})$. Pour cela, nous allons comparer le comportement du filtre particulaire régularisé (cf. 2.3), avec un filtre particulaire régularisé dont la fonction d'importance est une gaussienne multivariée centrée sur une estimation du MAP, et de matrice de covariance P_{rot} définie dans la section 4.5.3.1 : ce dernier algorithme est noté RPF-MAP. Les paramètres du capteur sont fixés à $\sigma_r = 5\,m$ et $\sigma_\theta = 1°$. L'erreur en distance est faible, ce qui est un cas difficile pour le filtrage particulaire classique.

La cible a pour vecteur position-vitesse initiale $x_{cible,0} = \begin{bmatrix} 10000 & 10000 & 40 & -40 \end{bmatrix}^T$. On suppose également que x_0 suit une loi gaussienne de moyenne $\bar{x}_0$ et de covariance $P_0 = \mathrm{diag}\left((1000\mathrm{m})^2 \quad (1000\mathrm{m})^2 \quad (10\mathrm{m/s})^2 \quad (10\mathrm{m/s})^2\right)$. La moyenne $\bar{x}_0$ est tirée selon une gaussienne centrée sur la position initiale de la cible $x_{cible,0}$ de covariance P_0.
Nous avons calculé l'erreur quadratique moyenne des trajectoires non-divergentes sur la base de 200 simulations ainsi que le taux de non-divergence des deux algorithmes : le RPF-MAP affiche un taux de non-divergence de 97% tandis que le RPF standard a un taux de non-divergence de 24%. Nous avons considéré qu'un filtre a divergé si l'état final vrai sort de l'ellipsoide de confiance associée à la Borne de Cramer-Rao a posteriori au seuil 99.99% et centrée sur l'état estimé, ceci pendant les 5 derniers instants de la trajectoire.
Par ailleurs, on observe (figure 4.12) que la position horizontale est estimée avec plus de précision par le RPF-MAP pour ce capteur : l'erreur quadratique moyenne des trajectoires convergentes du RPF-MAP s'approche rapidement de la borne de Cramer-Rao à posteriori tandis que celle du RPF ne suit pas tout à fait cette borne.



(a) RMSE x_1 (b) RMSE x_2

FIGURE 4.12 – Erreur quadratique moyenne pour la position ; bleu : RPF, noir : RPF-MAP, traits pointillés rouges : Borne de Cramer-Rao a posteriori



(a) RMSE $\dot{x}_1$ (b) RMSE $\dot{x}_2$

FIGURE 4.13 – Erreur quadratique moyenne pour la vitesse ; bleu : RPF, noir : RPF-MAP, traits pointillés rouges : Borne de Cramer-Rao a posteriori

4.5.3.3 Application au filtrage particulaire pour les modèles d'observation partiellement linéaires

Dans cette section nous détaillons comment définir une fonction d'importance basée sur les maxima locaux estimés de la densité a posteriori $p(x_k|y_{0:k})$. Pour cela nous nous appuyons sur les résultats établis précédemment pour l'échantillonnage d'importance dans un cadre statique. Nous faisons l'hypothèse que l'équation d'observation est partiellement linéaire, ce qui est le cas en recalage altimétrique :

$$y_k = Ax_k^{(2)} - h_k^n(x_k^{(1)}) + v_k \tag{4.72}$$

où $v_k \sim \mathcal{N}(0, R)$ et A est une matrice indépendante de x_k. Le vecteur d'état se décompose sous la forme d'une partie linéaire $x_k^{(2)}$ et non-linéaire $x_k^{(1)}$, $x_k = \left[x_k^{(1),T}, x_k^{(2),T}\right]^T$. Enfin, h_k^n est une fonction non-linéaire de $x_k^{(1)}$.

Nous nous plaçons dans le cadre du MRPF (Mixture Regularized Particle Filter) introduit précédemment.

Rappelons dans le MRPF, la densité a posteriori $p(x_k|y_{0:k})$ s'écrit comme un mélange de densités :

$$p_{k|k} = p(x_k|y_{0:k}) = \sum_{j=1}^{M} \alpha_{j,k} p_j(x_k|y_{0:k}) \tag{4.73}$$

avec $\sum_{j=1}^{M} \alpha_{j,k} = 1$. On a

$$p(x_k|y_{0:k}) = \frac{p(y_k|x_k)p(x_k|y_{0:k-1})}{\int p(y_k|x_k)p(x_k|y_{0:k-1})\,\mathrm{d}x_k} \tag{4.74}$$

Supposons les coefficients $\alpha_{j,k-1}$ connus. La densité prédite est :

$$p_{k|k-1} = \sum_{j=1}^{M} \alpha_{j,k-1} p_j(x_k|y_{0:k-1}) \tag{4.75}$$

avec $p_j(x_k|y_{0:k-1}) = \int p(x_k|x_{k-1}) p_j(x_{k-1}|y_{0:k-1}) \mathrm{d}x_{k-1}$ Introduisons M densités d'importance $\tilde{q}_j(x_k)$, $j = 1, \ldots, M$ respectivement centrées sur le MAP de $p_j(x_k|y_{0:k})$ noté $\hat{x}_j^*$.

$$\begin{aligned}
p_{k|k} &= \frac{\sum_{j=1}^{M} \alpha_{j,k-1} p(y_k|x_k) p_j(x_k|y_{0:k-1})}{\sum_{l=1}^{M} \alpha_{l,k-1} \int p(y_k|x_k) p_l(x_k|y_{0:k-1})\,\mathrm{d}x_k} \\
&= \frac{\sum_{j=1}^{M} \alpha_{j,k-1} p(y_k|x_k) \frac{p_j(x_k|y_{0:k-1})}{\tilde{q}_j(x_k)} \tilde{q}_j(x_k)}{\sum_{l=1}^{M} \alpha_{l,k-1} \int p(y_k|x_k) \frac{p_l(x_k|y_{0:k-1})}{\tilde{q}_l(x_k)} \tilde{q}_l(x_k)\,\mathrm{d}x_k} \\
&= \sum_{j=1}^{M} \frac{\alpha_{j,k-1} \int p(y_k|x_k) \frac{p_j(x_k|y_{0:k-1})}{\tilde{q}_j(x_k)} \tilde{q}_j(x_k)\,\mathrm{d}x_k}{\sum_{l=1}^{M} \alpha_{l,k-1} \int p(y_k|x_k) \frac{p_l(x_k|y_{0:k-1})}{\tilde{q}_l(x_k)} \tilde{q}_l(x_k)\,\mathrm{d}x_k} \times \underbrace{\left(\frac{p(y_k|x_k) \frac{p_j(x_k|y_{0:k-1})}{\tilde{q}_j(x_k)} \tilde{q}_j(x_k)}{\int p(y_k|x_k) \frac{p_j(x_k|y_{0:k-1})}{\tilde{q}_j(x_k)} \tilde{q}_j(x_k)\,\mathrm{d}x_k} \right)}_{p_j(x_k|y_{0:k})}
\end{aligned} \tag{4.76}$$

Par identification des termes,

$$\alpha_{j,k} = \frac{\alpha_{j,k-1} \int p(y_k|x_k) \frac{p_j(x_k|y_{0:k-1})}{\tilde{q}_j(x_k)} \tilde{q}_j(x_k)\,\mathrm{d}x_k}{\sum_{l=1}^{M} \alpha_{l,k-1} \int p(y_k|x_k) \frac{p_l(x_k|y_{0:k-1})}{\tilde{q}_l(x_k)} \tilde{q}_l(x_k)\,\mathrm{d}x_k} \tag{4.77}$$

Ceci suggère l'algorithme suivant,

1. On suppose que l'on dispose de l'approximation particulaire suivante à l'instant $k-1$:

$$p_{k-1|k-1} = \sum_{j=1}^{M} \alpha_{j,k-1} \sum_{i \in I_{j,k-1}} \omega_{k-1}^{i} \delta_{x_{k-1}^{i}} \tag{4.78}$$

2. Prédiction : tirer N échantillons $x_k^i \sim p(x_k|x_{k-1}^i)$
3. Pour $j = 1, \ldots, M$, former une approximation gaussienne $\varphi(x_k|\bar{x}_{j,k|k-1}, \hat{P}_{j,k|k-1})$ de la densité prédite $p_j(x_k|y_{0:k-1})$: calculer la moyenne prédite selon

$$\bar{x}_{j,k|k-1} = \sum_{i \in I_{j,k-1}} \omega_{k-1}^{i} x_k^i \tag{4.79}$$

et la covariance prédite selon

$$\hat{P}_{j,k|k-1} = \sum_{i \in I_{j,k-1}} \omega_{k-1}^{i} (x_k^i - \bar{x}_{j,k|k-1})(x_k^i - \bar{x}_{j,k|k-1})^T \tag{4.80}$$

4. Pour $j = 1, \ldots, M$, calculer le MAP $\hat{x}_j^*$ de $p_j(x_k|y_{0:k}) \propto p(y_k|x_k) p_j(x_k|y_{0:k-1})$.
Pour cela, utiliser l'approximation gaussienne $q_j = \varphi(x_k|\bar{x}_{j,k|k-1}, \hat{P}_{j,k|k-1})$ de $p_j(x_k|y_{0:k-1})$.
Utiliser ensuite la méthode exposée dans 4.5.2 : q_j joue ici le rôle de la loi a priori.
5. Définir pour $j = 1, \ldots, M$ une densité d'importance gaussienne $\tilde{q}_j$ de moyenne $\hat{x}_j^*$ et de covariance Σ_j^* égale à P_{rot}.

$$\tilde{q}_j(x) = \varphi(x|\hat{x}_j^*, \Sigma_j^*) \tag{4.81}$$

Un autre choix consiste à définir $\tilde{q}_j$ comme la densité de la loi de Student $T_d(\nu, \hat{x}_j^*, \frac{\nu-2}{\nu} J_j^{-1}(\hat{x}_j^*))$ où

$$J_j(x) = -\frac{\partial^2 \log g_k}{\partial x^2}(x) - \frac{\partial^2 \log q_j}{\partial x^2}(x)$$

Le degré de liberté est alors choisi expérimentalement.
Soit $N_j = \mathrm{Card}(I_{j,k-1})$. Tirer N_j échantillons i.i.d. $x_k^i \sim \tilde{q}_j$ avec $i \in I_{j,k-1}$.
6. Calculer les poids intra-composante ω_k^i. Pour $j = 1, \ldots, M$:
Pour $i \in I_{j,k-1}$ calculer

$$\tilde{\omega}_k^i = \frac{p(y_k|x_k^i) q_j(x_k^i)}{\tilde{q}_j(x_k^i)} \tag{4.82}$$

et poser

$$\omega_k^i = \frac{\tilde{\omega}_k^i}{\sum_{i \in I_{j,k-1}} \tilde{\omega}_k^i} \tag{4.83}$$

7. Calculer les coefficients de mélange $\alpha_{j,k}$ selon

$$\alpha_{j,k} \approx \frac{\frac{\alpha_{j,k-1}}{N_j} \sum_{i \in I_{j,k-1}} \tilde{\omega}_k^i}{\sum_{l=1}^{M} \frac{\alpha_{l,k-1}}{N_l} \sum_{i \in I_{l,k-1}} \tilde{\omega}_k^i} \tag{4.84}$$

L'approximation (4.84) peut se déduire par substitution des échantillons i.i.d. $x_k^i \sim \tilde{q}_j$, $i \in I_{j,k-1}$ dans (4.77) : on a alors

$$\int p(y_k|x_k)\frac{p_j(x_k|y_{0:k-1})}{\tilde{q}_j(x_k)}\tilde{q}_j(x_k)\,\mathrm{d}x_k \approx \frac{1}{N_j}\sum_{i\in I_{j,k-1}}\frac{p(y_k|x_k^i)p_j(x_k^i|y_{0:k-1})}{\tilde{q}_j(x_k^i)} \tag{4.85}$$

$$\approx \frac{1}{N_j}\sum_{i\in I_{j,k-1}}\frac{p(y_k|x_k^i)q_j(x_k^i)}{\tilde{q}_j(x_k^i)} \tag{4.86}$$

$$= \frac{1}{N_j}\sum_{i\in I_{j,k-1}}\tilde{\omega}_k^i \tag{4.87}$$

Echantillonnage d'importance adaptatif pour le MRPF Pour diminuer le surcoût algorithmique induit par la recherche des maxima locaux, il est judicieux d'utiliser un critère permettant de quantifier l'incohérence entre la fonction de vraisemblance $g_k(x_k) = p(y_k|x_k)$ et la densité prédite $p_{k|k-1}(x_k) = p(x_k|y_{0:k-1})$, de manière à recourir aux fonctions d'importances $\tilde{q}_j$ uniquement lorsque le critère indique que le recouvrement entre les deux densités est faible. Dans le cas contraire, le déroulement des étapes de prédiction et correction se fait comme exposé dans le chapitre 2, sous-section 2.6.2 et dans la section 4.4.
Un indicateur de cohérence entre g_k et $p_{k|k-1}$ est donné par

$$\delta_k = \frac{\sup_{x\in\mathbb{R}^d} g_k(x)}{\int g_k(x_k)p_{k|k-1}(x_k)dx_k} \tag{4.88}$$

$\delta_k \geq 1$ et δ_k est d'autant plus grand que g_k et $p_{k|k-1}$ ont un faible recouvrement [37, p. 256]. On peut évaluer δ_k pour un filtre particulaire donné en utilisant les particules et les poids associées. En effet, soit

$$\hat{p}_{k|k-1} = \sum_{k=1}^{N}\omega_{k-1}^i\delta_{x_k^i} \tag{4.89}$$

une approximation de la densité prédite $p_{k|k-1}$, avec $\sum_{k=1}^N \omega_{k-1}^i = 1$. Alors,

$$\delta_k \approx N\frac{\sup_{x\in\mathbb{R}^d} g_k(x)}{\sum_{k=1}^N \omega_{k-1}^i g_k(x_k^i)} \tag{4.90}$$

Dans le cas particulier où le bruit v est additif gaussien, (4.90) se réduit à

$$\delta_k \approx \frac{N}{\sum_{k=1}^{N}\omega_{k-1}^i \exp(-\frac{1}{2}\|y_k - h_k(x_k)\|^2_{R^{-1}})} \tag{4.91}$$

où $\|.\|^2_W$ est définie par $\|x\|^2_W = x^T W x$ lorsque W est une matrice $d \times d$.

On peut alors exploiter l'indicateur δ_k, pour contrôler les instants où l'échantillonnage d'importance basé sur les densités $\tilde{q}_1, \ldots, \tilde{q}_M$ est utilisé. Il suffit de définir un seuil $\delta_{max} > 1$ qui

contrôle le degré d'incohérence tolérable.

Souvenons-nous que dans le MPF et le MRPF, le passage de la densité prédite $p_j(x_k|y_{0:k-1})$ de la j-ème composante à la densité corrigée $p_j(x_k|y_{0:k})$ peut se faire indépendamment des autres composantes (cf. 2.6.2). Seul le calcul des coefficients de mélange $\{\alpha_{l,k}\}_{l=1}^M$ dépend de l'ensemble des poids des particules.
On peut alors définir pour chaque composante d'indice j, un indicateur de cohérence entre g_k et $p_j(x_k|y_{0:k})$, que nous noterons $\delta_{j,k}$. Si $\delta_{j,k} \geq \delta_{max}$, l'étape de correction pour la j-ème composante se fait selon les points 3, 4, 5 et 6. Dans le cas contraire, l'étape de correction se réduit au calcul des poids intra-composante ω_k^i comme décrit dans le point 2.(b)i. de l'algorithme MRPF.
Dans les deux cas, pour obtenir la densité a posteriori globale $p(x_k|y_{0:k})$, il faut calculer les coefficients $\{\alpha_{l,k}\}_{l=1}^M$ en fonction de $\{\alpha_{l,k-1}\}_{l=1}^M$. L'expression correspondante est :

$$\alpha_{j,k} = \frac{\alpha_{j,k-1}\beta_{j,k}}{\sum_{l=1}^M \alpha_{l,k-1}\beta_{l,k}} \tag{4.92}$$

avec

$$\beta_{j,k} = \begin{cases} \displaystyle\sum_{i \in I_{j,k-1}} \omega_{k-1}^i p(y_k|x_k^i) & \text{si } \delta_{j,k} < \delta_{max} \\ \frac{1}{N_j} \displaystyle\sum_{i \in I_{j,k-1}} \frac{p(y_k|x_k^i) q_j(x_k^i)}{\tilde{q}_j(x_k^i)} & \text{sinon} \end{cases} \tag{4.93}$$

Alternativement, on peut utiliser la taille effective de l'échantillon (4.33) comme indicateur d'incohérence entre la fonction de vraisemblance et la densité prédite. Le choix d'utiliser la densité d'importance $\tilde{q}_j$ s'opère lorsque la taille effective de la loi empirique $\hat{p}_j = \sum_{i \in I_{j,k}} \omega_k^i \delta_{x_k^i}$, notée N_{eff}^j, descend en dessous d'un seuil $N_{th,MAP}$ inférieur ou égal au seuil de ré-échantillonnage. Nous proposons de définir ce seuil comme une fraction $\zeta \in [0,1]$ du seuil de ré-échantillonnage :

$$N_{th,MAP} = \zeta N_{th} \tag{4.94}$$

L'idée est que, le ré-échantillonnage et la régularisation peuvent suffire à ramener les particules dans la zone d'intéret lorsque N_{eff}^j est juste légèrement en dessous du seuil de ré-échantillonnage. Le coefficient $\zeta < 1$ permet alors d'ajuster le degré de dégénérescence toléré : des valeurs petites de $\zeta < 1$ signifient que l'échantillonnage selon $\tilde{q}_j$ est effectué uniquement lorsque les particules appartenant à la composante d'indice j sont très dégénérées.
En résumé, si $N_{eff}^j < \frac{N_{th,MAP}}{N} N_j$, l'échantillonnage se fait selon $\tilde{q}_j$, et dans le cas contraire selon le noyau de transition.
Nous résumons ci-après l'algorithme fusionnant le MRPF et l'utilisation des modes locaux pour la phase de d'échantillonnage d'importance, que nous appellerons MRPF-MAP.

Dans la partie consacrée à l'évaluation des performances, nous distinguerons deux variantes du MRPF-MAP :

- le MRPF-MAP 1 : la densité instrumentale a pour covariance la covariance prédite,
- le MRPF-MAP 2 : la densité instrumentale a pour covariance la covariance P_{rot} définie par l'équation (4.64).

Algorithme 7 : MRPF-MAP

1. *[Initialisation]*
   - Tirer N particules selon la loi initiale $x_0^i \sim \mu_0$
   - poser $\omega_0^i \propto p(y_0|x_0^i)$
   - Normaliser les ω_0^i
2. Pour $k \geq 1$
   (a) *[Prédiction]* Tirer N particules $x_k^i \sim p(x_k|x_{k-1}^i)$
   (b) *[Correction]*
      i. Pour $j = 1, \ldots, M_{k-1}$
         - calculer N_{eff}^j selon (4.33)
         - Si $N_{eff}^j/N_j \leq N_{th,MAP}/N$
            - Calculer $\bar{x}_{j,k|k-1}$ et $\hat{P}_{j,k|k-1}$ selon (4.80) et (4.79). Poser $q_j = \varphi(.|\bar{x}_{j,k|k-1}, \hat{P}_{j,k|k-1})$.
            - Calculer $\hat{x}_j^* = \arg\max_x p(y_k|x)q_j(x)$ selon la méthode de la section 4.5.2.
            - Poser $\tilde{q}_j(x) = \varphi(x|\hat{x}_j^*, \Sigma_j^*)$ avec $\Sigma_j^* = \hat{P}_{j,k|k-1}$ ou $\Sigma_j^* = P_{rot}$
            - Tirer N_j échantillons i.i.d. $x_k^i \sim \tilde{q}_j$ avec $i \in I_{j,k-1}$ et $N_j = \mathrm{Card}(I_{j,k-1})$
            - Calculer $\tilde{\omega}_k^i$ et ω_k^i selon (4.82) et (4.83)
            - Poser $\beta_{j,k} = \frac{1}{N_j}\sum_{i\in I_{j,k-1}} \tilde{\omega}_k^i$
         - Sinon
            - Calculer les poids non normalisés $\tilde{\omega}_k^i$ et les poids normalisés ω_k^i de la j-ème composante selon (4.37) et (4.38)
            - Poser $\beta_{j,k} = \sum_{i\in I_{j,k-1}} \tilde{\omega}_k^i$
      ii. Pour $j = 1, \ldots, M_{k-1}$, calculer $\alpha_{j,k}$ en fonction des $\{\alpha_{j,k-1}\}_{1\leq j\leq M_{k-1}}$ et des $\{\beta_{j,k-1}\}_{1\leq j\leq M_{k-1}}$ en utilisant l'équation (4.92)
   (c) *[Clustering]* Décomposer le nuage de particules en M_k clusters $C_{j,k}$. $[C'_{j,k}, I'_{j,k}] = \texttt{meanshiftClustering}(\{x_k^i\}_{i=1}^N, \{\omega_k^i\}_{i=1}^N, K, h^{MS}, R_m, n_{itermax}, \varepsilon^{MS})$
   (d) Mettre à jour les coefficients de mélange $\alpha_{j,k}$ et les poids des particules intra-cluster ω_k^i pour la nouvelle partition $C'_{j,k}$ en utilisant les équations (2.84) et (2.85).
3. Poser $j = 1$. Tant que $j \leq M_k$
   - Si $\alpha_{j,k} < \alpha_{min}$ supprimer le cluster $C_{j,k}$ et échantillonner N_j nouvelles particules selon la procédure décrite dans la section 4.3
   - Sinon
      - Calculer N_{eff}^j selon (4.33)
      - Si $\frac{N_{eff}^j}{N_j}$ est inférieur au seuil de ré-échantillonnage, ré-échantillonner et régulariser les particules du cluster $C_{j,k}$ selon l'algorithme décrit dans la section 4.2
      - $j = j + 1$
4. Mettre à jour la fenêtre de lissage h^{MS} de l'algorithme de clustering mean-shift

4.5.4 Extension au filtre particulaire Rao-Blackwellisé : algorithme MRBPF-MAP

L'algorithme précédent s'étend naturellement au RBPF (cf. section 2.4.1). L'avantage du RBPF par rapport au RPF est que l'échantillonnage des particules est restreint à la partie non-linéaire du vecteur d'état (composantes de la position horizontale dans le cas du recalage altimétrique). De plus, nous avons observé dans les simulations que les ambiguïtés de terrain sont résolues plus rapidement par le RBPF que par le RPF : cela permet en principe d'appliquer la recherche des modes locaux du RBPF pour un coût de calcul moindre puisque le nombre de modes dans le RBPF est en pratique plus faible que celui du RPF.
Rappellons que le MRBPF est l'extension du MRPF où le RPF est remplacé par le RBPF en conservant l'étape de régularisation lors du ré-échantillonnage. De même, le MRBPF-MAP est l'extension du MRPF-MAP. Deux adaptations sont nécessaires afin d'obtenir cet algorithme :

- le calcul de la matrice de covariance de la densité prédite correspondant à chaque cluster se fait en considérant la décomposition du vecteur d'état en la partie linéaire et non-linéaire
- la correction de Kalman au niveau d'un cluster donné est effectuée différemment selon que l'échantillonnage d'importance fondée sur les modes locaux est appliquée ou non.

Algorithme 8 : MRBPF-MAP

---

si $k = 1$

- tirer $x_k^{n,i} \sim p(x_0^n)$, $i = 1, \ldots, N$
- poser $x_{0|0}^{l,i} = \bar{x}_0^l$ et $P_{0|0}^l = \bar{P}_0$

fin si

si $k = 2, \ldots, n$,

- *[Prédiction particulaire]* Générer un échantillon $x_{k|k-1}^{n,i} \sim p(x_k^n | x_{0:k-1}^n, y_{0:k-1})$ (cf. (2.39)), $i = 1, \ldots, N$. Poser $\omega_{k|k-1}^i = \omega_{k-1}^i$, $i = 1, \ldots, N$.
- *[Prédiction de Kalman]* Pour $i = 1, \ldots, N$, calculer $\hat{x}_{k|k-1}^{l,i}$ et $P_{k|k-1}^l$ d'après (2.34).
- *[Clustering]* Décomposer le nuage de particules en M_k clusters $C_{j,k}$.
  $[C'_{j,k}, I'_{j,k}] = \texttt{meanshiftClustering}(\{x_k^i\}_{i=1}^N, \{\omega_k^i\}_{i=1}^N, K, h^{MS}, R_m, n_{itermax}, \varepsilon^{MS})$
- *[Correction particulaire]* Calculer les poids non-normalisés $\tilde{\omega}_k^i = \omega_{k-1}^i p(y_k | x_k^{n,i}, y_{0:k-1})$, $i = 1, \ldots, N$ où la vraisemblance $p(y_k | x_k^{n,i}, y_{0:k-1})$ est donnée d'après (2.40).
- *[Echantillonnage d'importance selon critère de dégénérescence]* Pour $j = 1, \ldots, M_k$
  - $\beta_j = \sum_{i \in I_{j,k}} \tilde{\omega}_k^i$
  - $\omega_k^i = \frac{\tilde{\omega}_k^i}{\beta_j}$, $i \in I_{j,k}$
  - Calculer N_{eff}^j selon (4.33)
  - $\text{indexMAP}_j = 0$
  - Si $N_{eff}^j / N_j \leq N_{th,MAP} / N$
    - Poser $\bar{x}_{k|k-1,j}^n = \sum_{i \in I_{j,k}} \omega_{k|k-1}^i x_{k|k-1}^{n,i}$
    - Poser $\bar{x}_{k|k-1,j}^l = \sum_{i \in I_{j,k}} \omega_{k|k-1}^i x_{k|k-1}^{l,i}$
    - Poser $\hat{P}_{j,k|k-1}^n = \sum_{i \in I_{j,k}} \omega_{k|k-1}^i (x_{k|k-1}^{n,i} - \bar{x}_{k|k-1,j}^n)(x_{k|k-1}^{n,i} - \bar{x}_{k|k-1,j}^n)^T$

- Poser $\hat{P}^l_{j,k|k-1} = P^l_{k|k-1} + \sum_{i\in I_{j,k}} \omega^i_{k|k-1}(x^{l,i}_{k|k-1} - \bar{x}^l_{k|k-1,j})(x^{l,i}_{k|k-1} - \bar{x}^l_{k|k-1,j})^T$
- Poser $\hat{P}^{nl}_{j,k|k-1} = \sum_{i\in I_{j,k}} \omega^i_{k|k-1}(x^{n,i}_{k|k-1} - \bar{x}^n_{k|k-1,j})(x^{l,i}_{k|k-1} - \bar{x}^l_{k|k-1,j})^T$
- Poser

$$\hat{P}_{j,k|k-1} = \begin{pmatrix} \hat{P}^n_{j,k|k-1} & \hat{P}^{nl}_{j,k|k-1} \\ \left(\hat{P}^{nl}_{j,k|k-1}\right)^T & \hat{P}^l_{j,k|k-1} \end{pmatrix}$$

et $\bar{x}_{j,k|k-1} = \left[\left(\bar{x}^n_{k|k-1,j}\right)^T \quad \left(\bar{x}^l_{k|k-1,j}\right)^T\right]^T$.

- Poser $q_j = \varphi(.|\bar{x}_{j,k|k-1}, \hat{P}_{j,k|k-1})$.
- Calculer $\hat{x}^*_j = \arg\max_x p(y_k|x)q_j(x)$ selon la méthode de la section 4.5.2.
- Poser $\tilde{q}_j(x) = \varphi(x|\hat{x}^*_j, \Sigma^*_j)$ avec $\Sigma^*_j = \hat{P}_{j,k|k-1}$ ou $\Sigma^*_j = P_{rot}$
- Tirer N_j échantillons i.i.d. $\xi^i \sim \tilde{q}_j$ avec $i \in I_{j,k-1}$ et $N_j = \mathrm{Card}(I_{j,k-1})$
- Poser $x^{n,i}_k = \xi^{n,i}$ et $x^{l,i}_k = \xi^{l,i}$ avec $\xi^i = \begin{bmatrix} \xi^{n,i} \\ \xi^{l,i} \end{bmatrix}$.
- Calculer $\tilde{\omega}^i_k$ et ω^i_k selon (4.82) et (4.83)
- Poser $\beta_{j,k} = \frac{1}{N_j}\sum_{i\in I_{j,k-1}} \tilde{\omega}^i_k$
- $\mathrm{indexMAP}_j = 1$

- Pour $j = 1, \ldots, M_k$, calculer $\alpha_{j,k}$ en fonction des $\{\alpha_{j,k-1}\}_{1\leq j\leq M_{k-1}}$ et des $\{\beta_{j,k-1}\}_{1\leq j\leq M_{k-1}}$ en utilisant l'équation (4.92)
- Calculer la moyenne pondérée partie non-linéaire $\hat{x}^n_k = \sum_{j=1}^{M_k}\sum_{i\in I_{j,k}} \omega^i_k x^{n,i}_k$.
- *[Ré-échantillonnage]* Poser $j = 1$. Tant que $j \leq M_k$
  - Si $\alpha_{j,k} < \alpha_{min}$ supprimer le cluster $C_{j,k}$ et échantillonner N_j nouvelles particules selon la procédure décrite dans la section 4.3
  - Sinon
    - Calculer N^j_{eff} selon (4.33)
    - Si $\frac{N^j_{eff}}{N_j}$ est inférieur au seuil de ré-échantillonnage, ré-échantillonner et régulariser les particules du cluster $C_{j,k}$ selon l'algorithme décrit dans la section 4.2
    - $j = j + 1$
- Mettre à jour la fenêtre de lissage h^{MS} de l'algorithme de clustering mean-shift
- *[Correction de Kalman]*
- Calculer P^l_k selon (2.36).
  - Pour $j = 1, \ldots, M_k$
    - si $\mathrm{indexMAP}_j = 0$, calculer $\hat{x}^{l,i}_k$, $i \in I_{j,k-1}$ d'après (2.35)
- calculer la moyenne a posteriori de la partie linéaire $\hat{x}^l_k = \sum_{j=1}^{M_k}\sum_{i\in I_{j,k}} \omega^i_k x^{l,i}_k$

fin pour

Les performances de cet algorithme seront illustrées dans la section suivante et comparées avec le RBPF standard.

4.6 Evaluation des performances du MRPF, du MRPF-MAP et du MRBPF-MAP

Nous présentons dans cette section les performances des filtres développés sur la base de trois simulations :

- Simulations A : comparaison du MRPF avec le RPF et le RBPF dans un cas nominal d'intensité de bruit de mesure,
- Simulations B : comparaison du MRPF avec le MRPF-MAP dans un cas où la variance du bruit de mesure est petite,
- Simulations C : comparaison du MRBPF-MAP avec le MRBPF et le RBPF dans les mêmes conditions que les simulations B.

Pour chaque simulation, les résultats numériques issus des tirages Monte Carlo sont rapportés, ceci pour trois types de terrain.
Dans un premier temps, nous introduisons les scénarios de référence (terrain et trajectoire associée) qui serviront de base aux simulations. Ensuite, nous rappelons le modèle d'état et d'observation. Enfin, nous présentons les performances en termes d'erreur quadratique moyenne et de taux de convergence.

4.6.1 Profils de terrain pour le recalage radio-altimétrique

Afin d'évaluer les performances des algorithmes de recalage, nous nous appuyerons sur trois scénarios de navigation qui diffèrent essentiellement par le niveau d'ambigüité et/ou de dureté du profil de terrain. La dureté d'un terrain est une caractéristique locale du relief utilisée pour évaluer son degré de rugosité. Plus celle ci est grande, plus l'information contenue dans le terrain est importante. Il existe diverses définitions du facteur f de dureté, selon les industriels, la suivante est couramment utilisée en navigation par corrélation de terrain :

$$f = \sqrt{\frac{\sigma_T \sigma_p}{2\pi}} \tag{4.95}$$

où

- σ_T désigne l'écart-type de l'altitude locale
- σ_p désigne l'écart-type de la pente locale

Une valeur de f faible suggère que le terrain est localement plat tandis qu'une valeur de f élevée indique que le terrain est localement accidenté. On peut montrer que $f \propto \frac{\sigma_T}{X_T}$, où X_T est la longueur de corrélation du terrain et le rapport $\frac{\sigma_T}{X_T}$ caractérise la rugosité de terrain d'après [27]. En pratique, la dureté f en un point de la carte est évaluée en moyennant la dureté nord f_n et la dureté est f_e. Afin de calculer f_n et f_e, les termes σ_T et σ_p sont calculées sur une croix nord-est de 5 km centrée sur le point d'intérêt.

Type de terrain	plat	vallonné	accidenté	très accidenté	extrêmement accidenté
f	0 à 0.35	0.35 à 0.75	0.75 à 2	2 à 3.6	> 3.6

L'ambiguïté de terrain désigne la présence de portions de terrain dont les profils en altitude sont semblables. En terme de filtrage, plus un terrain est ambigu, plus l'intervalle de temps pendant lequel le filtre est multimodal est grand.

Lors de l'évaluation et la comparaison des différents filtres de navigations, nous considérons des trajectoires rectilignes. Le radio-altimètre fournit une mesure à la fréquence de $10\,Hz$ Les scénarios de navigation que nous considérons ont les caractéristiques suivantes :

- **Scénario 1** : Terrain peu ambigu, présentant une dureté moyenne.
- **Scénario 2** : Terrain ambigu en début de recalage, avec une dureté élevée.
- **Scénario 3** : Terrain très ambigu sur la première moitié de la trajectoire, moyennement ambigu sur la deuxième moitié de la trajectoire. La dureté le long de la trajectoire considérée est faible ($f < 0.35$).

4.6.2 Modèle d'état et d'observation

Nous souhaitons estimer à chaque instant $k \geq 0$, le vecteur d'erreur inertielle

$$x_k = \begin{bmatrix} x_{k,1} & x_{k,2} & x_{k,3} & x_{k,4} & x_{k,5} & x_{k,6} \end{bmatrix}^T$$

défini par

$$x_k = \begin{bmatrix} \delta \boldsymbol{x} \\ \delta \boldsymbol{v} \end{bmatrix}_k \tag{4.96}$$

où

$$\delta \boldsymbol{x}_k = \begin{bmatrix} \delta x_n \\ \delta x_e \\ \delta x_d \end{bmatrix}_k \tag{4.97}$$

désigne le vecteur d'erreur en position horizontale dans le repère TGL $(\boldsymbol{n}, \boldsymbol{e}, \boldsymbol{d})$ défini au chapitre 3, section 3.1.1. De même,

$$\delta \boldsymbol{v}_k = \begin{bmatrix} \delta v_n \\ \delta v_e \\ \delta v_d \end{bmatrix}_k \tag{4.98}$$

est le vecteur d'erreur en vitesse horizontale dans ce même repère.

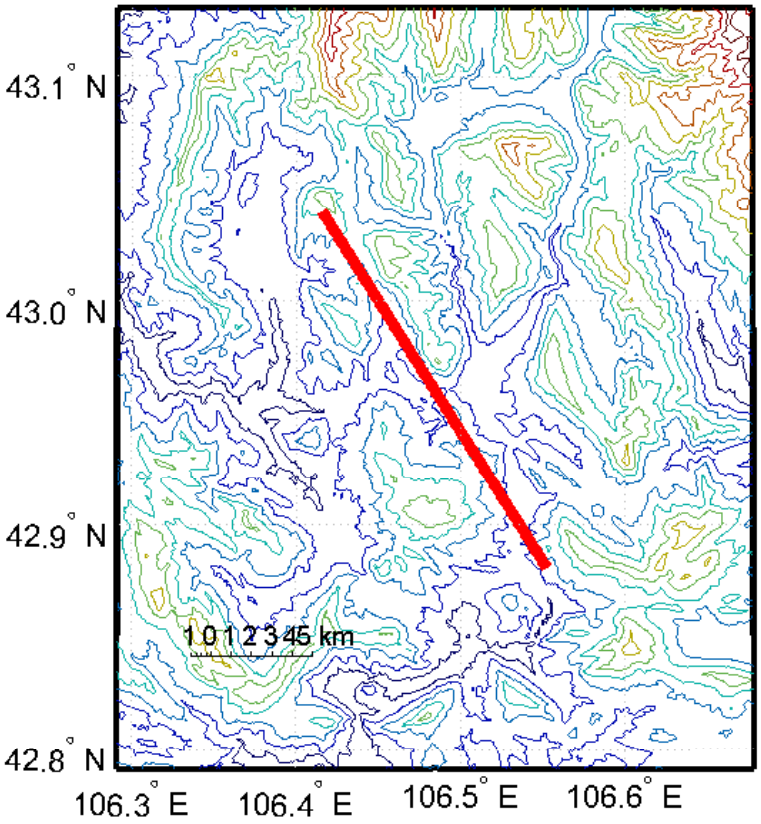
La dynamique que nous considérons est un modèle double intégrateur défini par

$$x_{k+1} = F x_k + G_k w_k \tag{4.99}$$

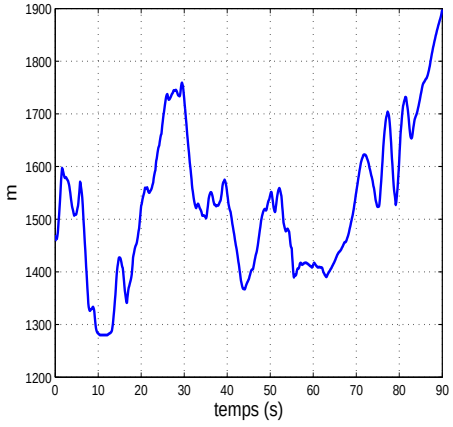
avec

$$F = \begin{pmatrix} I_3 & \Delta I_3 \\ 0_3 & I_3 \end{pmatrix} \qquad G_k = \begin{pmatrix} \frac{\Delta^2}{2} I_3 \\ \Delta I_3 \end{pmatrix} \tag{4.100}$$

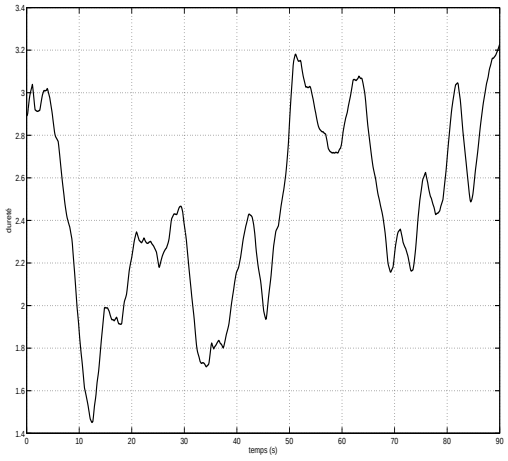
- I_3 est la matrice identité de taille
- 0_3 est la matrice nulle de taille 3×3
- Δ est la période d'échantillonnage.
- w_k est un bruit blanc gaussien de covariance Q



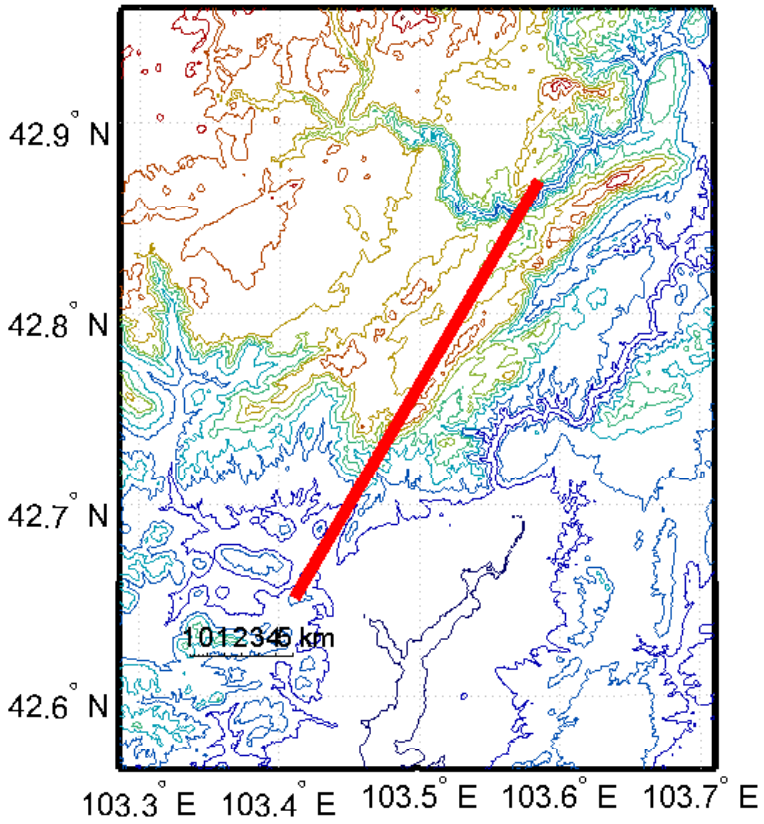
(a) Profil de terrain et trajectoire (trait rouge)



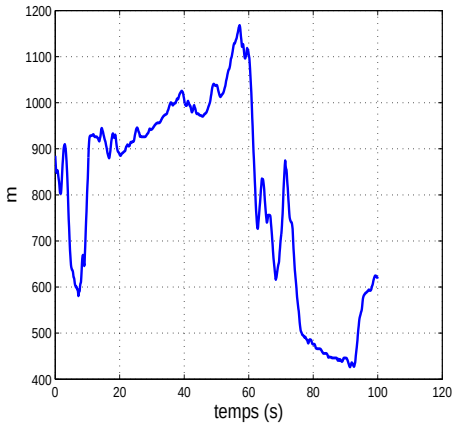
(b) Hauteur de terrain le long de la trajectoire



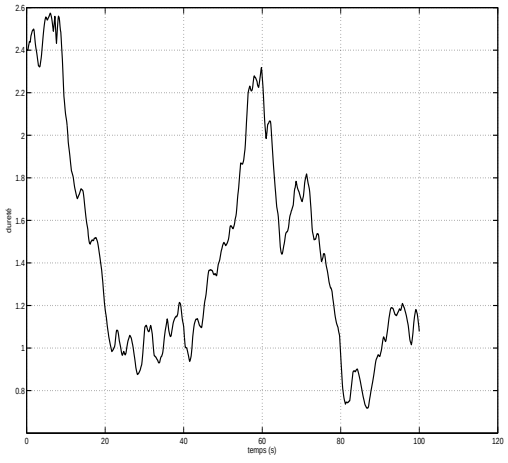
(c) Dureté du terrain le long de la trajectoire



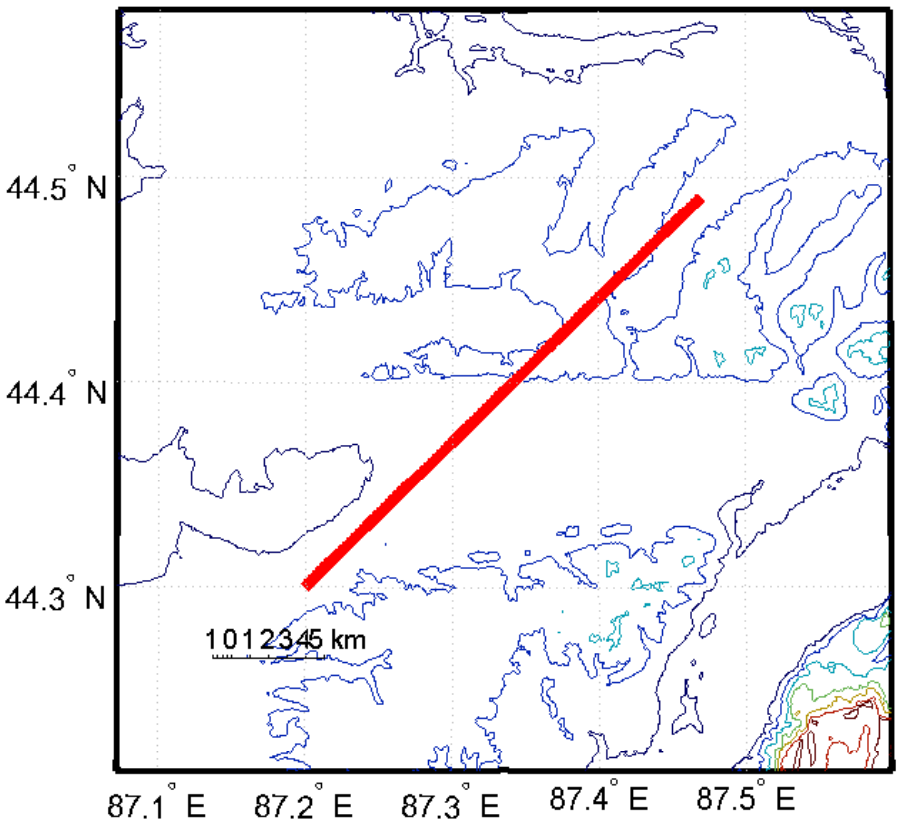
(a) Profil de terrain et trajectoire (trait rouge)



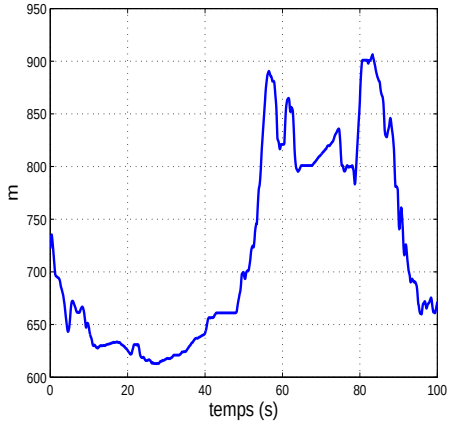
(b) Hauteur de terrain le long de la trajectoire



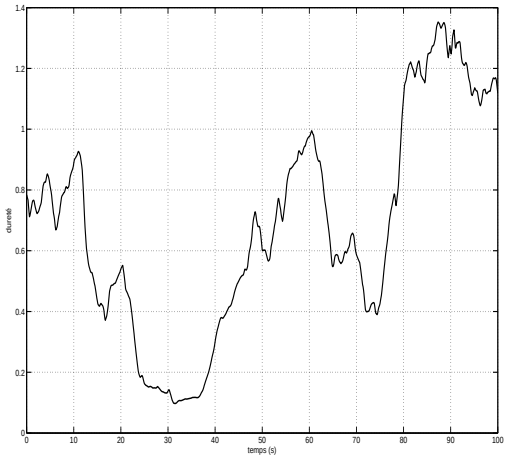
(c) Dureté du terrain le long de la trajectoire



(a) Profil de terrain et trajectoire (trait rouge)



(b) Hauteur de terrain le long de la trajectoire



(c) Dureté du terrain sous la trajectoire

Notons que dans cette étude, nous n'estimons pas les angles d'attitude étant donné que les trajectoires considérées sont assez courtes et ne permettent pas aux erreurs d'estimation d'attitude de converger suffisamment.

Le modèle d'observation que nous considérons est celui d'une mesure altimétrique correspondant à la hauteur-sol bruitée par un bruit blanc additif gaussien v de variance σ_v^2. Soit $\tilde{\lambda}$, $\tilde{\phi}$ et $\tilde{z}$, la latitude, longitude et altitude inertielle de l'aéronef. On introduit également le modèle numérique de terrain h_{MNT} qui à une latitude λ et une longitude ϕ, fait correspondre la hauteur du relief $h_{MNT}(\lambda, \phi)$. La mesure altimétrique y correspond à la différence entre l'altitude z du mobile et la hauteur du MNT correspond à la position horizontale (λ, ϕ) du mobile, à laquelle vient s'ajouter un bruit d'observation $v \sim \mathcal{N}(0, \sigma_v^2)$.

$$y = z - h_{MNT}(\lambda, \phi) + v \tag{4.101}$$

Compte tenu de l'orientation de l'axe $\boldsymbol{d}$ du TGL, l'altitude réelle z du mobile est reliée à l'altitude inertielle $\tilde{z}$ selon

$$z = \tilde{z} - \delta x_d$$

De même, la latitude réelle λ et la longitude réelle ϕ de l'aéronef s'obtiennent en fonction de leurs équivalents inertiels d'après les expressions suivantes :

$$\begin{aligned} \lambda =& \tilde{\lambda} + \frac{\delta x_n}{R_{\tilde{\lambda}} + \tilde{z}} \\ \phi =& \tilde{\phi} + \frac{\delta x_e}{\left(R_{\tilde{\phi}} + \tilde{z}\right) \cos \tilde{\phi}} \end{aligned} \tag{4.102}$$

Finalement, l'équation d'observation est

$$y_k = \tilde{z}_k - x_{k,3} - h_{MNT}\left(\tilde{\lambda} + \frac{x_{k,1}}{R_{\tilde{\lambda}} + \tilde{z}_k}, \tilde{\phi}_k + \frac{x_{k,2}}{\left(R_{\tilde{\phi}} + \tilde{z}_k\right) \cos \tilde{\phi}_k}\right) + v_k \tag{4.103}$$

où $R_{\tilde{\lambda}}$ et $R_{\tilde{\phi}}$ sont respectivement les rayons nord et est défini par les équations (3.9) et (3.10).

4.6.3 Critères de performances

Pour chaque scénario nous calculons l'erreur quadratique moyenne (MSE), le taux de non divergence ainsi que la vitesse de convergence des algorithmes.
Etant donné un estimateur $\hat{x}_k$ de x_k issu d'un algorithme particulaire, nous définissons momentanément ici l'erreur quadratique moyenne comme

$$MSE(\hat{x}_k) = \mathbb{E}_{p(x_k, y_{0:k})}\left[(\hat{x}_k - x_k)(\hat{x}_k - x_k)^T\right] \tag{4.104}$$

Cette définition diffère légèrement de celle donnée dans le chapitre 1 et nous permet d'estimer l'erreur quadratique moyenne commise sur chacun des axes *nord*, *est* et *down* en position et vitesse. Dans les trois algorithmes considérés l'estimateur $\hat{x}_k$ est la moyenne pondérée

$$\hat{x}_k = \sum_{i=1}^{N} \omega_k^i x_k^i$$

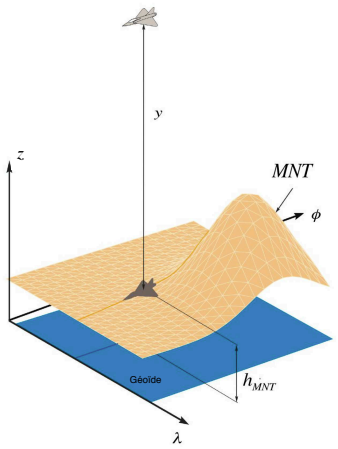


FIGURE 4.14 – Mesure altimétrique

La vraie densité jointe $p(x_k, y_{0:k})$ étant inconnue, nous approchons le MSE sur la base de $N_{MC} = 200$ réalisations du filtre :

$$MSE(\hat{x}_k) \approx \frac{1}{N_{MC}} \sum_{j=1}^{N_{MC}} (\hat{x}_k^j - x_k^j)(\hat{x}_k^j - x_k^j)^T \tag{4.105}$$

où les termes $\hat{x}_k^j$ et x_k^j représentent respectivement l'estimée du filtre et l'état vrai pour la j-ème réalisation du filtre. L'état vrai correspond à la vraie erreur inertielle simulée sur la bases des paramètres données dans la section 4.6.4.1.
Afin d'évaluer l'efficacité des algorithmes en terme d'erreur, le MSE est comparé à la borne de Cramer-Rao a posteriori (BCR) qui correspond à une borne inférieure sur l'erreur quadratique du filtre optimal. Pour les modèles à espace d'état de type (2.6), l'erreur quadratique moyenne d'un estimateur $\hat{x}_{0:k}$ de $x_{0:k}$ est minorée par l'inverse de la matrice d'information de Fisher du modèle i.e.

$$MSE(\hat{x}_{0:k}) = \mathbb{E}\left[(\hat{x}_{0:k} - x_{0:k})(\hat{x}_{0:k} - x_{0:k})^T\right] \geq J_{0:k}^{-1} \tag{4.106}$$

où la matrice d'information de Fisher $J_{0:k}$ est définie de la manière suivante :

$$J_{0:k} = -\mathbb{E}_{p(x_{0:k}, y_{0:k})}\left[\frac{\partial^2}{\partial x_{0:k}^2} \log p(x_{0:k}, y_{0:k})\right] \tag{4.107}$$

En ce qui concerne cette étude, nous souhaitons plutôt une borne sur $MSE(\hat{x}_k)$. En décomposant $J_{0:k}$ en blocs A_k, B_k et C_k de taille respectives $kd \times kd$, $kd \times d$ et $d \times d$

$$J_{0:k} = \begin{pmatrix} A_k & B_k \\ B_k^T & C_k \end{pmatrix} \tag{4.108}$$

on a [114]

$$MSE(\hat{x}_k) \geq J_k^{-1} \tag{4.109}$$

$J_k = C_k - B_k^T A_k B_k$ est la matrice d'information de Fisher associée à l'estimation de x_k.
En pratique, l'algorithme de Tichavsky et al. [114] qui permet de calculer J_k récursivement hors ligne.

Enfin, nous estimons pour chaque filtre le taux de non-divergence qui est calculé comme le rapport du nombre de réalisations non-divergentes sur le nombre total de réalisations du filtre. On considère qu'il y a divergence d'une réalisation d'un filtre lorsque l'erreur entre l'état x_k et l'estimée $\hat{x}_k$ est grande devant la covariance de l'estimation. Plus précisément, si n désigne l'instant final de la trajectoire, une divergence est déclarée lorsque x_n se trouve en dehors de l'ellipsoïde de confiance au seuil α associée à la distribution gaussienne centrée en $\hat{x}_n$ et de covariance $\hat{P}_n$ où $\hat{P}_n$ désigne la covariance empirique du filtre. Autrement dit, une réalisation sera considérée divergente si

$$(\hat{x}_n - x_n)^T \hat{P}_n^{-1} (\hat{x}_n - x_n) > \chi^2_{d,1-\alpha} \tag{4.110}$$

où $\chi^2_{2,1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la distribution du khi-deux à d degrés de libertés. Dans les simulations, nous avons pris la valeur $\alpha = 0.001$ pour le niveau de confiance.
Ce critère est motivé par l'utilisation d'une approximation gaussienne de la distribution empirique $\hat{p}_n = \sum_{i=1}^N \omega_n^i \delta_{x_n^i}$: en cas de non-divergence, l'état vrai x_k doit se trouver dans l'ellipsoïde de confiance associée à la distribution gaussienne dont les moments sont ceux de $\hat{p}_n$. Nous reviendrons sur la divergence du filtre particulaire au chapitre 5, dans le cadre de sa détection en ligne.

4.6.4 Simulations A : évaluation des performances du MRPF

Le MRPF est ici comparé à deux algorithmes particulaires classiques :
- le filtre particulaire régularisé (RPF) présenté au chapitre 2, section 2.3
- le filtre particulaire Rao-Blackwellisé (RBPF) exposé au chapitre 2, section 2.4.1

4.6.4.1 Conditions de simulations

Simulation de trajectoires inertielles Les trajectoires inertielles sont simulées en prenant les paramètres accélérométriques et gyrométriques suivants :
- écart-type du biais accélérométrique : $\sigma_a = 3.10^{-5}\, m.s^{-2}$
- période de corrélation du biais accélérométrique : $\tau_a = 60\, s$
- écart-type du biais gyrométrique : $\sigma_a = 10^{-6}\, rad.s^{-1}$
- période de corrélation du biais gyrométrique : $\tau_g = 60\, s$

L'erreur initiale en position, vitesse et attitude suit une distribution gaussienne centrée de matrice de covariance

$$P_0^{INS} = \mathrm{diag}(\sigma^2_{0,x_n},\, \sigma^2_{0,x_e},\, \sigma^2_{0,x_d},\, \sigma^2_{0,v_n},\, \sigma^2_{0,v_e},\, \sigma^2_{0,v_d},\, \sigma^2_{0,\varphi},\, \sigma^2_{0,\theta},\, \sigma^2_{0,\psi})$$

avec :

- $\sigma_{0,x_n} = 1000\,m$ (écart-type en position Nord)
- $\sigma_{0,x_e} = 1000\,m$ (écart-type en position Est)
- $\sigma_{0,x_d} = 100\,m$ (écart-type en position Verticale)
- $\sigma_{0,v_n} = 3\,m.s^{-1}$ (écart-type en vitesse Nord)
- $\sigma_{0,v_e} = 3\,m.s^{-1}$ (écart-type en vitesse Est)
- $\sigma_{0,v_d} = 1\,m.s^{-1}$ (écart-type en vitesse Verticale)
- $\sigma_{0,\varphi} = \sigma_{0,\theta} = \sigma_{0,\psi} = 0.05\frac{\pi}{180}^{\circ}$ (écarts-type en roulis, tangage et lacet)

Paramètres du modèle de dynamique et de mesure

- covariance du bruit de dynamique $Q = \mathrm{diag}\begin{pmatrix}1^2 & 1^2 & 0.01^2\end{pmatrix}$
- écart-type du bruit de mesure du radio-altimètre $\sigma_v = 15\,m$
- période d'échantillonnage $\Delta = 0.1\,s$

Paramètres propres aux filtres

RPF

▷ $N_{RPF} = 5000$ particules

▷ seuil de ré-échantilonnnage : $N_{th} = \frac{N_{RPF}}{2}$

MRPF

▷ $N_{MRPF} = 5000$ particules

▷ Paramètres de l'algorithme mean-shift clustering :

- rayon de fusion $R_m = 100\,m$
- nombre d'itérations maximum par procédure mean-shift : $n_{itermax} = 50$
- seuil de convergence $\varepsilon^{MS} = 1\,m$

▷ seuil de suppression de cluster $\alpha_{min} = 10^{-8}$

L'algorithme de clustering est appliqué toutes les 5 itérations afin de limiter le surcoût algorithmique

RBPF

▷ $N_{RBPF} = 5000$ particules

▷ seuil de ré-échantillonnage : $N_{th} = \frac{N_{RBPF}}{2}$

4.6.4.2 Résultats numériques

Scénario 1

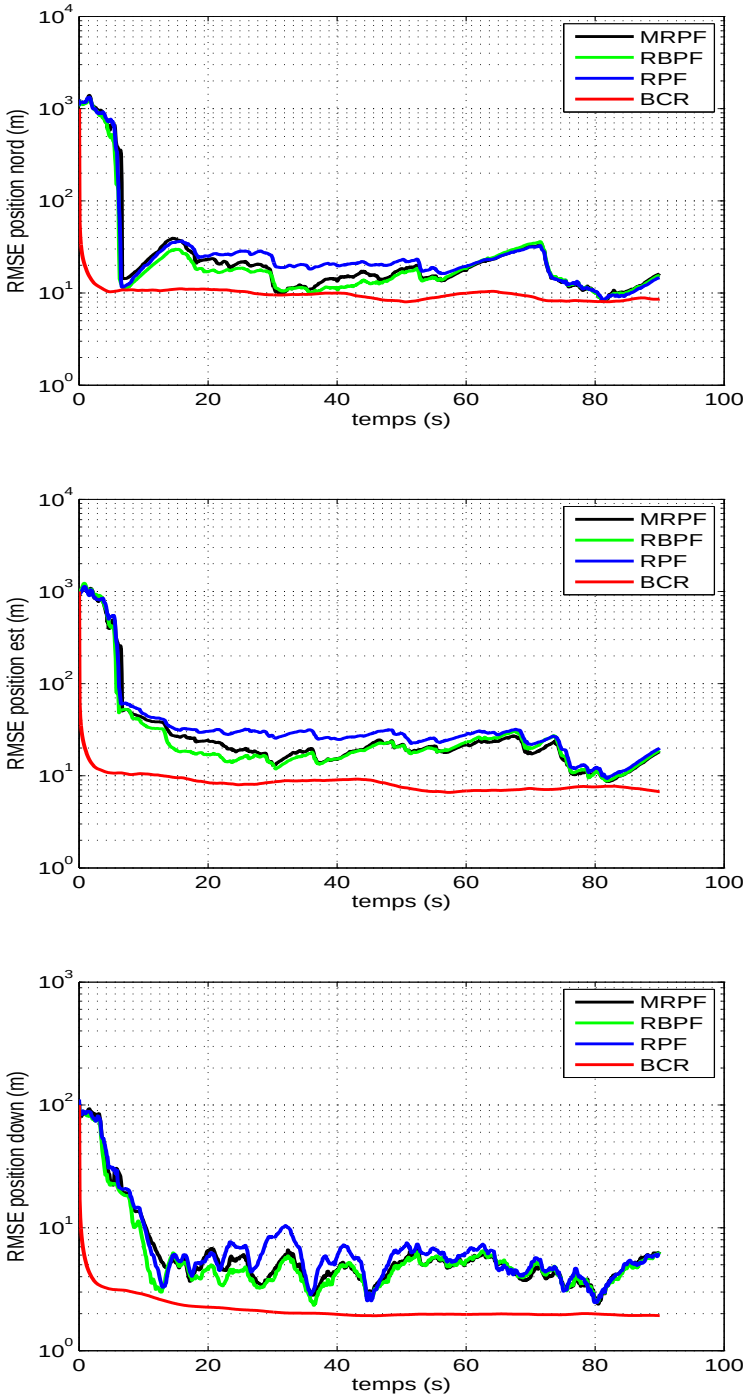


Figure 4.15 – RMSE en position ; noir : MRPF, vert : RBPF, bleu : RPF

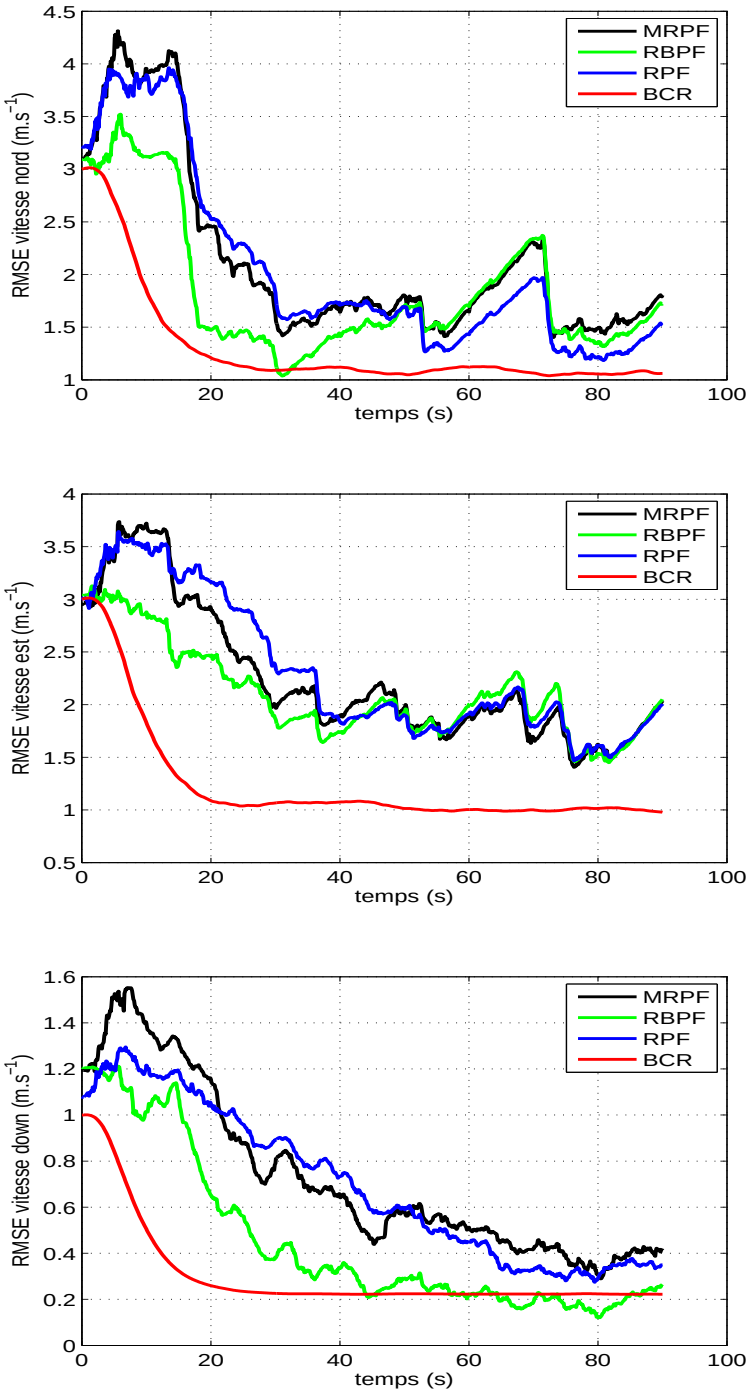


FIGURE 4.16 – RMSE en vitesse ; noir : MRPF, vert : RBPF, bleu : RPF

Scénario 2

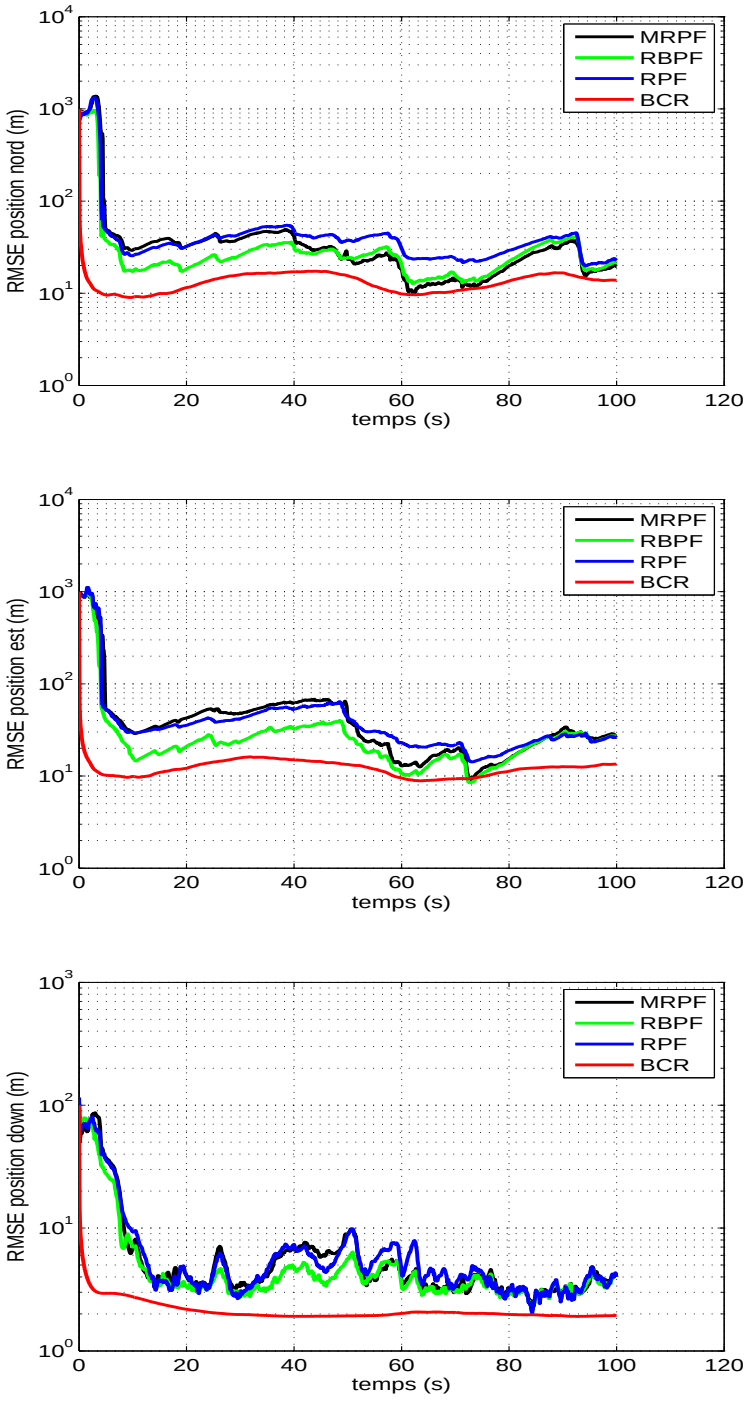


Figure 4.17 – RMSE en position ; noir : MRPF, vert : RBPF, bleu : RPF

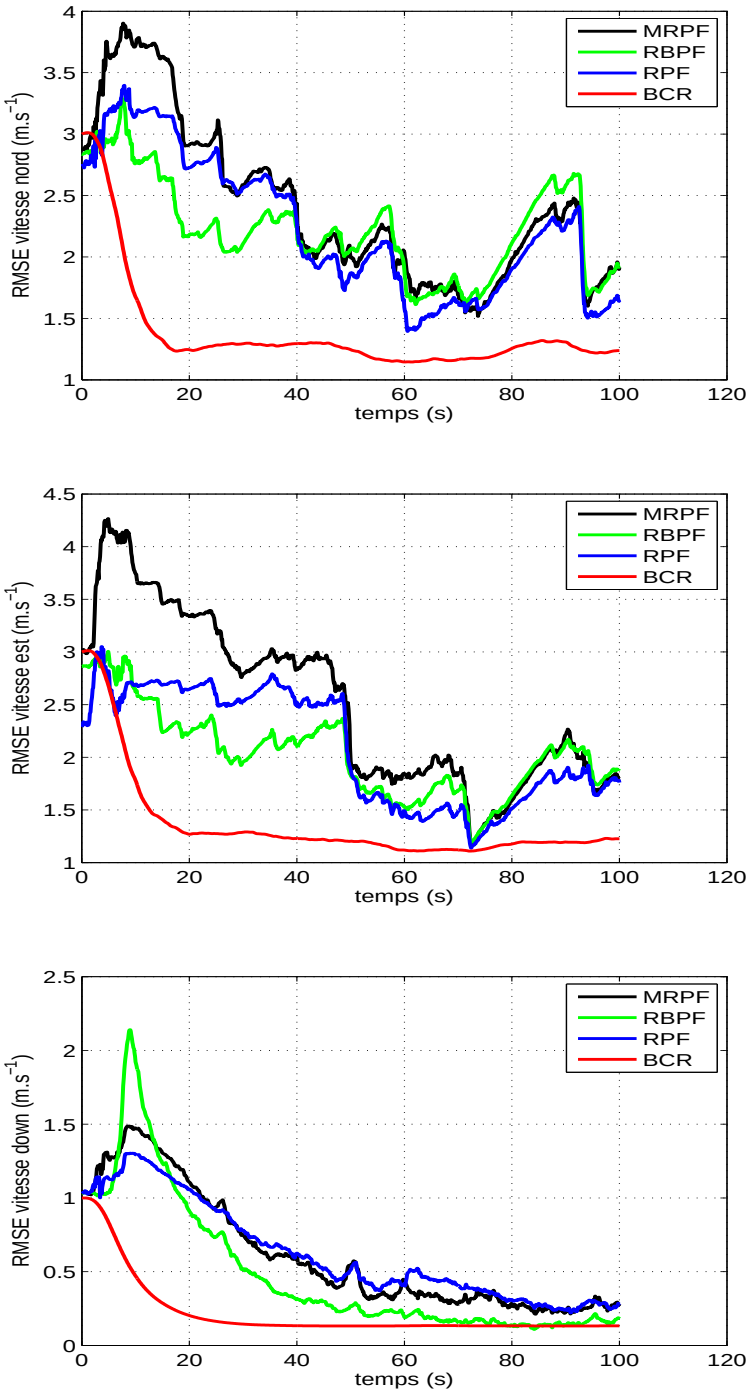


FIGURE 4.18 – RMSE en vitesse ; noir : MRPF, vert : RBPF, bleu : RPF

Scénario 3

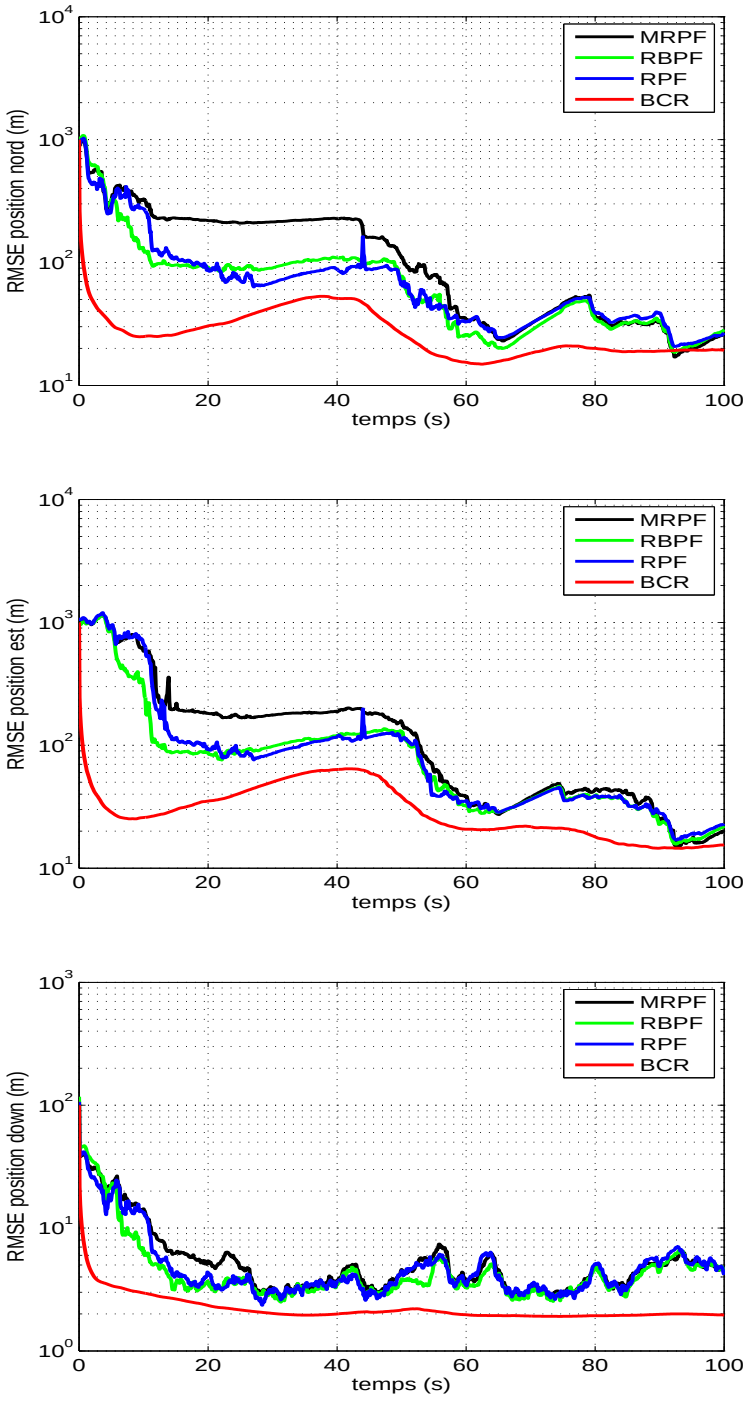


Figure 4.19 – RMSE en position ; noir : MRPF, vert : RBPF, bleu : RPF

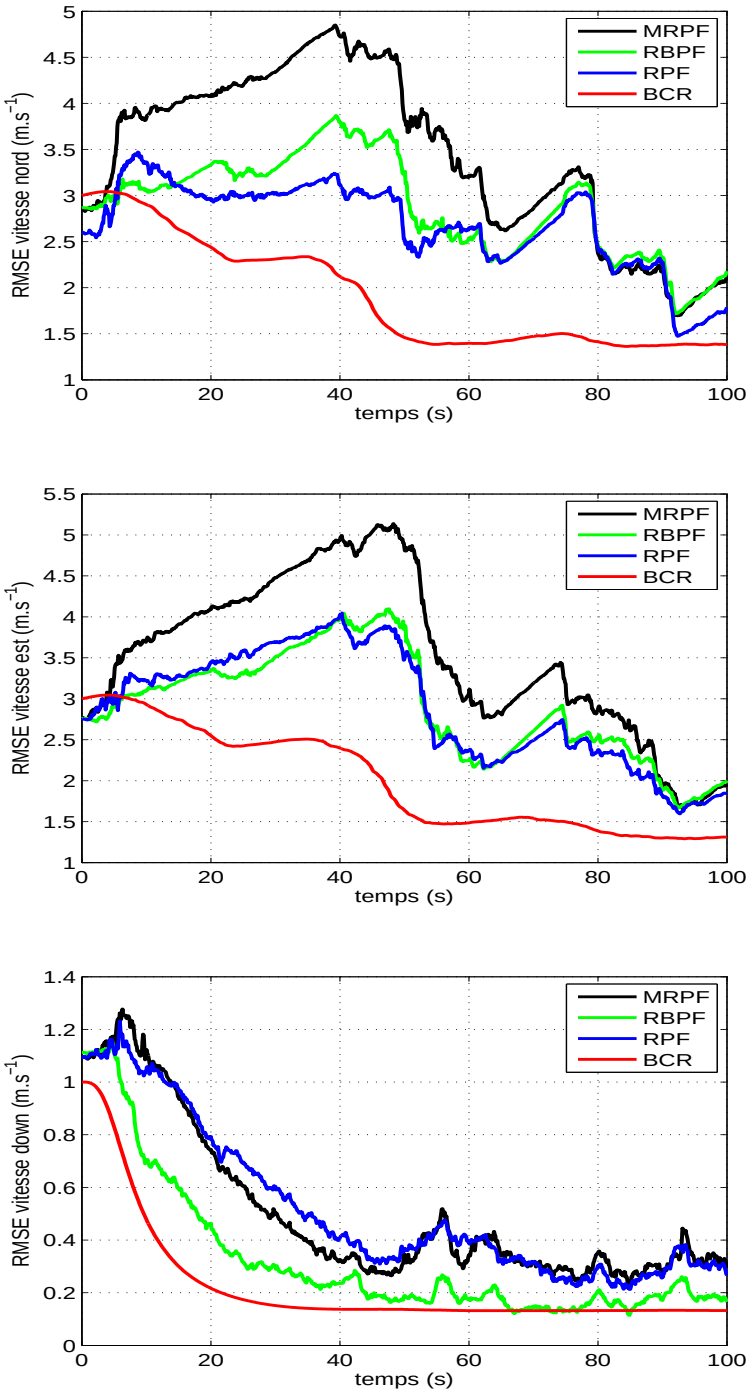


FIGURE 4.20 – RMSE en vitesse ; noir : MRPF, vert : RBPF, bleu : RPF

	MRPF	RBPF	RPF
scénario 1	94	93	73
scénario 2	93	97	60
scénario 3	90	94	50

Table 4.2 – Simulations A : Taux de non divergence (%)

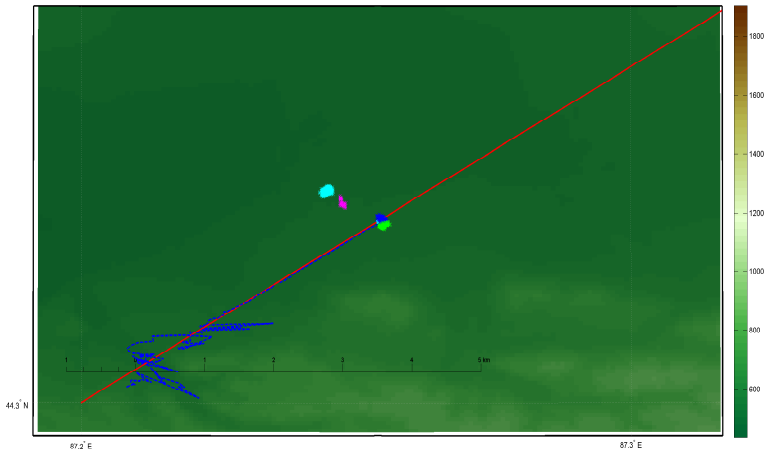


Figure 4.21 – Conservation des modes du MRPF dans le scénario 3 : chaque mode est représenté par un nuage de point de couleur différente

Analyse des résultats

- ▷ scénario 1 : L'analyse du RMSE fait apparaître une précision légèrement meilleure en position du MRPF par rapport au RPF sur la partie centrale de la trajectoire, les erreurs de vitesse étant comparables. De plus, on constate que le MRPF est plus robuste aux divergences que RPF (21% de cas non divergents en plus pour le MRPF).
- ▷ scénario 2 : ce terrain cumule deux types de difficultés pour le filtre particulaire : la présence d'ambiguïtés en début de recalage et l'apparition d'un dénivelé abrupt (canyon) sur la première portion du terrain, qui en terme de filtrage, correspond à un pic d'information élevé dont on a vu qu'il pouvait provoquer la divergence du filtre lorsque peu de particules tombent dans la zone d'intérêt. Ici le gain en robustesse du MRPF par rapport au RPF est bien plus marqué (33% de divergences en moins). L'erreur quadratique moyenne de la position pour les réalisations convergentes est la plus faible dans le RBPF en début de trajectoire ; sur la deuxième moitié de la trajectoire, le MRPF a un avantage en terme de précision sur le RPF et son RMSE est comparable à celui du RBPF.
- ▷ Ce scénario est le plus ambigu des trois : le MRPF converge à une fréquence plus élevée

que le RPF. On peut néanmoins constater une moindre précision du MRPF sur l'estimation des composantes de précision sur la la première moitié de la trajectoire. Ceci correspond au fait que l'algorithme maintient les modes plus longtemps que le RPF et le RBPF et que l'ambiguité persiste jusqu'à 45 secondes environ après le début du vol (cf. figure 4.21). Il se trouve que sur la portion de terrain considérée, les modes maintenus peuvent être relativement éloignés ce qui explique que l'estimation du MRPF (la moyenne pondérée) est moins précise sur la première moitié de la trajectoire.

Nous observons sur les figures 4.15, 4.17 et 4.19 que le RMSE est éloigné de la BCR dans les phases multimodales, la BCR étant un critère de performance locale. En outre, le RMSE n'est pas le critère de performance le plus pertinent dans les phases multimodales car il est calculé avec l'état vrai et non avec l'espérance conditionnelle exacte qui est inconnue.

Enfin, le temps de calcul du MRPF est de 1.8 fois supérieur à celui du RPF et 1.4 fois supérieur à celui du RBPF.

4.6.5 Simulations B : évaluation des performances du MRPF-MAP

Afin d'illustrer les performances du MRPF-MAP, nous avons considéré un cas où la dynamique de l'aéronef est faiblement bruitée avec un radio-altimètre délivrant des mesures précises. Comme nous l'avons évoqué dans la section 4.5.1, les méthodes particulaires où la prédiction se fait selon le noyau de transition sont souvent peu efficaces, puisque beaucoup de particules sont propagées dans des zones de faible vraisemblance.

Nous comparons ici le MRPF-MAP au MRPF en terme d'erreur quadratique moyenne et de robustesse aux divergences. 2 versions du MRPF-MAP utilisant deux densités d'importance légèrement différentes sont présentées-ci après :

- MRPF-MAP 1 : utilise une loi d'importance locale $\tilde{q} = \mathcal{N}(\hat{x}_j^*, P_j)$ (cf. 4.5.3.1) en cas d'incohérence entre la densité prédite et la fonction de vraisemblance. $\hat{x}_j^*$ et P_j désigne le MAP local et la covariance prédite du cluster d'indice j
- MRPF-MAP 2 : utilise une loi d'importance locale $\tilde{q} = \mathcal{N}(\hat{x}_j^*, P_{rot})$ en cas d'incohérence entre la densité prédite et la fonction de vraisemblance où P_{rot} est défini par (4.64).

Les performances du RPF standard étant peu satisfaisantes dans le cas qui nous intéressent, elles ne sont pas exposées dans les résultats numériques.

Afin de caractériser les performances du MRPF-MAP, nous distinguons deux situations :

Cas faiblement multimodal Nous établissons les performances du MRPF-MAP lorsque le nombre de modes de la densité a posteriori est petit (inférieur à 10). L'objectif est d'évaluer l'apport de l'échantillonnage d'importance, lorsque le coût lié à la recherche des maxima locaux est modeste.

Cette situation est simulée en considérant les paramètres suivants :

- écart-type du bruit de mesure du radio-altimètre $\sigma_v = 1\,m$
- L'état initial x_0 suit une distribution gaussienne centrée de covariance

$$P_0^{INS} = \mathrm{diag}((100\mathrm{m})^2,\ (100\mathrm{m})^2,\ (100\mathrm{m})^2,\ (3\mathrm{m}/s)^2,\ (3\mathrm{m}/s)^2,\ (1\mathrm{m}/s)^2)$$

Cas fortement multimodal Cette situation est caractérisée par la présence d'un nombre important de modes en début de recalage de sorte que l'échantillonnage d'importance basé sur les maxima locaux occasionne un coût de calcul non-négligeable s'il est appliqué brutalement. Afin de limiter ce surcoût algorithmique, l'échantillonnage d'importance de densité $\tilde{q}$ n'est activé uniquement que lorsque le filtre présente moins de 30 modes. Ce cas est simulé en prenant les paramètres suivants :

- écart-type du bruit de mesure du radio-altimètre $\sigma_v = 5\,m$
- L'état initial x_0 suit une distribution gaussienne centrée de covariance

$$P_0^{INS} = \mathrm{diag}((1000\mathrm{m})^2,\, (1000\mathrm{m})^2,\, (100\mathrm{m})^2,\, (3\mathrm{m}/s)^2,\, (3\mathrm{m}/s)^2,\, (1\mathrm{m}/s)^2)$$

4.6.5.1 Conditions de simulations

Paramètres propres aux filtres

MRPF

▷ $N_{MRPF} = 4000$ particules
▷ seuil de ré-échantilonnage : $N_{th} = \frac{N_{MRPF}}{2}$

MRPF-MAP

▷ $N_{MRPF-MAP} = 3000$ particules
▷ seuil de déclenchement de l'échantillonnage selon la loi d'importance centrée sur le MAP : $N_{th,MAP} = \frac{3}{4}N_{th}$
▷ Algorithme d'optimisation : BFGS [10]
▷ Paramètres de l'algorithme mean-shift clustering :
  - rayon de fusion $R_m = 25\,m$
  - nombre d'itérations maximum par procédure mean-shift : $n_{itermax} = 50$
  - seuil de convergence $\varepsilon^{MS} = 1\,m$
▷ seuil de suppression de cluster $\alpha_{min} = 10^{-10}$

4.6.5.2 Résultats numériques

Cas faiblement multimodal (scénario 2)

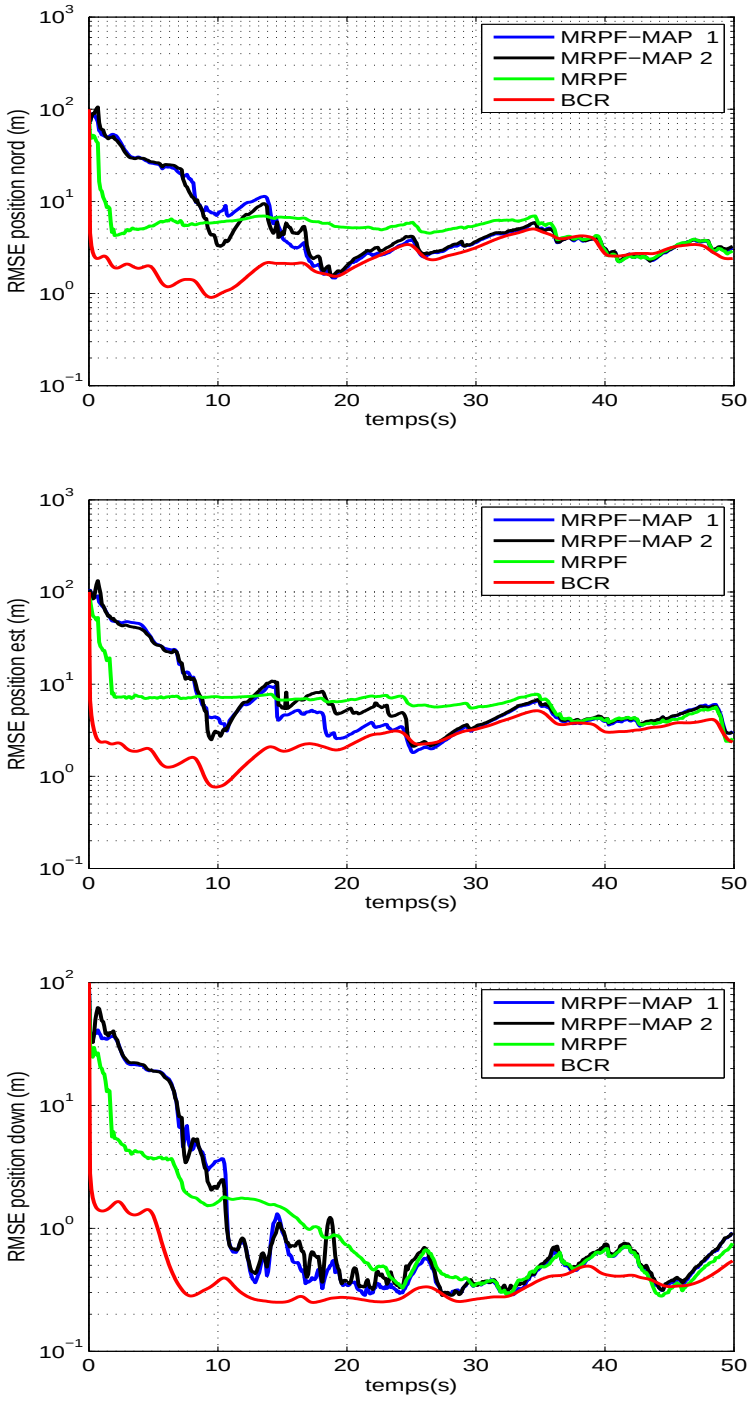


FIGURE 4.22 – RMSE en position

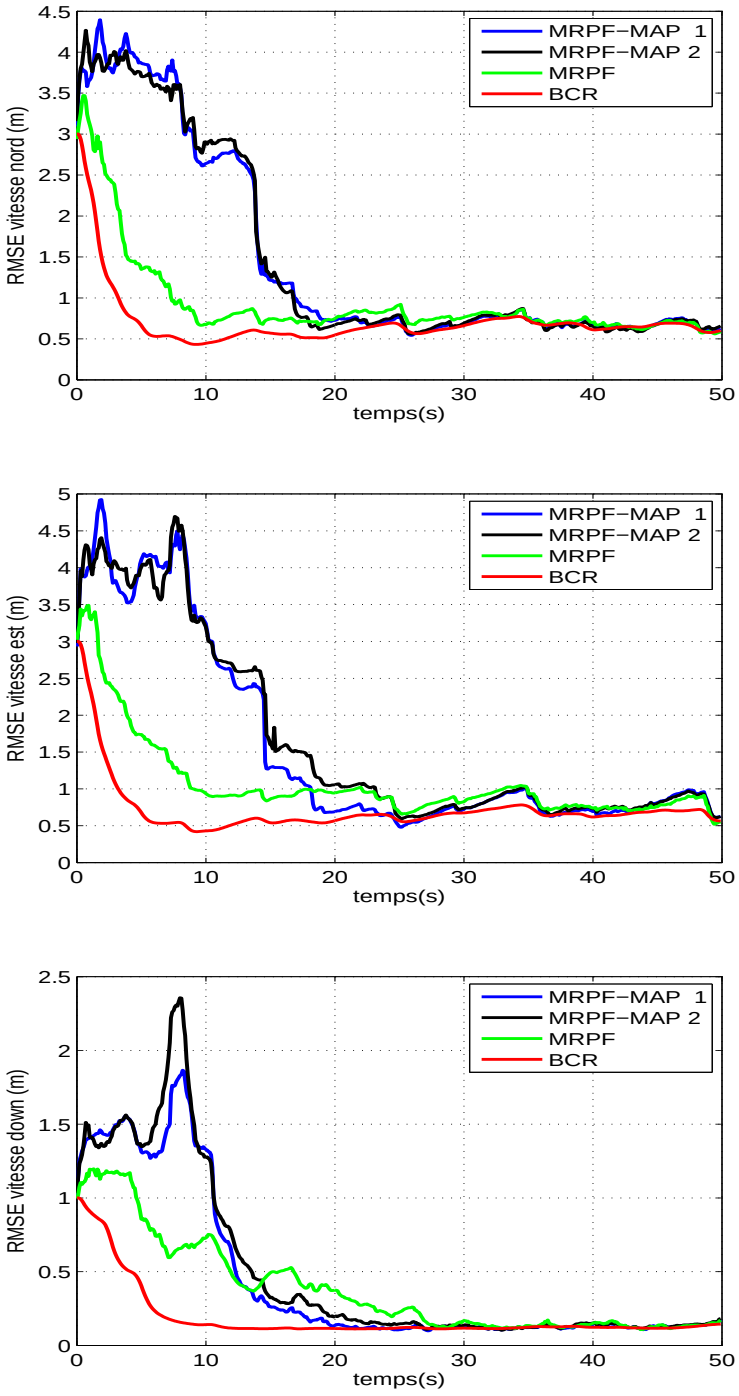


Figure 4.23 – RMSE en vitesse

Analyse des résultats

Les taux de non-divergence sont de 92% pour le MRPF-MAP 1 et le MRPF-MAP 2 et de 78% pour le MRPF standard. Les courbes d'erreur quadratique moyenne de la figure 4.22 montrent que la borne de Cramer-Rao a posteriori sur l'erreur de postition est atteinte plus rapidement dans le cas de l'utilisation du MRPF. En revanche, le RMSE est plus élevé en début de recalage, ce qui peut s'expliquer par le fait que le MRPF-MAP maintient plus longtemps des clusters dont le recouvrement avec la fonction de vraisemblance est initialement faible mais augmente petit à petit. Les choix différents de la covariance de la densité d'importance ne semblent pas avoir d'incidence sur la robustesse ou la précision du MRPF-MAP.

Cas fortement multimodal

Scénario 1

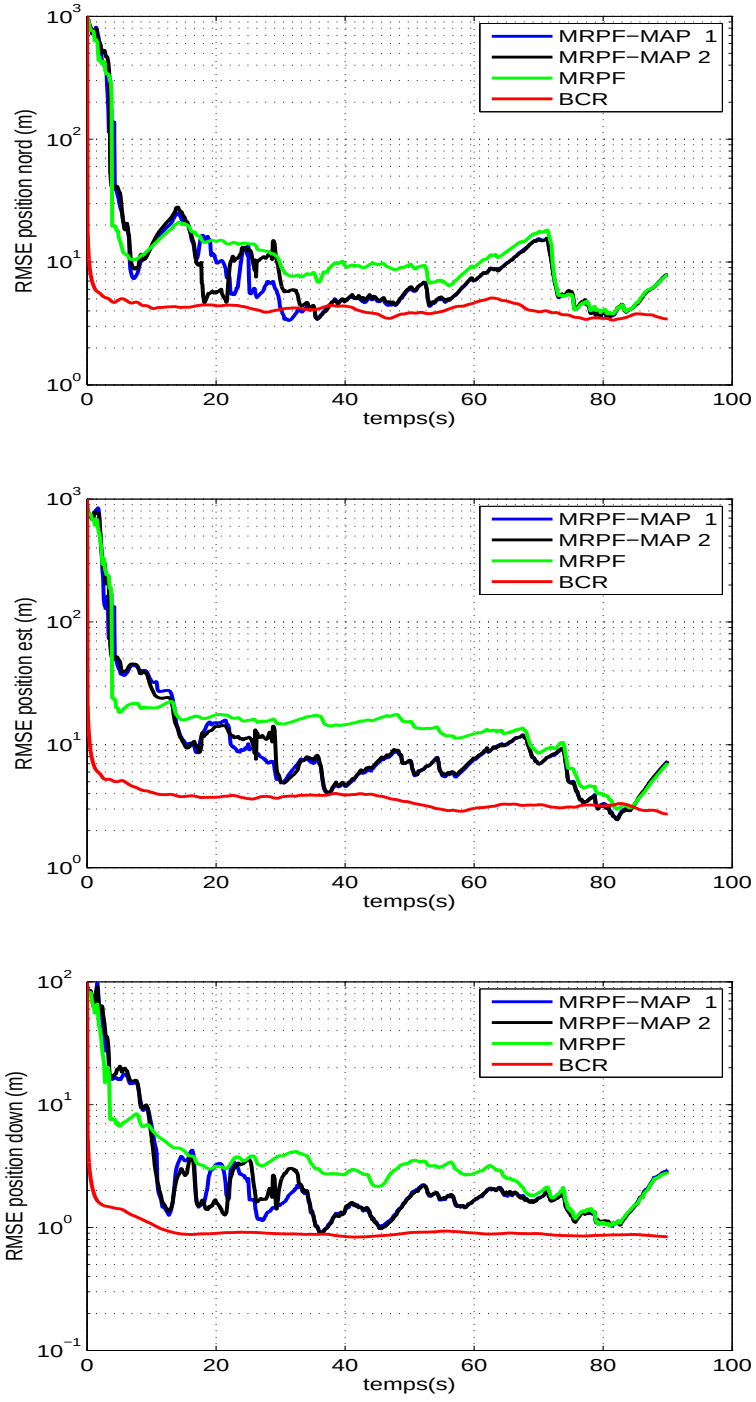


Figure 4.24 – RMSE en position

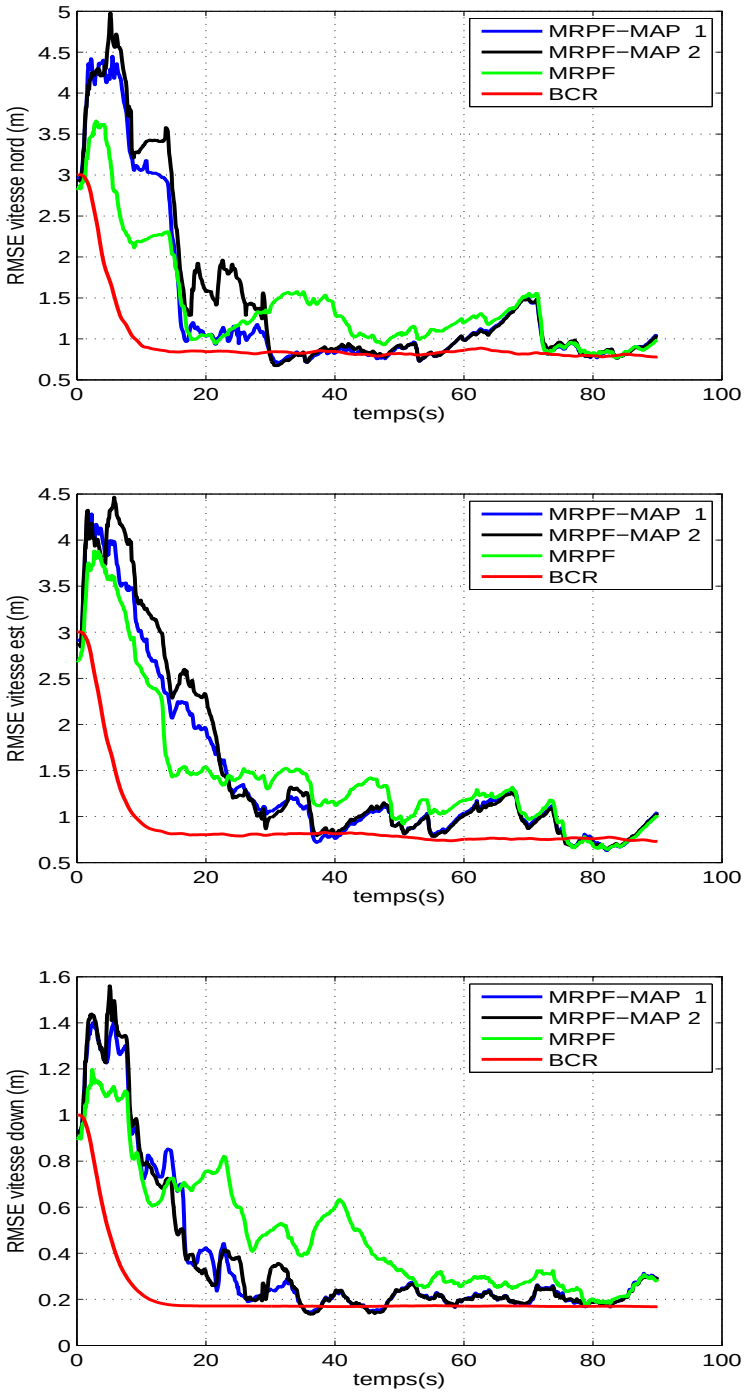


FIGURE 4.25 – RMSE en vitesse

Scénario 2

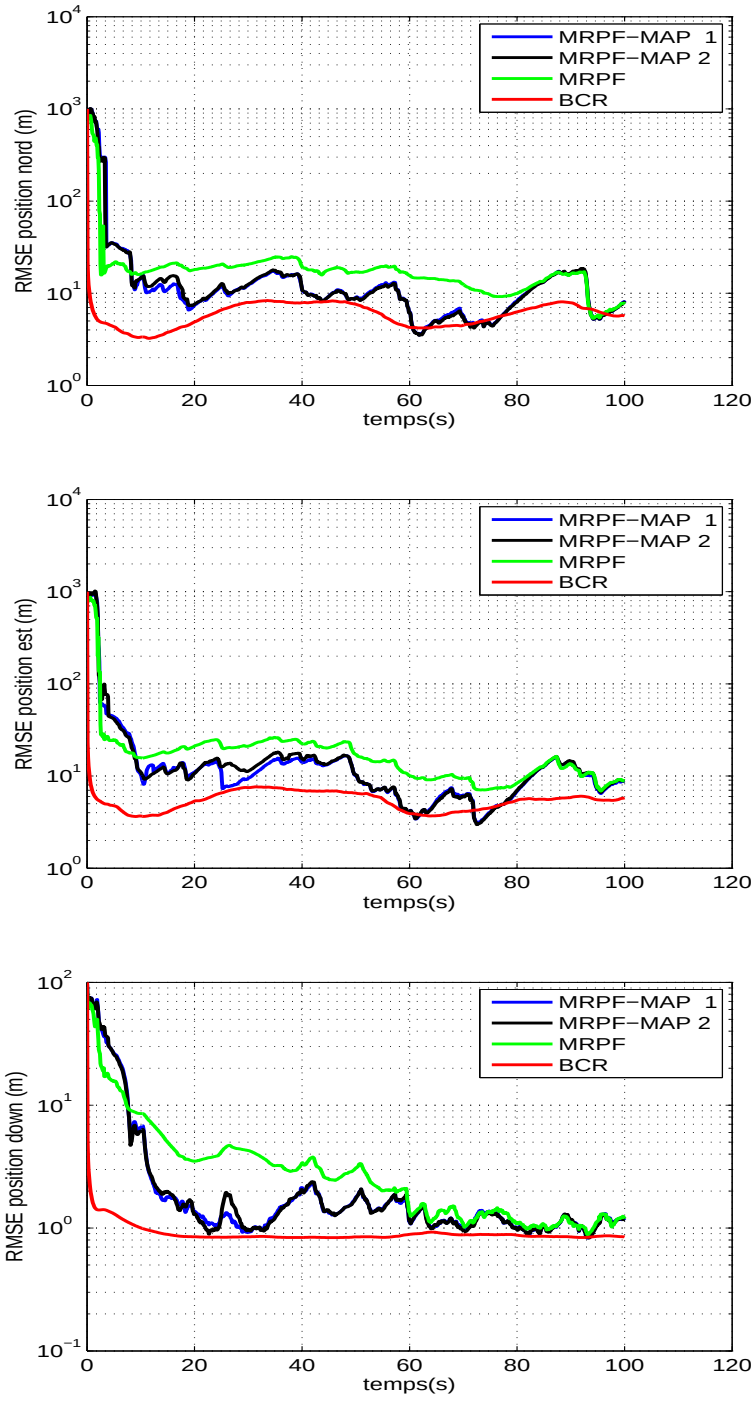


Figure 4.26 – RMSE en position

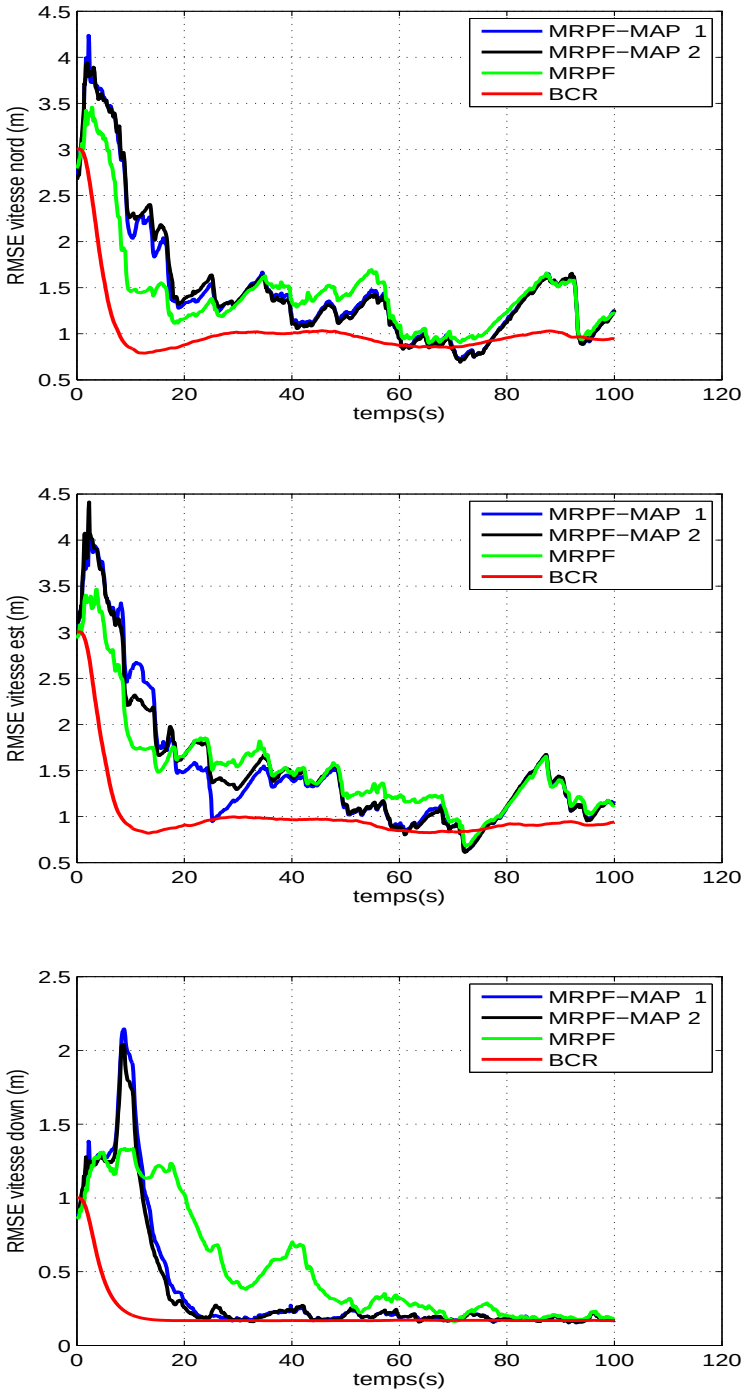


FIGURE 4.27 – RMSE en vitesse

Scénario 3

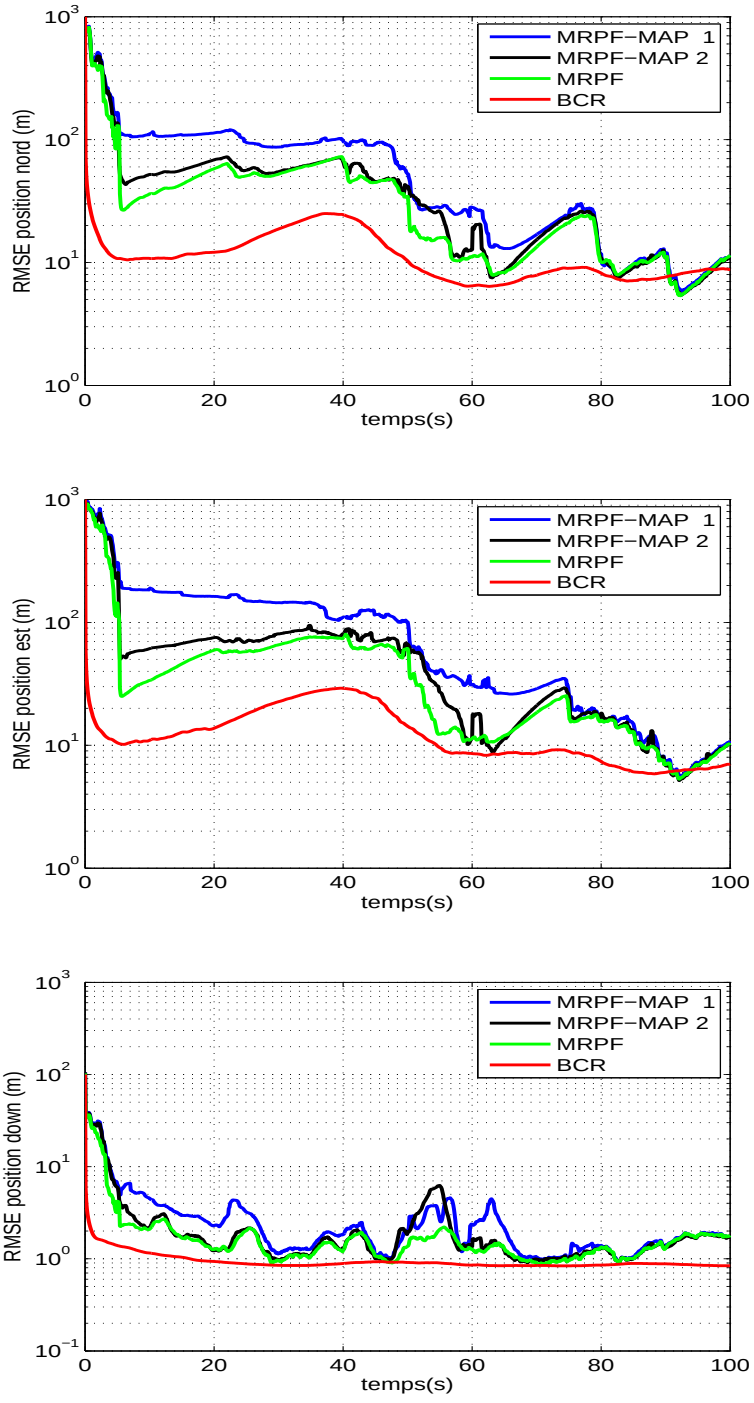


Figure 4.28 – RMSE en position

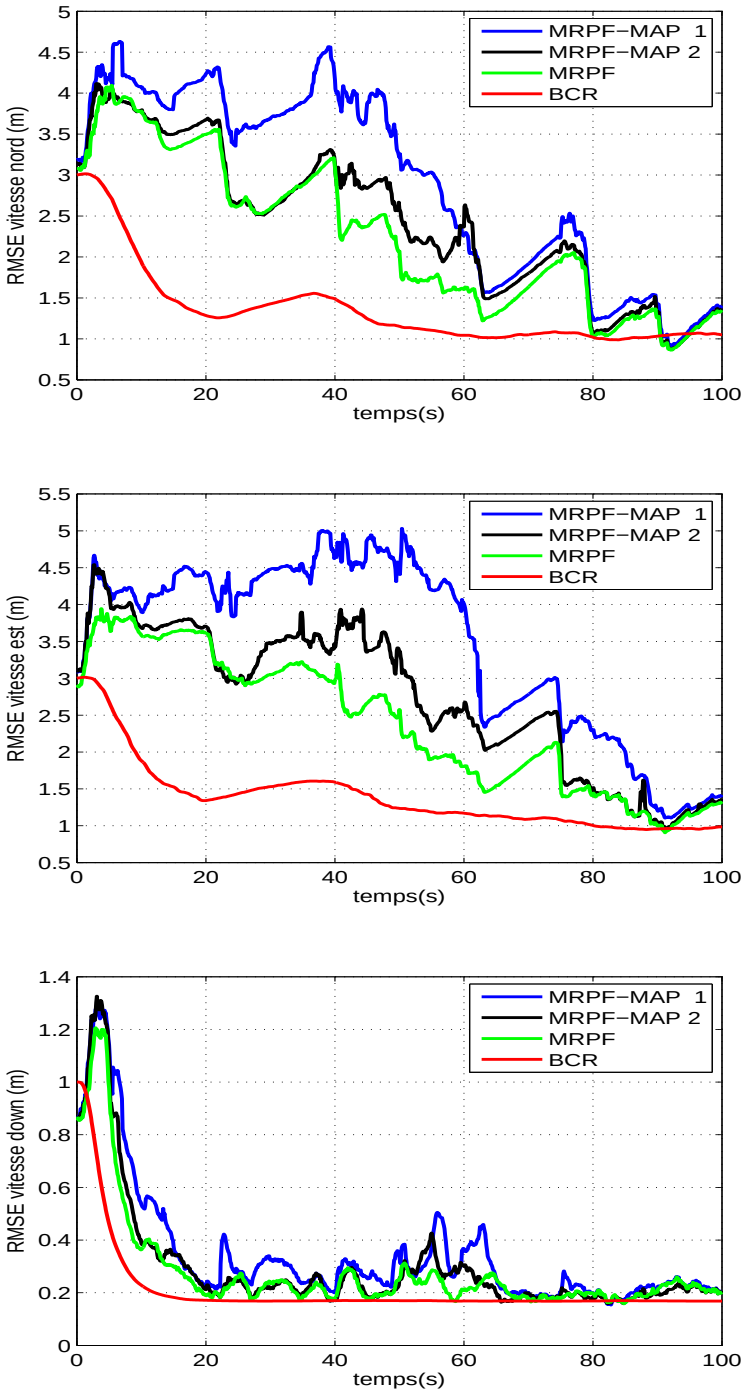


FIGURE 4.29 – RMSE en vitesse

	MRPF-MAP 1	MRPF-MAP 2	MRPF
scénario 1	78	78	72
scénario 2	80	82	73
scénario 3	78	76	73

Table 4.3 – Simulations B : Taux de non divergence (%)

Analyse des résultats

- ▷ scénario 1 : les deux versions du MRPF-MAP sont plus précises en termes de RMSE comparativement au MRPF sur une bonne partie de la trajectoire. En revanche, il semble que l'utilisation de la densité d'importance $\tilde{q}_2$ ne permet pas de gagner en précision sur l'estimation de la position et de la vitesse de l'aéronef par rapport à $\tilde{q}_1$ (voir la figure 4.24). Les deux versions du MRPF-MAP sont également plus robustes aux divergences (6% de cas convergents supplémentaires) que le MRPF standard.
- ▷ scénario 2 : même constat ici ; les deux versions du MRPF-MAP permettent de gagner jusqu'à 10 mètres de précision supplémentaire en position nord et est dans la phase monomodale, et 3 mètres en erreur verticale (cf. figure 4.26). Le MRPF-MAP 1 et le MRPF-MAP 2 présentent ici des gains en taux de non-divergence de 7% et 9% respectivement relativement au MRPF. L'utilisation de la densité d'importance $\tilde{q}_2$ dans le MRPF-MAP permet d'obtenir une modeste augmentation des cas convergents par rapport à $\tilde{q}_1$.
- ▷ scénario 3 : Le MRPF apporte un gain modéré en terme de robustesse (5% de cas convergents en plus pour le MRPF-MAP 1, 3% pour le MRPF-MAP 2).

4.6.6 Simulations C : évaluation des performances du MRBPF-MAP

Le MRBPF-MAP est l'extension du MRPF-MAP obtenu en considérant un filtre particulaire Rao-Blackwellisé à la place du RPF standard. Dans l'algorithme MRBPF-MAP que nous avons implémenté, la procédure de régularisation a été conservé afin d'assurer la robustesse au petit bruit d'état. L'extension au RBPF est justifié expérimentalement par le fait que le filtre Rao-Blackwellisé élimine plus rapidement les modes non-pertinents de la densité a posteriori. Dans ce cas, on peut espérer utiliser l'échantillonnage d'importance fondé sur les modes locaux dès les premiers instants de filtrage, là où il faut attendre quelques itérations dans le cas du MRPF-MAP.
Dans les simulations suivantes, l'échantillonnage selon la densité d'importance $\tilde{q} = \tilde{q}_2$ n'est enclenché que lorsque le filtre comporte moins de 20 modes afin de limiter le temps de calcul lié à la recherche des modes. Les paramètres du modèles sont les suivants :

- ▷ écart-type du bruit de mesure du radio-altimètre : $\sigma_v = 5\,m$
- ▷ l'état initial x_0 suit une distribution gaussienne centrée de covariance

$$P_0^{INS} = \mathrm{diag}((1000\mathrm{m})^2,\ (1000\mathrm{m})^2,\ (100\mathrm{m})^2,\ (3\mathrm{m/s})^2,\ (3\mathrm{m/s})^2,\ (1\mathrm{m/s})^2)$$

Paramètres propres aux filtres

RBPF

▷ $N_{RBPF} = 4000$ particules
▷ seuil de ré-échantilonnage : $N_{th} = \frac{N_{RBPF}}{3}$

MRBPF et MRBPF-MAP

▷ $N_{MRBPF} = 4000$ particules
▷ $N_{MRBPF-MAP} = 3000$ particules
▷ seuil de déclenchement de l'échantillonnage selon la loi d'importance centrée sur le MAP : $N_{th,MAP} = \frac{3}{4} N_{th}$
▷ seuil de suppression de cluster $\alpha_{min} = 10^{-20}$

4.6.6.1 Résultats numériques

Scénario 1

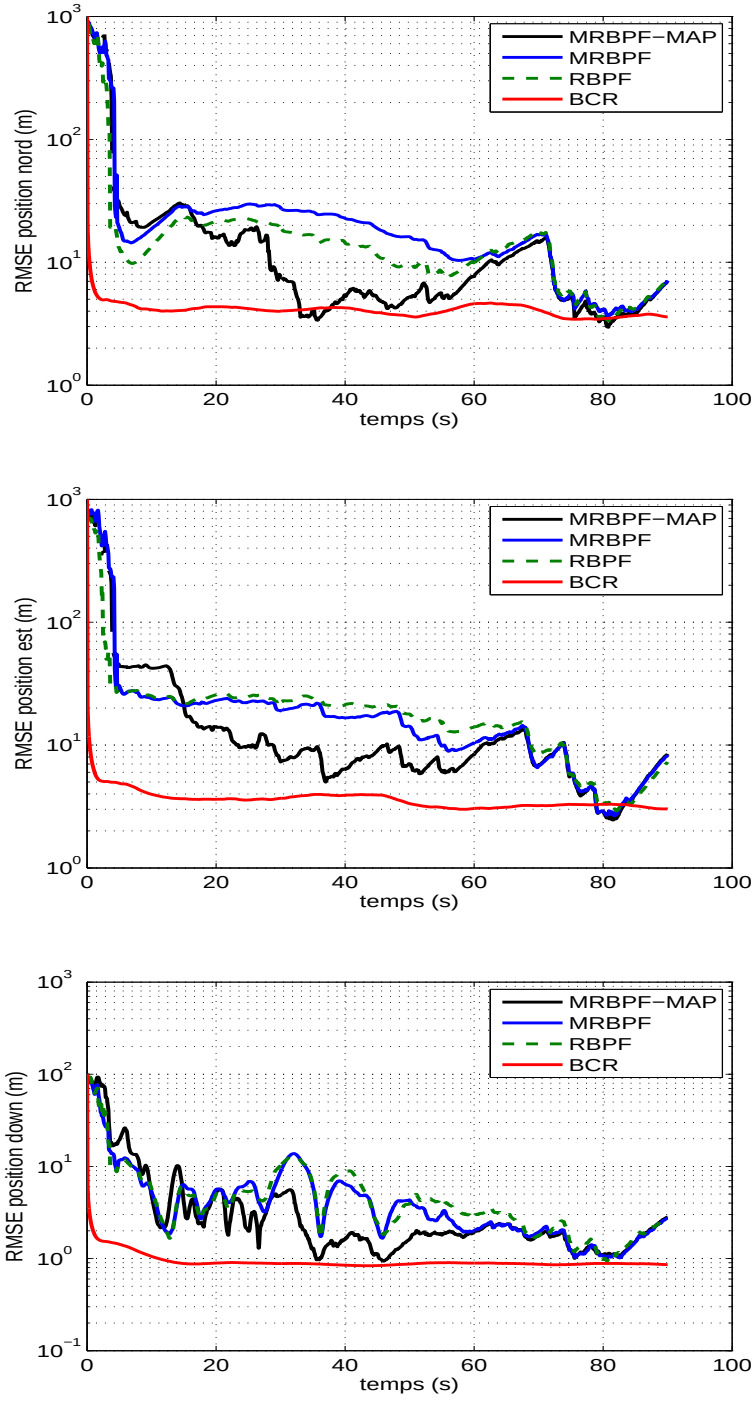


Figure 4.30 – RMSE en position

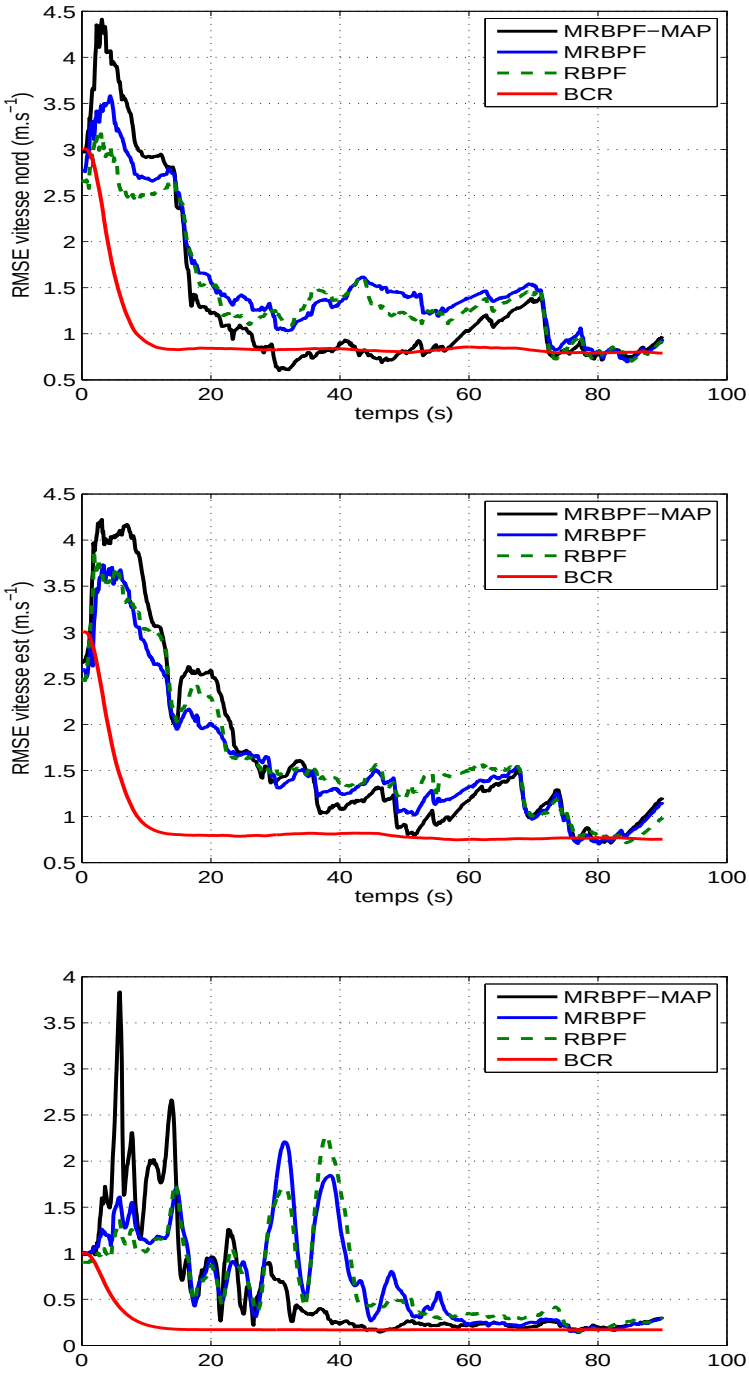


FIGURE 4.31 – RMSE en vitesse

Scénario 2

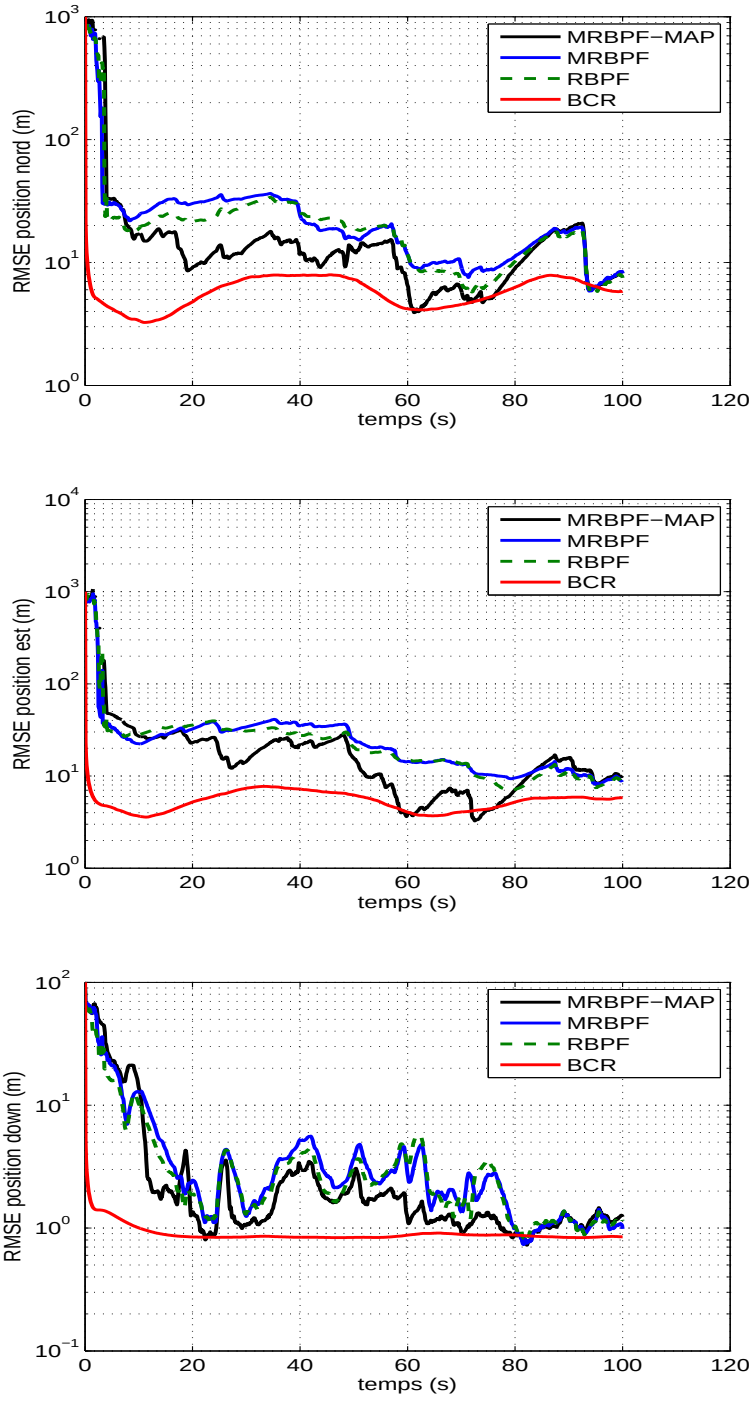


Figure 4.32 – RMSE en position

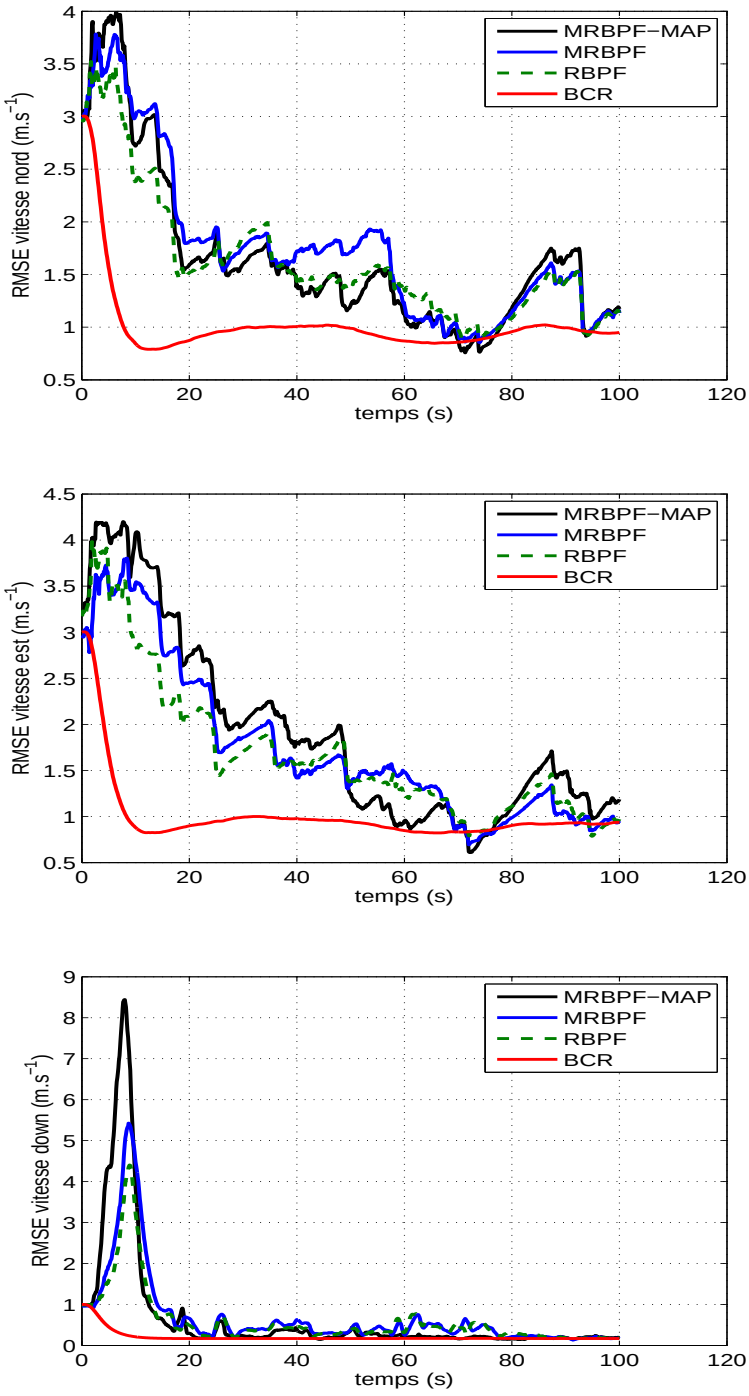


FIGURE 4.33 – RMSE en vitesse

Scénario 3

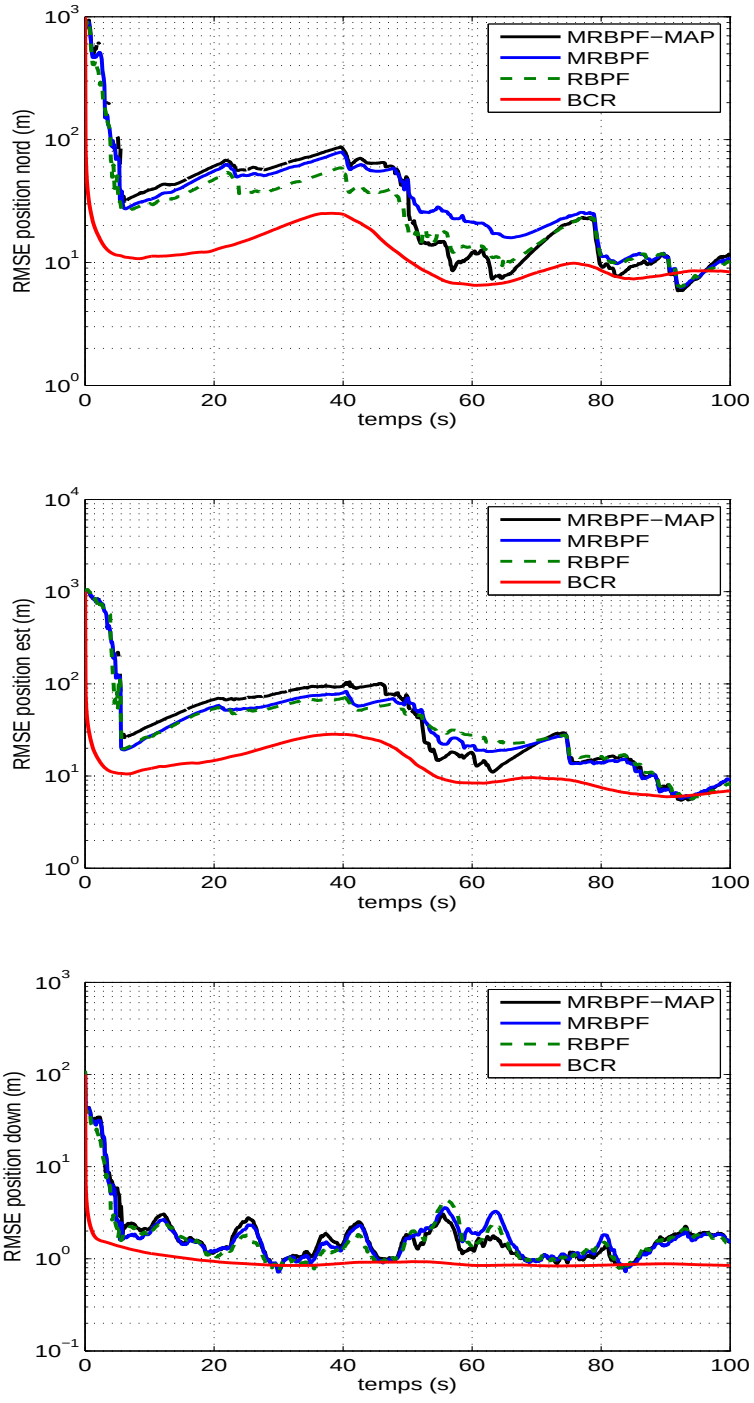


Figure 4.34 – RMSE en position

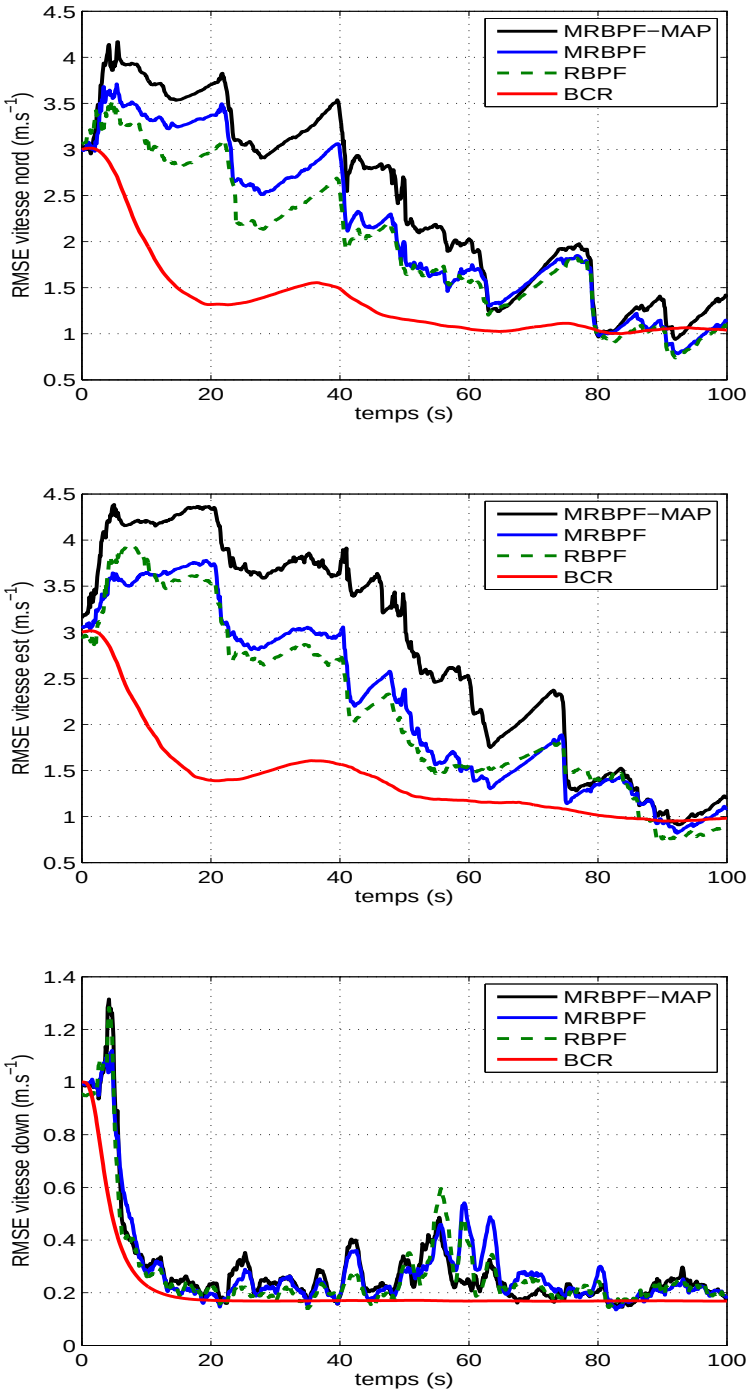


FIGURE 4.35 – RMSE en vitesse

	RBPF	MRBPF	MRBPF-MAP
scénario 1	71	74	86
scénario 2	83	79	93
scénario 3	71	73	84

Table 4.4 – Simulations C : Taux de non divergence (%)

Analyse des résultats

- ▷ scénario 1 : de la 30ème à la 60ème seconde de la trajectoire, le RMSE des composantes de position du MRBPF-MAP est significativement inférieur à celui du RBPF et du MRBPF comme le montre la figure 4.30. Le gain en précision sur les composantes de vitesse est peu significatif excepté pour la composante de vitesse verticale. Enfin, le gain du MRBPF-MAP en terme de taux non-divergence est de 12% comparé au MRBPF et de 15% par rapport au RBPF standard.
- ▷ scénario 2 : Ici, on remarque également que le RMSE en position du MRBPF-MAP est inférieur à celui du RBPF et du MRBPF à partir de la 20ème seconde (fig. 4.32). L'analyse de la figure 4.33 révèle que les composantes de vitesses ne sont pas estimées avec plus de précision par le MRBPF-MAP. Le taux de non-divergence du MRBPF-MAP est supérieur de 14% par rapport au MRBPF et 10% par rapport au RBPF.
- ▷ scénario 3 : La figure 4.34 ne montre pas d'amélioration appréciable en précision pour le MRBPF-MAP pour les composantes de position. Le RMSE du MRBPF-MAP est légèrement plus élevé pour les composantes de vitesse horizontale (cf. fig. 4.35). En revanche le MRBPF-MAP converge à 84% contre 73% pour le MRBPF et 71% pour le RBPF.

Pour le choix de paramètres précédent, le temps de calcul du MRBPF-MAP est environ 2.3 fois celui du RBPF.

Conclusion du Chapitre 4

Nous avons développé dans ce chapitre un nouvel algorithme, le MRPF, permettant de gérer le caractère fortement multimodal de la loi a posteriori en début de recalage. Cet algorithme est étendu au filtre Rao-Blackwellisé et donne lieu au MRBPF. L'un des apports du MRPF est l'identification des modes de la loi empirique du filtre particulaire grâce à l'algorithme mean-shift. La formulation du MRPF sous forme de mélange de densités permet la régularisation locale du nuage de particules qui est plus robuste que l'implémentation classique de la régularisation qui suppose implicitement une loi a posteriori monomodale. Enfin, nous avons introduit une procédure d'échantillonnage d'importance fondée sur les modes locaux de la densité a posteriori. Chaque mode local est calculé en approximant la densité prédite de chaque composante du mélange par une densité gaussienne dont les moments sont calculés empiriquement. La matrice de covariance de la densité d'importance est orientée suivant les axes principaux de la loi a posteriori, ce qui permet de générer le nuage de particules dans les zones d'intérêt. Le déclenchement de cette procédure d'échantillonnage d'importance est fait adaptativement en contrôlant la taille effective de l'échantillon au niveau de chaque cluster. L'algorithme résultant, le MRPF-MAP, a été comparé au MRPF dans un cas où les techniques d'échantillonnage classiques sont mises en difficulté, par exemple lorsque la variance du bruit de mesure est petite. Des gains en robustesse et en précision des filtres proposés ont été observés sur ces scénarios. L'extension du MRPF-MAP au filtre Rao-Blackwellisé, notée MRBPF-MAP, a également été implémentée. Les simulations ont montré que l'utilisation de la procédure d'échantillonnage d'importance précédente permet d'obtenir des gains en terme de taux non-divergence encore plus significatifs que ceux obtenus avec le MRPF-MAP.

Chapitre 5

Test d'intégrité pour le filtrage particulaire : application au recalage radio-altimétrique

5.1 Introduction

Les algorithmes de filtrage particulaire permettent d'obtenir une estimation sur l'état d'un système donné avec une confiance associée. Cette estimation peut servir d'entrée à une architecture dont la fonction est potentiellement critique : il est donc nécessaire de contrôler tout au long de la procédure d'estimation la fiabilité de l'estimée particulaire. L'objectif de ce chapitre est de proposer un cadre de détection séquentielle de divergence du filtre particulaire. En première approximation, nous définirons la divergence du filtre comme un état tel que l'erreur d'estimation est grande devant le domaine de confiance représenté par le système de particules et tel que la succession des étapes de prédiction, de correction et de ré-échantillonnage ne permet pas au filtre de converger vers l'état vrai.
Nous ouvrons ce chapitre avec un aperçu rapide des méthodes statistiques pour la détection de changement. Après avoir rappelé quelques résultats sur les détecteurs de divergence en filtrage de Kalman, nous formulons un test d'hypothèse pour la détection de divergence adapté aux filtres particulaires, que nous mettons en oeuvre grâce à l'algorithme CUSUM [98]. Nous illustrons ensuite les performances du détecteur de divergence sur la problématique du recalage altimétrique par filtrage particulaire et nous évaluons l'apport de la ré-initialisation du filtre en cas de détection de divergence.

5.2 Outils statistiques pour la détection de changement

5.2.1 Test d'hypothèses

Dans les applications de détection de pannes ou de contrôle d'intégrité, on s'intéresse à une suite de données suivant un modèle paramétrique dont les paramètres subissent des changements

brusques à des instants inconnus [4] : ces changements indiquent typiquement un évènement anormal (panne, ...). L'objectif est alors de déterminer au plus tôt si un changement a eu lieu et éventuellement à quel instant il est survenu.

Formellement, c'est un problème de test d'hypothèse qui peut être séquentiel ou non. Soit $(u_k)_{k\geq 1}$ une suite variables aléatoires de $\mathbb{R}^r$, caractéristiques du système que l'on souhaite contrôler. On suppose que les u_k sont indépendants et distribués selon une loi paramétrique de densité p_θ, $\theta \in \Theta$ où Θ est un sous-ensemble de $\mathbb{R}$. Soit Θ_0 et Θ_1 deux sous-ensembles disjoints de l'espace des paramètres Θ. On souhaite décider entre les hypothèses H_0 et H_1 définis par

$$\begin{cases} H_0 & : \quad \theta \in \Theta_0 \\ H_1 & : \quad \theta \in \Theta_1 \end{cases} \tag{5.1}$$

Deux cas peuvent se produire :

1. $\Theta_0 = \{\theta_0\}$ et $\Theta_1 = \{\theta_1\}$ avec θ_0, $\theta_1 \in \mathbb{R}$. C'est un cas de test d'hypothèses simple.
2. Au moins l'un des ensembles Θ_i, $i = 1, 2$ contient plus d'un élément : c'est un test d'hypothèses composites

Définition 1. *Etant donné un n-échantillon $u_1, \ldots, u_n$, un test statistique entre deux hypothèses H_0 et H_1 est une application $D : \mathbb{R}^{r\times n} \mapsto \{0, 1\}$ qui à $(u_1, \ldots, u_n) \stackrel{\triangle}{=} \mathcal{U}_1^n$ associe*

$$D(\mathcal{U}_1^n) = \begin{cases} 0 & \textit{si on accepte } H_0 \\ 1 & \textit{si on accepte } H_1 \end{cases}$$

$D(\mathcal{U}_1^n)$ est une variable aléatoire appelée règle (ou fonction) de décision.

La qualité d'un test d'hypothèses est habituellement évaluée grâce aux erreurs de première et deuxièmes espèces. L'erreur de première espèce $\alpha_0(\theta)$ est la probabilité de rejeter H_0 sachant que H_0 est vrai tandis que l'erreur de deuxième espèce $\alpha_1(\theta)$ est la probabilité de rejeter H_1 à tort, i.e.

$$\alpha_0 = \mathbb{P}(D(\mathcal{U}_1^n) \neq 0 | H_0)$$
$$\alpha_1 = \mathbb{P}(D(\mathcal{U}_1^n) \neq 1 | H_1)$$

Enfin la puissance d'un test β est définie comme la probabilité d'accepter H_1 lorsque $\theta \in \Theta_1$, càd $\beta = 1 - \alpha_1$.
Lorsque l'hypothèse est composite α_0 dépend de la valeur inconnue du paramètre $\theta \in \Theta_0$. On définit alors le niveau α du test comme $\alpha = \sup_{\theta\in\Theta_0} \alpha_0(\theta)$.

5.2.2 Test séquentiel entre hypothèses simples

L'analyse séquentielle consiste à déterminer un test entre deux hypothèses H_0 et H_1, lorsque les données $(u_k)_{k\geq 1}$ sont observées les unes après les autres [119, 4].

Définition 2. *Un test séquentiel entre H_0 et H_1 est un couple (D, T) où T est un temps d'arrêt et $D(\mathcal{U}_1^T)$ une règle de décision.*

En général, les test séquentiels exigent une taille de données moindre pour des risques α_0 et α_1 donnés par rapport aux tests non-séquentiels.

Définition 3. *Etant donné deux hypothèses simples $H_0 : \theta = \theta_0$ et $H_1 : \theta = \theta_1$, on définit le rapport de vraisemblance $S_k = \ln \frac{p_{\theta_1}(\mathcal{U}_1^k)}{p_{\theta_0}(\mathcal{U}_1^k)}$. (D, T) est un test séquentiel du rapport de probabilités (ou* SPRT *pour* Sequential Probability Ratio Test*) entre H_0 et H_1 lorsque les données $(u_k)_{k\geq 1}$ sont observées séquentiellement et qu'à l'instant k :*

- *on accepte H_0 lorsque $S_k \leq -a$*
- *on accepte H_1 lorsque $S_k \geq h$*
- *on recueille les données et on teste tant que $-a < S_k < h$*

où a et h sont des seuils finis.

La règle de décision correspondante est

$$D(\mathcal{U}_1^T) = \begin{cases} 0 & \text{si } S_T \leq -a \\ 1 & \text{si } S_T \geq h \end{cases} \tag{5.2}$$

où $T = \min\{n \geq 1 \mid \{S_n \leq -a\} \text{ ou } \{S_k \geq h\}\}$ est le temps d'arrêt.

5.2.3 Algorithme CUSUM pour la détection de changement en ligne

L'algorithme CUSUM (CUmulative SUM) de Page [98] est un cas particulier de SPRT pour le test de changement de moyenne d'une distribution gaussienne univarié très utilisé en détection de changement. Dans sa version originale, on suppose que la suite de v.a. $(u_k)_k$ suit une distribution gaussienne de densité p_θ où θ est la moyenne et σ^2 la variance. De plus, on suppose qu'à un instant t_0 inconnu, le paramètre θ change de valeur,

$$\begin{cases} \theta = \theta_0 & \text{si } t < t_0 \\ \theta = \theta_1 & \text{si } t \geq t_0 \end{cases} \tag{5.3}$$

L'objectif étant de détecter la présence d'un changement de paramètre, le principe du CUSUM est de tester de manière répétée entre $H_0 : \theta = \theta_0$ et $H_1 : \theta = \theta_1$ en utilisant le SPRT. L'idée de Page est alors de réinitialiser le CUSUM tant que la décision est 0. Dès que la décision vaut 1, la procédure s'arrête et l'instant t_a correspondant est le temps d'alarme. Le SPRT nécessite de préciser deux seuils $-a$ et h : Page a montré de manière intuitive que le seuil a optimal est $a = 0$, ce qui fut démontré de manière plus rigoureuse par Shiryaev [111].
En notant $D_k = D(\mathcal{U}_1^k)$ la règle de décision, celle-ci peut s'exprimer de manière récursive pour une distribution de densité p_θ quelconque, comme :

$$D_k = \begin{cases} D_{k-1} + \ln \frac{p_{\theta_1}(u_k)}{p_{\theta_0}(u_k)} & \text{si } D_{k-1} + \ln \frac{p_{\theta_1}(u_k)}{p_{\theta_0}(u_k)} > 0 \\ 0 & \text{sinon} \end{cases} \tag{5.4}$$

Le temps d'arrêt t_a est alors défini par

$$t_a = \inf \{k \geq 1 \mid D_k \geq h\} \tag{5.5}$$

Dans le cas où p_θ est une densité gaussienne de moyenne θ, la fonction de décision se ré-écrit de la manière suivante :

$$D_k = \left[D_{k-1} + \frac{\theta_1 - \theta_0}{\sigma^2} \left(u_k - \frac{\theta_1 + \theta_0}{2} \right) \right]^+ \tag{5.6}$$

où pour $x \in \mathbb{R}$, $(x)^+ = \max(0, x)$.

Supposons que le temps de changement t_0 soit inconnu mais déterministe.

Définition 4 (Temps moyen entre deux fausses alarmes)**.** *Le temps moyen $\bar{T}$ entre deux fausses alarmes est défini par la quantité suivante :*

$$\bar{T} = \mathbb{E}_{\theta_0}(t_a) \tag{5.7}$$

Définition 5 (Essentiel supremum)**.** *Soit $(Y_i)_{i \in I}$, une famille de v.a réelles sur $(\Omega, \mathcal{F}, \mathbb{P})$, bornée par une autre variable. On dit que Y est l'essentiel supremum et on écrit $Y = \operatorname{ess\,sup}_{i \in I} Y_i$ si*

- *$\forall i \in I$, $\mathbb{P}(Y_i \leq Y) = 1$*
- *pour toute v.a. réelle Z, $\mathbb{P}(Y_i \leq Y) = 1 \Leftrightarrow \mathbb{P}(Y \leq Z) = 1$*

Définition 6 (Délai moyen de détection conditionnel)**.** *Le délai moyen de détection conditionnel est défini par*

$$\mathbb{E}_{t_0}(t_a - t_0 + 1 | t_a \geq t_0, \mathcal{U}_1^{t_0 - 1}) \tag{5.8}$$

où $\mathbb{E}_{t_0}$ est l'espérance mathématique découlant de $\mathbb{P}_{t_0}^1$, la distribution des observations $u_1, \ldots, u_{t_0-1}$, $u_{t_0}, \ldots, u_{t_a}$, lorsque $t_0 = 1, 2 \ldots$.

Le délai moyen de détection conditionnel permet de donner un sens au pire retard moyen. Le pire retard moyen correspond à une combinaison d'un temps de changement t_0 et d'une trajectoire d'observations qui maximisent le délai moyen de détection conditionnel.

Définition 7 (Pire retard moyen [78])**.** *Le pire retard moyen $\overline{\tau}^*$ est défini par*

$$\overline{\tau}^* = \sup_{t_0 \geq 1} \operatorname{ess\,sup} \mathbb{E}_{t_0} \left[t_a - t_0 + 1 | t_a \geq t_0, \mathcal{U}_1^{t_0 - 1} \right] \tag{5.9}$$

En détection de changement en-ligne, l'analogue de la puissance du test est la fonction ARL (*Average Run Length*).

Définition 8 (Fonction ARL)**.** *On définit la fonction ARL comme*

$$\mathrm{ARL}(\theta) = \mathbb{E}_\theta(t_a) \tag{5.10}$$

La fonction ARL est directement liée à deux indicateurs de performances de l'algorithme de détection :

- Lorsque $\theta = \theta_0$, $\mathrm{ARL}(\theta_0) = \mathbb{E}_{\theta_0}(t_a)$ et la fonction ARL indique le temps moyen entre deux fausses alarmes. Idéalement, cette quantité doit être la plus grande possible.
- Lorsque $\theta = \theta_1$, $\mathrm{ARL}(\theta_1) = \mathbb{E}_{\theta_1}(t_a)$ et la fonction ARL indique le délai moyen de détection, que l'on souhaite le plus petit possible.

L'algorithme CUSUM est asymptotiquement optimal au sens où il minimise le pire retard moyen $\overline{\tau}^*$ lorsque le temps moyen entre deux fausses alarmes tend vers l'infini [78]. Un résultat non asymptotique de Moustakides [87] montre que le CUSUM minimise le pire retard moyen parmi les algorithmes dont le temps moyen $\bar{T}$ entre deux fausses alarmes vérifie $\bar{T} > \bar{T}_0$ où $\bar{T}_0$ est petit dans la plupart des cas.

5.3 Un aperçu sur les tests d'intégrité pour les filtres de Kalman

Rappelons que le filtre de Kalman est le filtre optimal pour le modèle linéaire gaussien défini par :

$$\left\{ \begin{array}{rcl} x_k & = & F_k x_{k-1} + w_k \\ y_k & = & H_k x_k + v_k \end{array} \right. \tag{5.11}$$

où w_k et v_k sont des v.a. décrivant les bruits de dynamique et de mesure, w_k et v_k sont des bruits blanc centrés de matrices de covariance respectives Q_k et R_k, indépendante de la condition initiale x_0. Enfin, F_k, B_k, H_k et D_k sont des matrices déterministes et connues. Le filtre de Kalman consiste à calculer récursivement le couple moyenne/covariance prédite $(\hat{x}_{k|k-1}, P_{k|k-1})$ ainsi que la moyenne et covariance corrigée $(\hat{x}_k, P_k)$, pour $k \geq 1$ (voir le chapitre 1, section 1.3.1.1).

5.3.1 Définition et causes de divergences pour le filtre de Kalman et ses variantes

Dans la littérature sur le filtre de Kalman, la divergence du filtre correspond au fait que la matrice de covariance de l'erreur d'estimation calculée par le filtre devient trop petite par rapport à la vraie matrice de covariance [40, 8]. Pour un modèle effectivement linéaire gaussien, le filtre de Kalman est le filtre optimal. Pour autant, la divergence peut survenir à cause :

- d'une méconnaissance des paramètres du modèle (statistiques des bruits de dynamique et de mesure erronées par exemple)
- d'erreurs numériques (inversion de matrices mal conditionnées, arrondis, troncature,...)

Les deux types d'erreurs peuvent faire que la matrice de covariance calculée par le filtre devient exagérément petite de sorte que les mesures reçues par le filtre n'influencent que marginalement l'estimation du filtre lors de l'étape de correction de Kalman.
Les études sur la divergence du filtre de Kalman distinguent la divergence apparente (*apparent divergence*) et la vraie divergence (*true divergence*) [40]. La divergence apparente correspond à un filtre dont la matrice de covariance estimée reste bornée mais dont l'erreur associée est très grande. La vraie divergence correspond à une matrice de covariance calculée par le filtre qui

augmente infiniment en norme [40].
Le filtre de Kalman étendu (EKF) et le filtre de Kalman sans parfum (UKF) sont des filtres qui procèdent pour le premier, par une linéarisation des équations du modèle et pour le second, par une approximation des moments a posteriori fondée sur une quadrature de Gauss. Dans le cas où les non-linéarités du système sont importantes au voisinage des états prédits et corrigés, les termes d'ordre 2 et plus ne sont pas toujours négligeables et l'EKF est susceptible de diverger. L'approximation de la moyenne et de la covariance a posteriori fournie par l'UKF est correcte à l'ordre 3 [63], mais cela ne garantit aucunement la stabilité du filtre lorsque la fonction est fortement non-linéaire.

5.3.2 Tests de détection de divergence en filtrage de Kalman

Les tests de détection de divergence utilisés pour les filtres de Kalman reposent essentiellement sur les propriétés statistiques du processus d'innovation $(\epsilon_k)_{k\geq 1}$ défini comme la différence entre la mesure reçue à l'instant k et la mesure prédite par le filtre $\hat{y}_{k|k-1}$:

$$\epsilon_k = y_k - \hat{y}_{k|k-1} \tag{5.12}$$

ce qui peut être ré-écrit sous la forme

$$\epsilon_k = y_k - H_k \hat{x}_{k|k-1} \tag{5.13}$$

L'innovation quantifie l'information apportée par l'observation y_k sur l'estimation de $\hat{x}_k$. En théorie, ϵ_k est un bruit blanc gaussien de matrice de covariance $\Sigma_k = H_k P_{k|k-1} H_k^T + R_k$. Ceci permet de définir des tests d'hypothèses basés sur la distribution de ϵ_k.
L'un des tests les plus couramment utilisés est basé sur le fait que la v.a. $r_k = \epsilon_k^T \Sigma_k^{-1} \epsilon_k$ suit une distribution du khi-deux à m degrés de libertés [8] où m est taille du vecteur des observations : une divergence peut être déclarée lorsque r_k dépasse un seuil q_α correspondant au quantile d'ordre $\alpha = 0.95,\, 0.99, \ldots$ de la loi du khi-deux à m degrés de libertés.
Des tests d'hypothèse vérifiant le caractère "blanc" de l'innovation ont également été proposés [84] : en effet, ϵ_k est théoriquement un bruit blanc gaussien. Ces tests de blancheur sont basés sur la fonction d'autocorrélation empirique de ϵ_k.
D'un point de vue pratique, les test de divergence précédents ont été largement appliqués, notamment en hybridation GPS/INS par filtrage de Kalman [82, 2, 86, 61].
Un critère similaire relatif à l'EKF a été proposé par [109]. Dans ce critère la mesure prédite $\hat{y}_{k|k-1}$ est essentiellement approchée par $h_k(\hat{x}_{k|k-1})$ où h_k est la fonction d'observation non-linéaire. [109] définit alors le processus normalisé η_k comme

$$\eta_k = \left(y_k - h_k(\hat{x}_{k|k-1}\right)^T S_k^{-1} \left(y_k - h_k(\hat{x}_{k|k-1}\right), \tag{5.14}$$

où S_k est la matrice approchée de l'innovation obtenue par linéarisation autour de l'état prédit.

$$S_k = \nabla^T h_k(\hat{x}_{k|k-1}) P_{k|k-1} \nabla h_k(\hat{x}_{k|k-1}) + R_k \tag{5.15}$$

La divergence peut alors être détectée de la même manière que pour un filtre de Kalman standard en comparant η_k à un seuil.

5.4 Détection de divergence en filtrage particulaire

Etat de l'art

Contrairement au filtre de Kalman, la problématique de détection de divergence en filtrage particulaire n'a pas fait l'objet de beaucoup d'études publiées. Une recherche bibliographique fait apparaitre les travaux de Giremus [49] et Turgut [116]. Les premiers portent sur le filtrage particulaire dans le cadre d'hybridation serrée GPS/INS et les seconds sur la localisation indoor à partir de signaux radio.
Dans [49], l'auteur propose en réalité un critère de dégénérescence du filtre basé sur la somme des poids non-normalisés qui permet d'enclencher une procédure de régularisation. Ce n'est donc pas à strictement parler un critère permettant de détecter la divergence du filtre au sens d'une incohérence irrémédiable entre le nuage de particules et l'état vrai. Il est avancé que la taille effective de l'échantillon N_{eff} est un indicateur peu fiable de la dégénérescence du nuage de particule lorsque toutes les particules ont un poids non normalisé proche d'une valeur commune très faible. Ceci arrive par exemple lorsque le bruit de dynamique est petit, de sorte que les particules occupent une région étroite de l'espace. Ainsi, les vraisemblances sont comparables et par ricochet les poids normalisés aussi. N_{eff} peut alors être proche du nombre total de particules N, indiquant (à tort) que l'ensemble des particules n'est pas dégénéré. Etant donné l'ensemble des poids non normalisés $\{\tilde{\omega}_k^i\}_{i=1}^N$, la statistique de test proposée dans [49] est la somme des poids non normalisés L_k :

$$L_k = \sum_{k=1}^{N} \tilde{\omega}_k^i \tag{5.16}$$

L_k représente en fait une approximation de la vraisemblance prédite $p(y_k|y_{0:k-1})$ (voir l'équation (2.72), section 2.6.2). Sur la base de simulations, Giremus montre que le processus L_k oscille autour d'une moyenne $< L_k >$ en cas de fonctionnement normal et est soumis à une brusque décroissance lorsque le filtre devient dégénéré. Ce changement de moyenne est détecté au moyen de l'algorithme CUSUM présenté précédemment. La particularité de ce test est qu'il nécessite d'estimer la moyenne en fonctionnement normal (hypothèse H_0) en ligne : ceci au fait en prenant la moyenne de L_k sur une fenêtre glissante. A ce stade, nous notons que cette méthode a été utilisée dans [49] comme critère de ré-échantillonnage et régularisation et non pas de divergence au sens où le nuage de particules s'est éloigné irrémédiablement de l'état vrai. Par ailleurs, la statistique proposée n'est pas normalisée ce qui peut rendre son application dans un autre contexte difficile, notamment pour ce qui est du réglage des seuils intervenant dans le test.

Dans [116], la méthode de détection proposée repose sur une évaluation de la divergence de Kullback-Leibler (2.21) $D_{KL}(p, \hat{p})$ entre la distribution empirique $\hat{p}$ donnée par le système de particules pondérées et la loi conditionnelle inconnue de densité p. Le cadre d'application étant celui de la localisation indoor d'une cible, l'état à estimer est de dimension deux (position horizontale). Pour évaluer $D_{KL}(p, \hat{p})$, les auteurs cherchent une approximation de p, indépendante du filtre. La loi conditionnelle inconnue est exprimée selon la loi de Bayes, comme le produit de la loi a priori et de la fonction de vraisemblance. La loi a priori est représentée par une loi uniforme sur la surface admissible (par exemple, les corridors du bâtiment), tandis que la fonction de

vraisemblance est représentée par une somme pondérée de fonctions constantes sur un domaine rectangulaire. Finalement, la constante de normalisation peut être calculée étant donné que le produit de la loi a priori et de la vraisemblance est « constant par morceaux », les morceaux étant des rectangles du plan. La densité a posteriori p est donc représentée par un histogramme bi-dimensionnel, dans lequel les rectangles sont de taille variable. Afin d'obtenir la divergence de Kullback-Leibler $D_{KL}(p, \hat{p})$, les auteurs utilisent une technique analogue à l'intégration de Riemann en deux dimensions. La divergence est détectée lorsque $D_{KL}(p, \hat{p})$ dépasse un seuil. L'objectif final est de réinitialiser le filtre une fois que la divergence est détectée. Les auteurs reportent des gains en erreur lorsque le détecteur de divergence est couplée au filtre particulaire, mais affirment que le seuil sur $D_{KL}(p, \hat{p})$ doit être réglé précautionneusement de manière à ne ré-initialiser le filtre que lorsque le filtre ne peut plus se recaler lui même. De plus cette méthode n'est applicable que pour un vecteur d'état de dimension inférieure ou égale à 2, étant donné la nature de l'approximation de la divergence de Kullback-Leibler, ce qui restreint son utilisation au recalage altimétrique au cas où l'altitude est supposée connue.

5.5 Formulation d'un test d'hypothèse pour la détection de divergence d'un filtre particulaire

Dans cette section, nous définissons un nouveau test d'hypothèse pour la détection de la divergence d'un filtre particulaire quelconque. Comme d'habitude nous considérons le système non-linéaire suivant :

$$\begin{cases} x_k = & f(x_{k-1}) + w_k \\ y_k = & h(x_k) + v_k \end{cases} \tag{5.17}$$

avec les hypothèses habituelles présentées au chapitre 2, section 2.2, auxquelles nous rajoutons les hypothèses suivantes :

- la densité $p(x_k|x_{k-1})$ du noyau de transition est continue par rapport à x_k et x_{k-1}
- la fonction d'observation h est continue et bornée sur $\mathbb{R}^d$.

Ces hypothèses ne sont pas très restrictives et sont vérifiées dans l'application qui nous intéresse, puisque la densité du noyau de transition est gaussienne et la fonction h correspond à la hauteur-sol qui est bien sûr bornée. Dans la suite de cette section, sauf mention contraire, nous supposerons par souci de simplicité que les observations $(y_k)_{k\geq 0}$ sont scalaires et que $(v_k)_k$ est une suite de v.a. gaussiennes indépendantes de variance σ_v^2.
Enfin, l'hypothèse nulle H_0 sera celle du fonctionnement normal du filtre, tandis que l'hypothèse alternative de divergence du filtre sera notée H_1.

En s'inspirant des résultats énoncés précédemment sur le filtre de Kalman, nous allons définir un test statistique basé sur la suite des innovations. L'innovation du filtre particulaire est ici définie comme

$$\epsilon_k^N = y_k - \hat{y}_{k|k-1}^N \tag{5.18}$$

où $\hat{y}^N_{k|k-1} = \sum_{i=1}^{N} \omega^i_{k-1} h(x^i_k)$ est la mesure prédite par le filtre.

Déterminer la loi de l'innovation pour une fonction d'observation h quelconque semble difficile dans le cas général. Afin de définir le test, nous proposons de calculer la moyenne et la variance de l'innovation ϵ^N_k sous l'hypothèse de fonctionnement nominal H_0 du filtre particulaire et conditionnellement aux mesures passées $y_0, \ldots, y_{k-1}$, en approchant ces deux quantités par leur limite lorsque le nombre de particules est infiniment grand.

5.5.1 Approximation de la moyenne de l'innovation particulaire sous l'hypothèse de fonctionnement normal du filtre

L'objectif de cette section est de déterminer une valeur approchée de la moyenne conditionnelle de l'innovation particulaire sous l'hypothèse de non-divergence H_0, que l'on notera $\mathbb{E}_{H_0}\left(\epsilon^N_k | y_{0:k-1}\right)$.

Décomposons l'innovation ϵ^N_k comme :

$$\begin{aligned} \epsilon^N_k &= y_k - \hat{y}^N_{k|k-1} \\ &= y_k - \hat{y}_{k|k-1} + \hat{y}_{k|k-1} - \hat{y}^N_{k|k-1} \\ &= \epsilon_k + \hat{y}_{k|k-1} - \hat{y}^N_{k|k-1} \end{aligned} \tag{5.19}$$

où $\hat{y}_{k|k-1} = \mathbb{E}\left(y_k | y_{0:k-1}\right) = \int h(x_k) p(x_k | y_{0:k-1})\, dx_k$ est l'observation prédite par le filtre optimal et $\epsilon_k = y_k - \hat{y}_{k|k-1}$ est l'innovation correspondante. On a,

$$\mathbb{E}\left(\epsilon_k | y_{0:k-1}\right) = 0 \tag{5.20}$$

Soit ζ^N_k la différence entre la mesure prédite par le filtre optimal et celle prédite par le filtre particulaire.

$$\zeta^N_k = \hat{y}_{k|k-1} - \hat{y}^N_{k|k-1} \tag{5.21}$$

On a :

$$\begin{aligned} \hat{y}_{k|k-1} &= \int h(x_k) p(x_k | y_{0:k-1})\, \mathrm{d}x_k \\ &= \iint h(x_k) p(x_k | x_{k-1}) p(x_{k-1} | y_{0:k-1})\, \mathrm{d}x_k \mathrm{d}x_{k-1} \\ &= \int \psi_{k-1}\left(x_{k-1}\right) p(x_{k-1} | y_{0:k-1})\, \mathrm{d}x_{k-1} \end{aligned} \tag{5.22}$$

où la fonction ψ_{k-1} est définie par $\psi_{k-1}\left(x\right) = \int h(x_k) p(x_k | x_{k-1} = x)\, \mathrm{d}x_k$

ζ^N_k se ré-écrit donc sous la forme

$$\zeta^N_k = \int h(x_k) \left[p(x_k | y_{0:k-1}) - \hat{p}(x_k | y_{0:k-1})\right] \mathrm{d}x_k \tag{5.23}$$

$$= \int \psi_{k-1}\left(x_{k-1}\right) \left[p(x_{k-1} | y_{0:k-1}) - \hat{p}(x_{k-1} | y_{0:k-1})\right] \mathrm{d}x_{k-1} \tag{5.24}$$

$$= \langle \psi_{k-1}, p_{k-1} - \hat{p}_{k-1} \rangle \tag{5.25}$$

où $\hat{p}_{k-1} = \sum_{i=1}^{N} \omega_{k-1} \delta_{x_{k-1}^i}$ est l'approximation particulaire de $p_{k-1} = p(x_{k-1}|y_{0:k-1})$.
ψ_{k-1} est continue d'après le théorème de continuité sous le signe somme étant donné que par hypothèse, l'intégrande $x_{k-1} \mapsto h(x_k)p(x_k|x_{k-1})$ est continue. ψ_{k-1} est également bornée car

$$\begin{aligned}\psi_{k-1}(x) &= \int h(x_k)p(x_k|x_{k-1}=x)\,\mathrm{d}x_k \\ &\leq \|h\|_\infty \int p(x_k|x_{k-1}=x)\,\mathrm{d}x_k = \|h\|_\infty\end{aligned}$$

D'après le théorème 1 de convergence faible,

$$\zeta_k^N = \langle \psi_{k-1}, p_{k-1} - \hat{p}_{k-1} \rangle \xrightarrow[N\to+\infty]{} 0 \quad p.s. \tag{5.26}$$

Finalement, comme $\epsilon_k^N = \epsilon_k + \zeta_k^N$, il vient

$$\begin{aligned}\mathbb{E}\left(\epsilon_k^N|y_{0:k}\right) &= \mathbb{E}(\epsilon_k|y_{0:k-1}) + \mathbb{E}(\zeta_k^N|y_{0:k-1}) \\ &= 0 + \zeta_k^N \xrightarrow[N\to+\infty]{} 0 \quad p.s.\end{aligned}$$

où l'égalité $\mathbb{E}(\zeta_k^N|y_{0:k-1}) = \zeta_k^N$ vient du fait que ζ_k^N est la différence de deux espérances conditionnellement à $y_{0:k-1}$.

En résumé, comme nous sommes sous l'hypothèse de fonctionnement normal du filtre, nous proposons d'approcher la moyenne conditionnelle de l'innovation du filtre par sa limite lorsque le nombre de particules est infiniment grand, c'est-à-dire,

$$\mathbb{E}_{H_0}\left(\epsilon_k^N|y_{0:k-1}\right) \approx 0 \tag{5.27}$$

5.5.2 Approximation de la variance de l'innovation particulaire sous l'hypothèse de fonctionnement normal du filtre

Le calcul de la variance de ϵ_k^N, conditionnellement à $y_{0:k-1}$, se fait en reprenant la décomposition de l'innovation sous la forme $\epsilon_k^N = \epsilon_k + \zeta_k^N$. On a donc

$$\begin{aligned}\sigma_\epsilon^2(k) &= \mathrm{Var}\left(\epsilon_k^N|y_{0:k-1}\right) \qquad (5.28)\\ &= \mathrm{Var}\left(\epsilon_k + \zeta_k^N|y_{0:k-1}\right) \\ &= \mathrm{Var}\left(\epsilon_k|y_{0:k-1}\right) + \mathrm{Var}\left(\zeta_k^N|y_{0:k-1}\right) \qquad (5.29)\end{aligned}$$

Puisque ϵ_k ne dépend que de y_k et ζ_k^N ne dépend que des particules prédites, ϵ_k et ζ_k^N sont indépendantes conditionnellement à $y_{0:k-1}$. En effet, v_k est supposée indépendante de x_k et de $y_{0:k-1}$.

Cherchons à présent une approximation pour $\mathrm{Var}\left(\zeta_k^N|y_{0:k-1}\right)$. D'après (5.25), on a

$$\zeta_k^N = \int \psi_{k-1}(x_{k-1})\, p(x_{k-1}|y_{0:k-1})\,\mathrm{d}x_{k-1} - \sum_{i=1}^{N} \omega_{k-1}^i \psi_{k-1}\left(x_{k-1}^i\right) \tag{5.30}$$

Le théorème suivant [31] est utile pour calculer la variance de ζ_k^N.

Théorème 5 (Théorème central limite (TCL) pour l'algorithme SIS)**.** *Soit* $\{x_k^i, \omega_k^i \; i = 1, \ldots, N\}_k$ *la suite de particules et poids obtenus avec l'algorithme SIS. Alors pour toute fonction continue bornée* ϕ*, la variable*

$$\sqrt{N}\left(\int \phi\left(x_k\right) p(x_k|y_{0:k})\,\mathrm{d}x_k - \sum_{i=1}^{N} \omega_k^i \phi\left(x_k^i\right)\right)$$

converge en loi vers une loi gaussienne $\mathcal{N}\left(0, v_k\left(\phi\right)\right)$ *lorsque* $N \to \infty$ *La variance asymptotique* $V_k\left(\phi\right)$ *s'exprime comme*

$$V_k\left(\phi\right) = \frac{\int \left(\phi\left(x_k\right) - \bar{\phi}_k\right)^2 g_{0:k}^2(x_{0:k}) p(x_{0:k})\,\mathrm{d}x_{0:k}}{\left(\int g_{0:k}(x_{0:k}) p(x_{0:k})\,\mathrm{d}x_k\right)^2} \tag{5.31}$$

où $g_{0:k}(x_{0:k}) = \prod_{p=0}^{k} g_p(x_p)$*,* g_k *étant la fonction de vraisemblance à l'instant* k*, et* $\bar{\phi}_k = \int \phi\left(x_k\right) p(x_k|y_{0:k})\,\mathrm{d}x_k$.

Sous l'hypothèse H_0, pour N suffisamment grand, $\mathrm{Var}\left(\zeta_k^N\right)$ peut être approchée en considérant $\phi = \psi_{k-1}$ dans le TCL :

$$\mathrm{Var}_{H_0}\left(\zeta_k^N | y_{0:k-1}\right) \approx \frac{V_{k-1}(\psi_{k-1})}{N} \tag{5.32}$$

Le calcul du terme $V_{k-1}(\psi_{k-1})$ peut s'effectuer au moyen du système de particules comme dans [90]. En raison du ré-échantillonnage, $V_{k-1}(\psi_{k-1})$ ne peut pas être obtenu directement par l'application du TCL. Cependant, le *SIR* peut être vu comme un *SIS* entre 2 instants successifs de ré-échantillonnage. Par conséquent, $V_{k-1}(\psi_{k-1})$ peut être calculé via l'équation (5.31) en accumulant les observations entre 2 instants de ré-échantillonnage.
Ainsi avant la première étape de ré-échantillonnage,

$$V_{k-1}(\psi_{k-1}) \approx \frac{N \sum_{i=1}^{N} (\psi_{k-1}(x_{k-1}^i) - \bar{\psi}_{k-1})^2 g_{0:k-1}^2(x_{0:k-1}^i)}{\left(\sum_{i=1}^{N} g_{0:k-1}(x_{0:k-1}^i)\right)^2}$$

avec $\psi_{k-1}(x_{k-1}^i) = \int h(x_k) p(x_k|x_{k-1}^i)\,\mathrm{d}x_k$ et $\bar{\psi}_{k-1} = \int \psi_{k-1}\left(x_{k-1}\right) p(x_{k-1}|y_{0:k-1})\,\mathrm{d}x_{k-1}$. On a $\bar{\psi}_{k-1} \approx \sum_{i=1}^{N} \omega_{k-1}^i \psi_{k-1}\left(x_{k-1}^i\right)$.

Posons $\hat{h}_{k|k-1} = \sum_{i=1}^{N} \omega_{k-1}^i h(x_k^i)$. D'après (5.22), $\sum_{i=1}^{N} \omega_{k-1}^i \psi_{k-1}\left(x_{k-1}^i\right) = \hat{h}_{k|k-1}$. La variance

asymptotique $V_{k-1}(\psi_{k-1})$ s'exprime donc comme suit :

$$V_{k-1}(\psi_{k-1}) \approx \frac{N\sum_{i=1}^{N}(h(x_k^i)-\hat{h}_{k|k-1})^2 g_{0:k-1}^2(x_{0:k-1}^i)}{\left(\sum_{i=1}^{N} g_{0:k-1}(x_{0:k-1}^i)\right)^2} \tag{5.33}$$

Notons k_0 le dernier instant de ré-échantillonnage avant l'instant k. Alors $V_{k-1}(\psi_{k-1})$ s'écrit :

$$V_{k-1}(\psi_{k-1}) \approx \frac{N\sum_{i=1}^{N}(h(x_k^i)-\hat{h}_{k|k-1})^2 g_{k_0:k-1}^2(x_{k_0:k-1}^i)}{\left(\sum_{i=1}^{N} g_{k_0:k-1}(x_{k_0:k-1}^i)\right)^2} \tag{5.34}$$

On a donc

$$\sigma_\epsilon^2(k) \approx \mathrm{Var}\left(\epsilon_k|y_{0:k-1}\right) + \frac{V_{k-1}(\psi_{k-1})}{N} \tag{5.35}$$

Dans la variance conditionnelle de l'innovation ϵ_k intervient également le terme $\mathrm{Var}\left(h(x_k)|y_{0:k-1}\right)$. Etant donné la nature du modèle (5.17),

$$\begin{aligned}\mathrm{Var}\left(\epsilon_k|y_{0:k-1}\right) &= \mathrm{Var}\left(v_k + h(x_k) - \mathbb{E}\left(h(x_k)|y_{0:k-1}\right)|y_{0:k-1}\right)\\ &= \sigma_v^2 + \mathrm{Var}\left(h(x_k) - \mathbb{E}\left(h(x_k)|y_{0:k-1}\right)|y_{0:k-1}\right)\\ &= \sigma_v^2 + \mathrm{Var}\left(h(x_k)|y_{0:k-1}\right)\end{aligned}$$

Sous l'hypothèse H_0 de fonctionnement nominal, le système poids-particules permet d'approcher la quantité $\mathrm{Var}\left(h(x_k)|y_{0:k-1}\right)$.

$$\mathrm{Var}\left(h(x_k)|y_{0:k-1}\right) \approx \sum_{i=1}^{N} w_{k-1}^i \left(h(x_k^i) - \hat{h}_{k|k-1}\right)^2$$

En effet,

$$\begin{aligned}&\mathrm{Var}\left(h(x_k)|y_{0:k-1}\right) - \sum_{i=1}^{N} w_{k-1}^i \left(h(x_k^i) - \hat{h}_{k|k-1}\right)^2\\ &= \langle h^2, p_{k|k-1}\rangle - \langle h, p_{k|k-1}\rangle^2 + \langle h^2, \hat{p}_{k|k-1}\rangle - \langle h, \hat{p}_{k|k-1}\rangle^2\\ &= \langle h^2, p_{k|k-1} - \hat{p}_{k|k-1}\rangle + \left[\langle h, p_{k|k-1}\rangle - \langle h, \hat{p}_{k|k-1}\rangle\right]\left[\langle h, p_{k|k-1}\rangle + \langle h, \hat{p}_{k|k-1}\rangle\right]\\ &= \langle h^2, p_{k|k-1} - \hat{p}_{k|k-1}\rangle + \langle h, p_{k|k-1} - \hat{p}_{k|k-1}\rangle\langle h, p_{k|k-1} + \hat{p}_{k|k-1}\rangle\end{aligned}$$

h étant continue bornée, les quantités $\langle h^2, p_{k|k-1} - \hat{p}_{k|k-1}\rangle$ et $\langle h, p_{k|k-1} - \hat{p}_{k|k-1}\rangle$ convergent presque sûrement vers 0 lorsque N tend vers $+\infty$ en vertu de 1, et donc $\sum_{i=1}^{N} w_{k-1}^i \left(h(x_k^i) - \hat{h}_{k|k-1}\right)^2$

tend p.s. vers $\mathrm{Var}\left(h(x_k)|y_{0:k-1}\right)$.

Finalement, de même que pour la moyenne conditionnelle, nous proposons d'approcher la variance conditionnelle de l'innovation du filtre particulaire sous H_0 par une approximation obtenue lorsque le nombre de particule tend vers $+\infty$:

$$\mathrm{Var}_{H_0}\left(\epsilon_k^N|y_{0:k-1}\right) \triangleq \sigma_\epsilon^2(k) \approx \sigma_v^2 + \sum_{i=1}^{N} w_{k-1}^i \left(h(x_k^i) - \hat{h}_{k|k-1}\right)^2 + \frac{V_{k-1}(\psi_{k-1})}{N} \tag{5.36}$$

5.5.3 Cas d'un vecteur de mesure multidimensionnel

Les calculs précédents se généralisent au cas où y est de dimension $m \geq 2$

$$\mathbb{E}_{H_0}\left(\epsilon_k^N|y_{0:k-1}\right) = 0 \tag{5.37}$$

$$\mathrm{Cov}_{H_0}\left(\epsilon_k^N|y_{0:k-1}\right) \triangleq \Sigma_{\epsilon,k} \approx R_k + \sum_{i=1}^{N} w_{k-1}^i \left(h(x_k^i) - \hat{h}_{k|k-1}\right)\left(h(x_k^i) - \hat{h}_{k|k-1}\right)^T \tag{5.38}$$

$$+ \frac{V_{k-1}(\psi_{k-1})}{N} \tag{5.39}$$

Une application de la détection de divergence dans le cas où la mesure est multidimensionnelle est donné en annexe D.1.

5.5.4 Définition du test d'hypothèse

Sur la base des résultats précédents, on peut définir l'hypothèse nulle comme

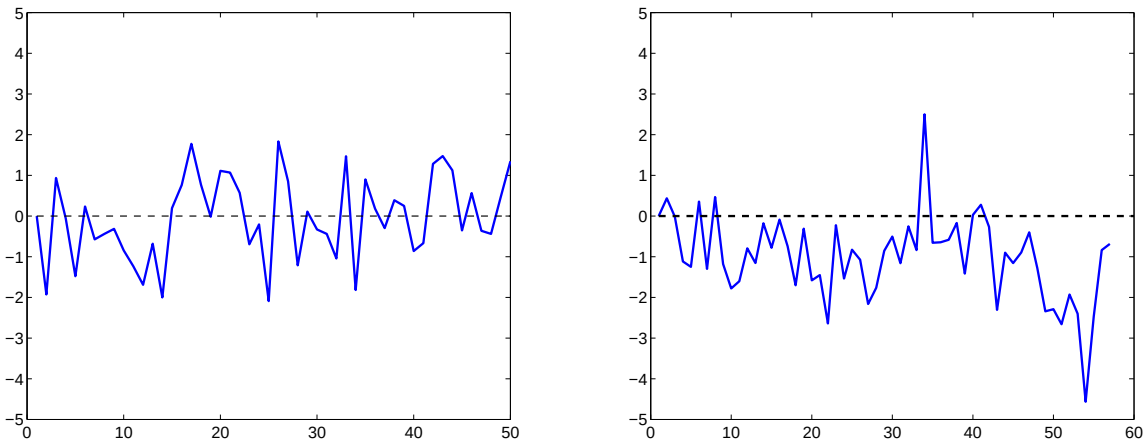
$$\begin{cases} \mathbb{E}_{H_0}\left(\epsilon_k^N|y_{0:k-1}\right) = 0 \\ \mathrm{Var}_{H_0}\left(\epsilon_k^N|y_{0:k-1}\right) = \sigma_\epsilon^2(k) \end{cases} \tag{5.40}$$

La première étape est de considérer le processus d'innovation normalisé défini par :

$$r_k = \frac{\epsilon_k^N}{\sigma_\epsilon(k)} \tag{5.41}$$

qui, conditionnellement à $y_{0:k-1}$, est de moyenne nulle et de variance unité sous l'hypothèse H_0. Posons $\theta = \mathbb{E}\left(r_k|y_{0:k-1}\right)$. Le test proposé consiste à tester $H_0 : \theta = 0$ contre $H_1 : \theta \neq 0$. Etant donné que nous n'avons pas accès à toute la distribution du résidu r_k mais seulement à une approximation des deux premiers moments, le test du rapport de vraisemblance généralisé [4, p. 121] est exclu. Nous proposons plutôt de définir l'hypothèse alternative comme un changement minimum dans la moyenne du résidu r_k, i.e.,

$$\begin{cases} \left|\mathbb{E}_{H_1}\left(r_k|\mathcal{Y}^{k-1}\right)\right| \geq \nu \\ \mathrm{Var}_{H_1}\left(r_k|\mathcal{Y}^{k-1}\right) = 1 \end{cases} \tag{5.42}$$



(a) Cas convergent (b) Cas divergent

Figure 5.1 – Comportement de l'innovation normalisée d'un filtre particulaire

En toute rigueur, la variance $\mathrm{Var}_{H_1}\left(r_k|\mathcal{Y}^{k-1}\right)$ est inconnue et dépend du type de dysfonctionnement de l'algorithme particulaire, mais on supposera qu'elle reste unitaire après divergence. Cette formulation nous permet entre autre d'utiliser le test CUSUM introduit précédemment.

5.6 Mise en oeuvre d'un test séquentiel pour la détection de divergence

5.6.1 Détection en ligne du changement de moyenne du processus d'innovation normalisé

Etant donné la formulation précédente du test d'hypothèse, nous devons détecter au plus tôt l'instant de changement t_0 qui vérifie

$$\begin{cases} \mathbb{E}\left(r_k|y_{0:k-1}\right) &= 0 \quad \text{si } k < t_0 \\ \left|\mathbb{E}\left(r_k|y_{0:k-1}\right)\right| &\geq \nu \quad \text{si } k \geq t_0 \end{cases} \tag{5.43}$$

où r_k est le du processus d'innovation normalisé.
Ce problème est analogue à la détection séquentielle d'une déviation de la moyenne dans une suite de variables aléatoires gaussiennes i.i.d $u_1, u_2, \ldots, u_k$ vérifiant :

$$\begin{cases} \mathrm{Var}(u_k) &= \sigma^2 \quad \forall k \\ m \triangleq \mathbb{E}(u_k) &= 0 \quad \text{for } k \leq t_0 \\ |m| &= \nu \quad \text{for } k \geq t_0 \end{cases} \tag{5.44}$$

Afin de détecter une éventuelle divergence du filtre, nous utilisons un test CUSUM bilatéral (*two-sided CUSUM*) [4] pour détecter à la fois les augmentations et les diminutions de la moyenne de r_k d'amplitude supérieure ou égale à ν. Pour la séquence $u_1, u_2, \ldots, u_k$, le temps d'arrêt du CUSUM bilatéral est défini par :

$$t_a^{TS} = \inf\left\{k \geq 1 \ : (D_k^+ \geq h) \text{ou } (D_k^- \geq h)\right\} \tag{5.45}$$

où D_k^+ est la fonction de décision correspondant au test d'une augmentation de la moyenne, D_k^- est la fonction de décision correspondant au test d'une diminution de la moyenne et h est un seuil de détection.

$$D_k^+ = \left(D_{k-1}^+ + \frac{\nu}{\sigma^2}\left(u_k - \frac{\nu}{2}\right)\right)^+ \tag{5.46}$$

$$D_k^- = \left(D_{k-1}^- + \frac{\nu}{\sigma^2}\left(-u_k - \frac{\nu}{2}\right)\right)^+ \tag{5.47}$$

avec $D_0^+ = D_0^- = 0$.

D'après [78], le CUSUM est optimal pour le critère du pire délai moyen $\overline{\tau}^*$. Le choix des paramètres h et ν influence fortement $\overline{\tau}^*$ et le temps moyen entre deux fausses alarmes $\bar{T}$. Nous avons rappelé (5.2.3) que le pire délai moyen de détection $\overline{\tau}^*$ et le temps moyen entre deux fausses alarmes $\bar{T}$ pouvaient tous les deux s'exprimer au moyen de la fonction $m \mapsto \mathrm{ARL}(m) = \mathbb{E}_m(t_a)$:

$$\bar{T} = \mathrm{ARL}(0) \tag{5.48}$$

$$\overline{\tau}^* = \mathrm{ARL}(\nu) \tag{5.49}$$

Pour un test CUSUM bilatéral symétrique, la fonction ARL, notée ARL_{TS}, peut s'exprimer grâce aux fonctions ARL des tests CUSUM unilatéraux dont les fonctions de décisions sont D_k^+ et D_k^-.

$$\frac{1}{\mathrm{ARL}_{TS}(\theta)} = \frac{1}{\mathrm{ARL}_-(\theta)} + \frac{1}{\mathrm{ARL}_+(\theta)}, \tag{5.50}$$

où $\mathrm{ARL}_+(\theta)$ (resp. $\mathrm{ARL}_-(\theta)$) est la fonction ARL du test *CUSUM* de fonction de décision D_k^+ (resp. D_k^-) défini par (5.46) (resp. (5.47)).

La fonction ARL d'un test CUSUM unilatéral vérifie [4] :

$$\mathrm{ARL}(\theta) = \frac{N_\theta(0)}{1 - \mathbb{P}_\theta(0)}, \tag{5.51}$$

où θ est le paramètre définissant la densité de probabilité du logarithme du rapport de vraisemblance z_k correspondant à une observation. Les u_k étant supposés gaussiens,

$$z_k = \frac{\nu}{\sigma^2}\left(u_k - \frac{\nu}{2}\right)$$

Les fonctions $z \mapsto N_\theta(z)$ et $z \mapsto \mathbb{P}_\theta(z)$ sont définies par les intégrales de Fredholm de deuxième espèce [4] :

$$\mathbb{P}_\theta(z) = \int_{-\infty}^{-z} f_\theta(x)dx + \int_0^h \mathbb{P}_\theta(x) f_\theta(x-z)dx \quad (5.52)$$

$$N_\theta(z) = 1 + \int_0^h N_\theta(x) f_\theta(x-z)dx \quad (5.53)$$

avec $0 \leq z \leq h$, et où la fonction f_θ est appelée noyau de l'équation intégrale. Dans le cadre de la détection de changement f_θ correspond à la densité de probabilité du logarithme du rapport de vraisemblance z_k. Etant donné que z_k est gaussien, f_θ sera une densité gaussienne de paramètre $\theta = (\mathbb{E}(z_k), \mathrm{Var}(z_k))^T$

$$\mathbb{E}(z_k) = \frac{\nu}{\sigma^2}(m - \frac{\nu}{2})$$
$$\mathrm{Var}(z_k) = \frac{\nu^2}{\sigma^2}.$$

La méthode couramment utilisée pour résoudre numériquement (5.52) et (5.53) consiste à remplacer les équations intégrales par un système d'équations algébriques en approchant les intégrales par quadrature de Gauss.
En guise d'illustration, nous avons étudié l'influence du seuil h sur $\overline{\tau}^*$ et $\bar{T}$ pour le test défini par (5.45). Les valeurs considérées ici sont $\sigma^2 = 1$ et $\nu = 3$. Pour un seuil h fixé, nous notons $m \mapsto \mathrm{ARL}(m; h)$ la fonction ARL correspondante. Les valeurs $\mathrm{ARL}(-3 + \frac{i}{2}; h)$ ont été calculées pour $i = 0, 1, \ldots, 6$. Les résultats sont illustrés sur la figure 5.2. Comme on peut s'y attendre, le temps moyen entre deux fausses alarmes $\bar{T}_h = \mathrm{ARL}(0; h)$ dépend fortement du seuil h ; il est d'autant plus important que h est grand. Le délai moyen de détection $h \mapsto \mathrm{ARL}(m; h)$ for $|m| \geq \nu$ croit également avec h mais à une vitesse plus lente. Nous confirmerons ce dernier point lors des simulations expérimentales.

5.6.2 Choix des paramètres

Jusqu'ici, nous avons supposé que les paramètres ν (saut de moyenne de l'innovation normalisée r_k) et σ^2 (variance de r_k) étaient entièrement connus. Nous avons vu que $\sigma^2 \approx 1$ sous l'hypothèse H_0 de non-divergence. La véritable variance après divergence étant inconnue, nous l'avons supposée égale à 1 sous H_1. De même, le saut ν n'est pas connu a priori mais dépend du problème de détection auquel on est confronté. On peut par exemple choisir $\nu = 3\sigma$ en partant du fait que si r_k était gaussien et centré $\mathbb{P}(|r_k| < 3\sigma) = 0.997$: l'objectif est alors de détecter un changement de moyenne situé en dehors de l'intervalle de confiance à 99.7% $[-3\sigma, 3\sigma]$. Une fois le paramètre ν, le seuil h peut être choisi grâce à la fonction ARL : on choisit le seuil h tel que le temps moyen entre deux fausses alarmes $\bar{T}_{min} = \mathrm{ARL}(0; h)$ vérifie $\mathrm{ARL}(0; h) \geq \bar{T}_{min}$ où $\bar{T}_{min}$ est défini par l'utilisateur.

Sensibilité du CUSUM à une erreur sur les paramètres

Un aspect important en détection de changement est la sensibilité de l'algorithme de détection

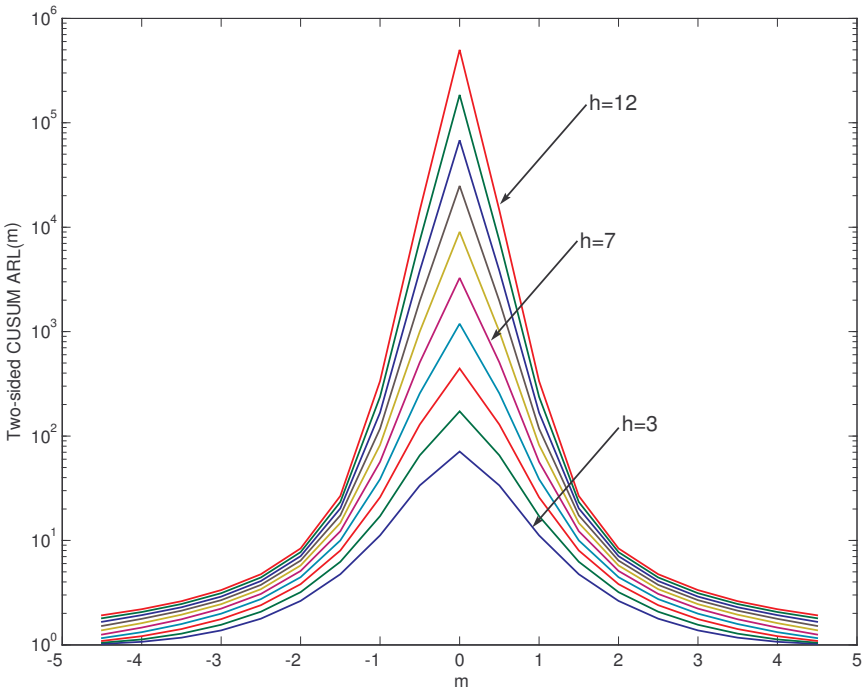


FIGURE 5.2 – Famille de fonctions ARL $m \mapsto \mathrm{ARL}(m;h)$ pour $h = \{3, 4, \ldots, 12\}$.

à une erreur sur l'un des paramètres. En l'occurence, le test de divergence s'appuie sur les valeurs nominales de la moyenne $m = 0$ et de la variance $\sigma^2 = 1$ du processus d'innovation normalisé r_k en l'absence de divergence. Ces valeurs ont été obtenues en remplaçant toutes les quantités faisant intervenir le filtre optimal et incalculables directement, par leurs équivalents particulaires, dont on suppose qu'ils fournissent de bonnes approximations dans l'hypothèse de fonctionnement normal du filtre. Or, nous avons fait comme si les valeurs exactes étaient utilisées dans le test de détection. A titre d'exemple, étudions l'influence d'une erreur sur σ en examinant la fonction ARL. On défini le ratio $R = \frac{\sigma_t^2}{\sigma^2}$ entre la vraie valeur σ_t^2 de la variance de r et sa valeur supposée. Les fonctions ARL $m \mapsto \mathrm{ARL}(m;R)$ sont calculées pour $R = 0.5, 0.6, \ldots, 1.5$ et $h = 6$ (cf. figure 5.3). On constate que le temps moyen avant une fausse alarme est très sensible à la valeur de R. En revanche, le retard moyen de détection est peu sensible à R.

5.6.3 Evaluation des performances du détecteur de divergence dans le cadre du recalage radio-altimétrique

L'objectif de cette section est d'évaluer la capacité de l'algorithme de détection présenté précédemment à repérer au plus tôt la divergence d'un filtre de navigation. Notons que le surcoût algorithmique de la module de détection est très modeste étant donné que le calcul le plus

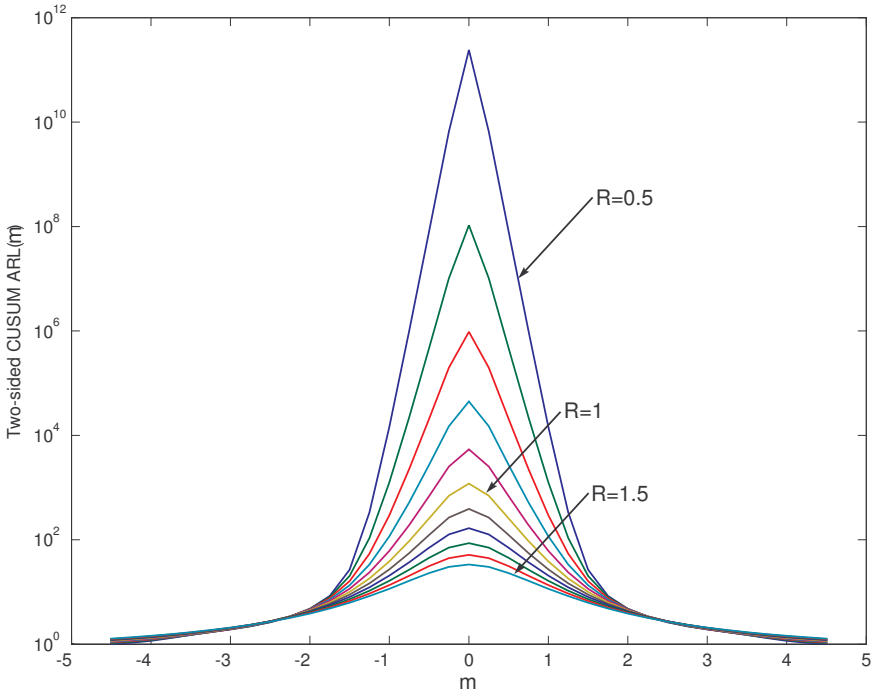


Figure 5.3 – Fonctions $m \mapsto \mathrm{ARL}(m)$ pour différents ratios $R = \{0.5, 0.6, \ldots, 1.5\}$.

complexe est celui de $\mathrm{Var}(h(x_k)|y_{0:k-1})$ qui se fait avec le système poids-particules $\left\{\omega_{k-1}^i, x_k^i\right\}_{i=1}^N$ dont dispose le filtre à l'instant k.
Notre étude se base sur trois principaux indicateurs de performances :

- Le taux de fausses alarmes TFA : il est défini comme la fréquence à laquelle le test CUSUM bilatéral déclenche une alarme à un instant $t_a^{TS} < t_0$, où t_0 est le vrai instant d'une éventuelle divergence, avec la convention $t_0 = \infty$ lorsque le filtre n'a pas divergé sur toute la durée de la simulation.
- le taux de non-détection TND : il s'agit de la fréquence avec laquelle l'algorithme de détection ne déclenche aucune alarme pendant toute la durée de la simulation alors que le filtre a réellement divergé.
- le délai de détection moyen $\tau = \mathbb{E}_{t_0}(t_a^{TS} - t_0 + 1 \mid t_a^{TS} \geq t_0)$

Les trois indicateurs seront évalués sur les trois scénarios de navigation définis au chapitre 4, section 4.6.1. La démarche consiste tout d'abord à simuler un nombre N_{MC} de trajectoires réelles de durée finie n. Pour chaque trajectoire, on simule également la suite d'observations correspondante selon l'équation de mesure du radio-altimètre. Enfin, un algorithme particulaire doté du détecteur de divergence est enclenché pour chaque couple trajectoire/observation : ceci donne lieu aux instants d'alarmes $t_{a,i}^{TS}$, $i = 1, \ldots, N_{MC}$ et aux vrais instants de divergences $t_{0,i}$, $i = 1, \ldots, N_{MC}$.

On note N_{FA} le nombre total de fausses alarmes et N_{ND} le nombre de divergences non détectées. Les taux de fausses alarmes et de non-détection s'expriment alors simplement comme :

$$\mathrm{TFA} = \frac{N_{FA}}{N_{MC}} \tag{5.54}$$

$$\mathrm{TND} = \frac{N_{ND}}{N_{TD}} \tag{5.55}$$

où N_{TD} désigne le nombre total de divergences réelles.
Le délai de détection moyen est calculé selon :

$$\tau = \frac{\sum_{i=1}^{N_{MC}} (t_{a,i}^{TS} - t_{0,i} + 1) \mathbb{1}_{\{t_{a,i}^{TS} \geq t_{0,i}\}}}{\sum_{i=1}^{N_{MC}} \mathbb{1}_{\{t_{a,i}^{TS} \geq t_{0,i}\}}} \tag{5.56}$$

Afin de calculer ces quantités efficacement, le vrai instant de divergence t_0 doit être identifié automatiquement. Pour cela, nous avons utilisé le MRPF car il permet d'identifier aisément l'instant de décrochage du filtre pour les raisons qui sont exposée ci-après.
On suppose qu'à l'instant k, la densité a posteriori empirique $\hat{p}_{k|k}$ se décompose en M_k composantes unimodales

$$\hat{p}_{k|k} = \sum_{j=1}^{M_k} \alpha_{j,k} \hat{p}_j(x_k|y_{0:k}) \tag{5.57}$$

Rappelons que chaque composante $\hat{p}_j(x_k|y_{0:k})$ est représenté par un cluster de particules, pour lequel nous pouvons calculer la moyenne conditionnelle intra-cluster $\hat{x}_k^j \approx \int x_k p_j(x_k|y_{0:k}) dx_k$ et la matrice de covariance empirique $\hat{\Sigma}_k^j \approx \int (x_k - \hat{x}_k^j)(x_k - \hat{x}_k^j)^T p_j(x_k|y_{0:k}) dx_k$. Nous dirons alors que le filtre a divergé à l'instant t_0 si : pour $j = 1, \ldots, M_k$ aucune des ellipsoïdes de confiance de

niveau β, associée à la distribution gaussienne de moyenne $\hat{x}_k^j$ et de covariance $\hat{\Sigma}_k^j$, ne contient l'état vrai x_k, ceci pendant L instants consécutifs, i.e. pour tout $k = t_0, t_0+1, \ldots, t_0+L-1$. L'introduction d'une fenêtre temporelle de taille L permet d'avoir un critère représentatif d'une "vraie" divergence dans la mesure où il est possible qu'aucune ellipse ne contienne l'état vrai à un instant t donné sans pour autant que cela signifie qu'il y ait une dérive irrémédiable de l'estimée particulaire. Dans les simulations numériques, nous avons fixé $L = 5$ et un niveau de confiance $\beta = 0.999$. Enfin le nombre de réalisations Monte Carlo a été fixé à $N_{MC} = 250$.

Paramètres du modèle

- covariance du bruit de dynamique $Q = \mathrm{diag}\begin{pmatrix} 1 & 1 & 0.01^2 \end{pmatrix}$
- écart-type du bruit de mesure du radio-altimètre $\sigma_v = 10\,m$
- période d'échantillonnage $\Delta = 0.1\,s$
- Loi initiale : $x_0 \sim \mathcal{N}(0_{6\times1}, P_0)$, avec $P_0 = \mathrm{diag}(1000^2,\, 1000^2,\, 100^2,\, 3^2,\, 3^2,\, 1^2)$

Le filtre utilisé est le MRPF avec $N_{MRPF} = 3000$ particules. Afin d'obtenir un nombre comparable de trajectoires convergentes et divergentes, le seuil de suppression de cluster α_{min} (cf. 4.3) a été fixé à une valeur artificiellement élevée $\alpha_{min} = 10^{-2}$, augmentant la probabilité de suppression du cluster correspondant au mode vrai.

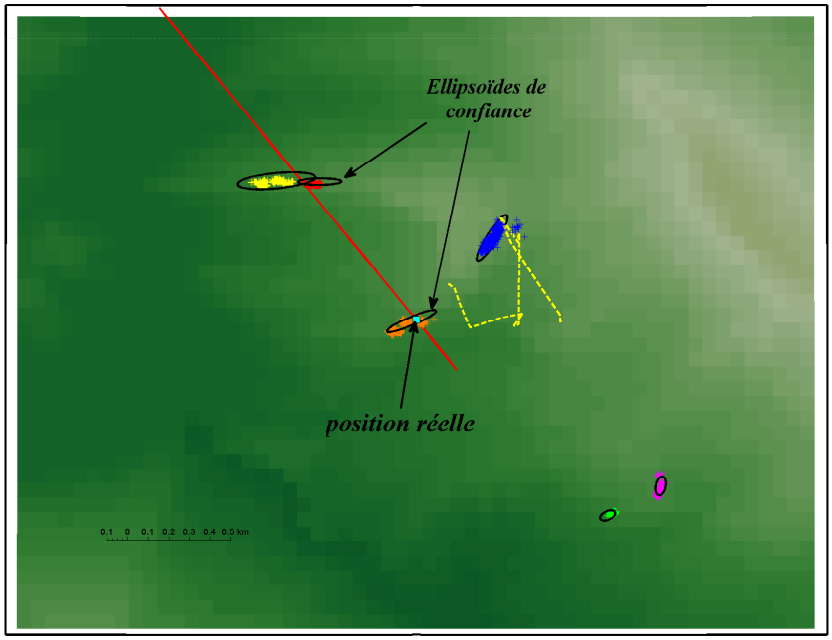
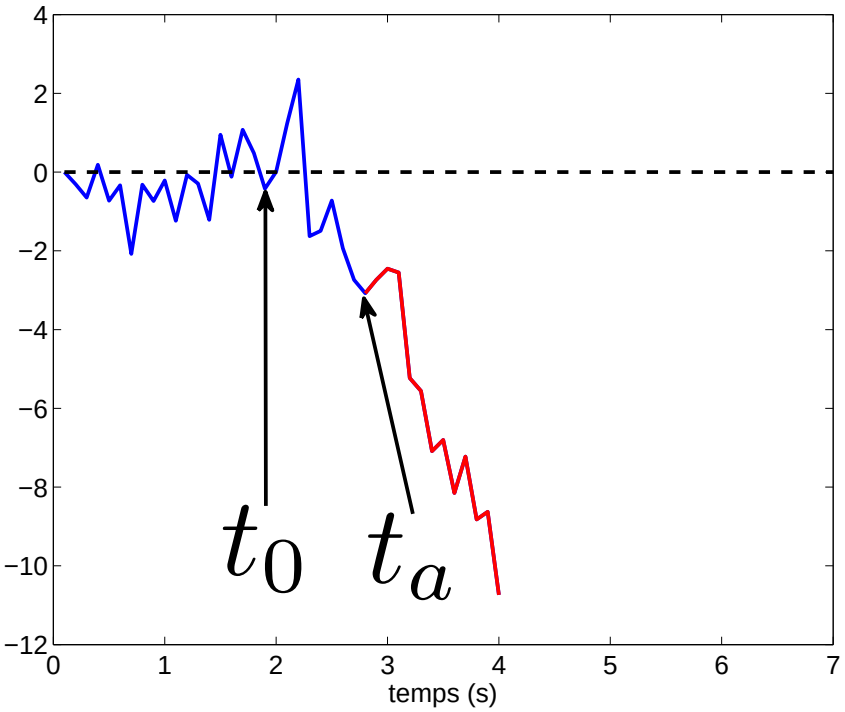


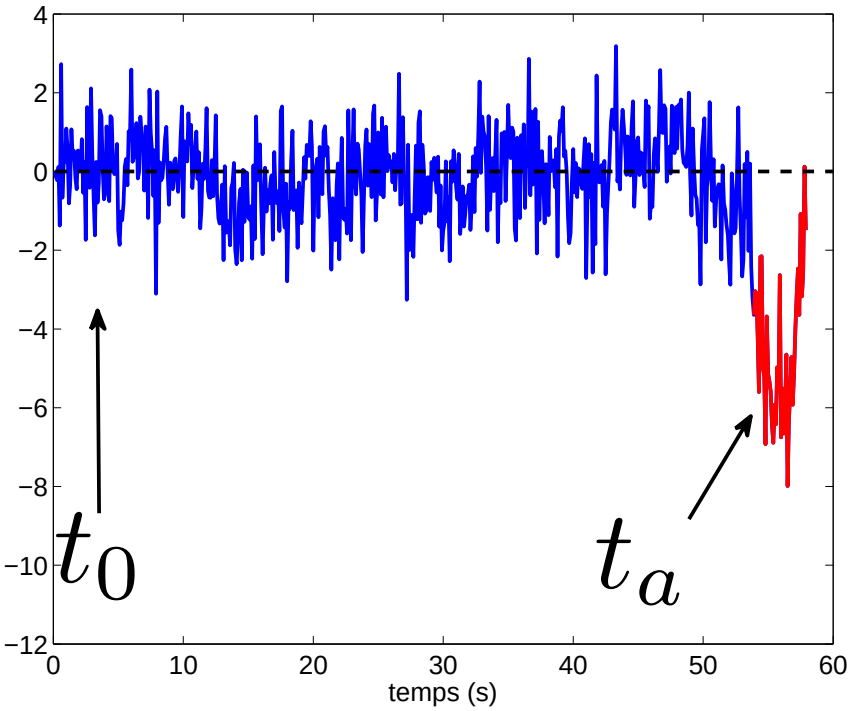
Figure 5.4 – Clusters et ellipsoïdes de confiance dans le MRPF

Analyse des résultats On rappelle que les trajectoires des scénarios 1,2 et 3 sont de durées respectives 901, 1001 et 1001 unités de temps (1 unité de temps = 0.1 s).
L'analyse des performances de l'algorithme de détection s'est faite en prenant en compte l'ambiguïté du terrain survolé. Afin de calculer les trois indicateurs TFA, τ et TND, nous avons fixé le

saut de moyenne à $\nu = 3$ et fait varier le seuil de détection h du CUSUM bilatéral entre 8 et 13. Les résultats sont présentés sur les tableaux 5.1, 5.2 et 5.3.
Lorsque le terrain est peu ambigu (scénario 1), la différence entre la mesure actuelle y_k et la mesure prédite $\hat{y}_{k|k-1}$ augmente rapidement en cas de divergence. On s'attend alors à ce que le détecteur CUSUM détecte rapidement le dysfonctionnement du filtre : ceci est illustré par la figure 5.5a. En revanche lorsque le terrain est ambigu (scénario 3), les similarités de profil sont telles que la mesure prédite $\hat{y}_{k|k-1}$ par le filtre peut être proche de y_k pendant un certain temps bien que le filtre ait divergé : un cas extrême est présenté dans la figure 5.5b (scénario 3) où le résidu oscille autour de zéro pendant une longue période même après l'instant de divergence. La détection ne peut alors se faire qu'une fois l'ambiguïté levée. Ceci se vérifie par simulation : en effet, on observe dans le tableau 5.2 que plus l'ambiguïté est grande, plus le délai de détection est important. En particulier pour le scénario 3, le délai de détection est 5 à 10 fois plus important que celui observé pour les scénarios 1 et 2. Ceci peut être pénalisant dans l'optique d'une ré-initialisation du filtre.
Nous avons également étudié l'influence du seuil de détection sur les indicateurs TFA, τ et TND. On observe qu'il est possible qu'il est possible d'atteindre des taux de fausses alarmes en dessous de 1% en prenant $h \geq 12$. De plus, l'évolution du taux de fausses alarmes TFA et du délai de détection en fonction de h est cohérente avec l'analyse faite sur le modèle de changement de moyenne dans une suite de variables aléatoires gaussiennes (section 5.6.1). En effet, le taux de fausses alarmes est plus sensible aux variations de h que le taux de non détection. Enfin le taux de non-détection (tableau 5.3) est également moins sensible à h que le taux de fausses alarmes.



(a) Terrain peu ambigu



(b) Terrain très ambigu

Figure 5.5 – Innovation normalisée d'un filtre de recalage : cas favorable (a)/défavorable (b) à la détection rapide de divergence

seuil CUSUM	scénario 1	scénario 2	scénario 3
$h = 8$	6	5.2	6.8
$h = 9$	3.6	2.4	1.6
$h = 10$	2	1.2	0.4
$h = 11$	1.2	0.8	0.4
$h = 12$	0.8	0.4	0
$h = 13$	0.4	0.4	0

TABLE 5.1 – Taux de fausses alarmes TFA (%)

seuil CUSUM	scénario 1	scénario 2	scénario 3
$h = 8$	1.9	4.1	14
$h = 9$	1.9	4.3	16.6
$h = 10$	2	4.3	18.9
$h = 11$	2.1	4.3	19.3
$h = 12$	2.2	4.3	20.5
$h = 13$	2.2	4.4	22.3

TABLE 5.2 – Délai moyen de détection τ (en secondes, mesures délivrées à 10 Hz) ; Durée des trajectoires : 100 s

seuil CUSUM	scénario 1	scénario 2	scénario 3
$h = 8$	2.4	3.2	3.2
$h = 9$	2.8	3.6	4.4
$h = 10$	3.6	5.2	4.4
$h = 11$	4	5.6	5.2
$h = 12$	4.4	6.8	5.6
$h = 13$	5.2	6.8	6.8

TABLE 5.3 – Taux de non-détection TND (%)

5.7 Ré-initialisation du filtre particulaire après divergence

L'algorithme de détection de divergence présenté précédemment présente un intérêt évident : la possibilité de déclencher la ré-initialisation du filtre une fois la divergence détectée. Nous proposons d'ajouter une étape de ré-initialisation prenant en compte l'incertitude courante de la centrale inertielle ainsi que la ou les dernières observations altimétriques accumulées. Le modèle de dérive inertielle considéré est donné par (4.99).

5.7.1 Choix de la loi a priori pour la ré-initialisation du filtre

L'option la plus directe consiste à ré-initialiser le filtre à partir de l'état courant donné par la centrale inertielle. Supposons qu'une divergence soit détectée à l'instant t_a. L'erreur inertielle initiale x_0 est modélisée comme une variable aléatoire gaussienne centrée et de covariance P_0. Etant donné la nature linéaire du modèle d'état (4.99), l'erreur inertielle x_k à un instant k quelconque est également une distribution gaussienne de moyenne nulle et de covariance donnée par :

$$P_k^x \triangleq \mathrm{Cov}(x_k) = F^k P_0 (F^k)^T + \sum_{j=0}^{k-1} F^j G_{k-j+1} Q G_{k-j+1}^T (F^j)^T \tag{5.58}$$

En effet, on montre par récurrence que

$$x_k = F^k x_0 + \sum_{j=1}^{k} F^{j-1} G_{k-j} w_{k-j}$$

d'où $\mathbb{E}(x_k) = 0$ puisque $\mathbb{E}(x_0) = 0$ et $\mathbb{E}(w_p) = 0$ pour tout $p \leq k$. De plus,

$$\begin{aligned}
P_k^x &= \mathbb{E}\left(F^k x_0 + \sum_{j=1}^{k} F^{j-1} G_{k-j} w_{k-j}\right)\left(F^k x_0 + \sum_{l=1}^{k} F^{l-1} G_{k-l} w_{k-l}\right)^T \\
&= F^k \mathbb{E}(x_0 x_0^T)(F^k)^T + \mathbb{E}\left(F^k x_0 \left(\sum_{l=1}^{k} F^{l-1} G_{k-l} w_{k-l}\right)^T\right) + \mathbb{E}\left(\left(\sum_{j=1}^{k} F^{j-1} G_{k-j} w_{k-j}\right)(F^k x_0)^T\right) \\
&\quad + \mathbb{E}\left(\sum_{j=1}^{k} F^{j-1} G_{k-j} w_{k-j}\right)\left(\sum_{l=1}^{k} F^{l-1} G_{k-l} w_{k-l}\right)^T \\
&= F^k P_0 (F^k)^T + 0 + 0 + \sum_{1 \leq j \leq l \leq k} F^{j-1} G_{k-j} \mathbb{E}(w_{k-j} w_{k-l}^T) G_{k-l}^T (F^{l-1})^T \\
&= F^k P_0 (F^k)^T + \sum_{1 \leq j \leq l \leq k} F^{j-1} G_{k-j} \delta_{j,l} Q G_{k-l}^T (F^{l-1})^T \\
&= F^k P_0 (F^k)^T + \sum_{j=1}^{k} F^{j-1} G_{k-j} Q G_{k-j}^T (F^{j-1})^T \\
&= F^k P_0 (F^k)^T + \sum_{j=0}^{k-1} F^j G_{k-j+1} Q G_{k-j+1}^T (F^j)^T
\end{aligned}$$

Le terme $\mathbb{E}\left(F^k x_0 \left(\sum_{l=1}^{k} F^{l-1} G_{k-l} w_{k-l}\right)^T\right)$ de la seconde ligne se décompose comme

$$F^k \mathbb{E}(x_0) \mathbb{E}\left(\sum_{l=1}^{k} F^{l-1} G_{k-l} w_{k-l}\right)^T = 0$$

étant donné l'indépendance de la condition initiale x_0 et du processus de bruit $(w_k)_k$. Enfin, le caractère i.i.d. gaussien de $w_0, \ldots, w_k$ de bruit donne $\mathbb{E}(w_{k-j} w_{k-l}^T) = \delta_{j,l} Q$.

La procédure de ré-initialisation peut se faire en tirant N particules $x_{t_a}^i$ en prenant comme loi a priori

1. la loi de densité $p(x_{t_a})$, gaussienne centrée et de covariance P_{t_a}
2. la loi conditionnelle de densité $p(x_{t_a}|y_{t_a-m+1:ta})$ qui prend en compte les m dernières mesures accumulées par le filtre

La première méthode est une ré-initialisation complètement aveugle puisqu'elle ne prend pas en compte les mesures reçues par le radio-altimètre. Elle est donc potentiellement pénalisante lorsque le délai de détection est grand, puisque dans ce cas la matrice de covariance P_{t_a} de la loi a priori est très grande ce qui veut dire que le filtre peut mettre un certain temps avant de reconverger. La seconde méthode permet de diminuer l'incertitude d'autant plus que le terrain varie fortement au voisinage de la position inertielle à l'instant t_a et que le nombre m de mesures est important. Pour la mettre en oeuvre il faut néanmoins échantillonner efficacement $p(x_{t_a}|y_{t_a-m+1:ta})$ ou une approximation de cette de cette densité.

Soit $Y \triangleq \begin{bmatrix} y_{ta} & y_{t_a-1} \cdots y_{t_a-m+1} \end{bmatrix}^T$. Dans le cas d'un véhicule dont le déplacement entre les instants $t_a - m + 1$ et t_a autorise la linéarisation du profil de terrain h_{MNT} entre ces instants, on peut montrer que (cf. annexe D)

$$Y = \tilde{Z} + H_1 x_{t_a}^{(2)} - \begin{bmatrix} h_{MNT}(\lambda(x_{t_a,1}), \phi(x_{t_a,2})) \\ \vdots \\ h_{MNT}(\lambda(x_{t_a,1}), \phi(x_{t_a,2})) \end{bmatrix} + V \tag{5.59}$$

où $\tilde{Z} = \begin{bmatrix} \tilde{z}_{t_a} & \tilde{z}_{t_a-1} \cdots \tilde{z}_{t_a-m+1} \end{bmatrix}^T$ est le vecteur des m dernières altitudes inertielles. $x_{t_a}^{(2)}$ est la partie linéaire de x_{t_a} et H_1 est défini par

$$H_1 = \begin{bmatrix} -1 & (m-1)\nabla h_{MNT}^T(\lambda(x_{t_a,1}), \phi(x_{t_a,2})) & (m-1) \\ \vdots & \vdots & \vdots \\ -1 & (m-1)\nabla h_{MNT}^T(\lambda(x_{t_a,1}), \phi(x_{t_a,2})) & 1 \\ -1 & 0 & 0 \end{bmatrix}$$

Le terme $h_{MNT}(\lambda(x_{t_a,1}), \phi(x_{t_a,2}))$ désigne la hauteur du terrain à la position courante dont la latitude et longitude s'exprime en fonction de la latitude et longitude inertielle $(\tilde{\lambda}, \tilde{\phi})$ et de l'écart

métrique nord-est $(x_{t_a,1}, x_{t_a,2})$ selon (4.102). Enfin V est un bruit additif décrit par :

$$V = \begin{bmatrix} v_{t_a-m+1} + \sum_{l=2}^{m}(l-1)[W_{t_a-m+l,3} - W^{(1)}_{t_a-m+l}\nabla h^T_{MNT}(\lambda(x_{t_a,1}),\phi(x_{t_a,2}))] \\ \vdots \\ v_{t_a-1} + [W_{t_a,3} - W^{(1)}_{t_a}\nabla h^T_{MNT}(\lambda(x_{t_a,1}),\phi(x_{t_a,2}))] \\ v_{t_a} \end{bmatrix}$$

avec $W_k = G_k w_k$, w_k étant le vecteur du bruit de dynamique du modèle (4.99) et $W^{(1)}_l = [W_{l,1} \quad W_{l,2}]^T$, $\forall l$.

L'objectif est de tirer un échantillon $x^i_{t_a}$, $i = 1, \ldots, N$ issu de $p(x_{t_a}|y_{t_a-m+1:ta}) = p(x_{t_a}|Y)$. Pour cela, on partitionne x_{t_a} sous la forme

$$x_{t_a} = \begin{bmatrix} x^{(1)}_{t_a} & x^{(2)}_{t_a} \end{bmatrix}^T$$

et la covariance P_{t_a} selon

$$P_{t_a} = \begin{pmatrix} P^{(11)}_{t_a} & P^{(12)}_{t_a} \\ P^{(21)}_{t_a} & P^{(22)}_{t_a} \end{pmatrix}$$

l'équation de mesure (5.59) se ré-écrit comme

$$Y = H_1 x^{(2)}_{t_a} - h_1(x^{(1)}_{t_a}) + V \tag{5.60}$$

Soit $x \mapsto \varphi(x|\mu, \Sigma)$ la densité gaussienne de moyenne μ et covariance Σ. D'après [28], l'échantillon $x^i_{t_a}$, $i = 1, \ldots, N$ peut être généré en deux étapes

(i) On tire

$$x^{(1),i}_{t_a} \sim p(x^{(1)}_{t_a}|Y) \propto \varphi(x^{(1)}_{t_a}|P^{(11)}_{t_a})\varphi(Y - h_1(x^{(1)}_{t_a}) - H_1 P^{(21)}_{t_a}(P^{(11)}_{t_a})^{-1}x^{(1)}_{t_a}|R_1 + H_1 P^{(22|1)}_{t_a} H_1^T)$$

selon une méthode d'acceptation-rejet ou une méthode approchée, avec

$$P^{(22|1)}_{t_a} = P^{(22)}_{t_a} - P^{(21)}_{t_a}(P^{(11)}_{t_a})^{-1}P^{(12)}_{t_a}$$

La matrice de covariance R_1 a pour terme i, j

$$R_{1,ij} = \delta_{i,j}\sigma_v^2 + \sum_{l=\max(i,j)+1}^{m}(l-i)(l-j)\left[Q_{3,3} - \nabla h^T_{MNT}(\lambda(x_{t_a,1}),\phi(x_{t_a,2}))Q^{(11)}\nabla h_{MNT}(\lambda(x_{t_a,1}),\phi(x_{t_a,2}))\right]$$

(ii) On complète le tirage en échantillonnant les parties linéaires $x^{(2),i}_{t_a} \sim \mathcal{N}(\hat{x}^{(2)}_{t_a}(Y, x^{(1),i}_{t_a}), C^{(22|1)}_{t_a})$ où

$$\hat{x}^{(2)}_{t_a}(Y, x^{(1),i}_{t_a}) = C^{(22|1)}_{t_a} H_1^T R_1[Y_1 - h_1(x^{(1),i}_{t_a})] + (P^{(22|1)}_{t_a})^{-1}P^{(21)}_{t_a}(P^{(11)}_{t_a})^{-1}x^{(1),i}_{t_a}$$

et

$$C^{(22|1)}_{t_a} = \left(H_1^T R_1^{-1} H_1 + P^{(22|1)-1}_{t_a}\right)^{-1}$$

Le principal inconvénient de cette méthode est son coût de calcul élevé, du notamment à la boucle d'acceptation-rejet dans la première étape. Pour cette raison, on préfèrera le choix de la loi a priori gaussienne centrée et de covariance P_{t_a}.

5.7.2 Simulations

Nous allons à présent combiner l'algorithme de ré-initialisation du filtre avec celui de détection de divergence et le comparer à un filtre sans module de ré-initialisation. Le filtre considéré ici est le MRPF mais il est tout à fait possible d'utiliser un autre filtre. Comme dans le chapitre 4, nous évaluons le RMSE sur les trajectoires non-divergentes des deux algorithmes ainsi que leur taux de non divergence. Les résultats des simulations sont évalués sur les même scénarios décrits dans la section 4.6.1. Pour chacun d'entre eux, 150 réalisations Monte Carlo ont été effectuées avec un paramétrage artificiel favorisant les divergences de l'algorithme (cf. 5.6.3).

Conditions de simulations Nous appellerons MRPF-R l'algorithme avec ré-initialisation. Nous reprenons les conditions de simulations de la section 4.6.4.1 avec les modifications suivantes :

- ▷ écart-type du bruit de mesure du radio-altimètre $\sigma_v = 10\,m$
- ▷ nombre de particules MRPF : $N_{MRPF} = 3000$
- ▷ nombre de particules MRPF-R : $N_{MRPF-R} = 3000$

Dans l'algorithme MRPF-R, une seule ré-initialisation est effectuée au maximum étant donné que la trajectoire a une durée finie et que l'on souhaite accorder au filtre le temps nécessaire pour se recaler. Enfin, conformément au chapitre 4, un filtre sera dit non-divergent si l'état estimé final est dans l'ellipsoïde à 99.9% associée à la distribution gaussienne centrée sur l'état final vrai et de covariance la covariance empirique donnée par le filtre particulaire.

Analyse des résultats Les figures 5.6 à 5.11 représentent le RMSE de chaque algorithme obtenu par tirage Monte Carlo. Le RMSE est calculé sur les réalisations convergentes. On observe

	nombre de ré-initialisations	nombre de ré-initialisations convergentes
scénario 1	59	36
scénario 2	83	63
scénario 3	67	35

TABLE 5.4 – Performance de la ré-initialisation dans le MRPF-R

que l'utilisation du module de détection de divergence permet au filtre de gagner en robustesse (voir tableau 5.5) : sur chacun des scénarios, on note une augmentation significative du taux de non divergence. L'erreur quadratique moyenne est naturellement plus élevée pour le MRPF-R dans la phase multimodale puisqu'il inclut les ré-initialisations. Pour les scénarios 1 et 2, cette

dégradation est peu pénalisante étant donné que la phase multimodale est relativement courte devant la durée de la trajectoire vu le caractère peu ambigu (scénario 1) et moyennement ambigu (scénario 2) des terrains survolés. En revanche, pour le scénario 3, l'ambiguïté pénalise le RMSE de l'algorithme avec ré-initialisation étant donné que le délai de détection est non négligeable et que la phase multimodale après ré-initialisation est plus longue.

	MRPF	MRPF-R
scénario 1	78	85
scénario 2	74	87
scénario 3	63	78

Table 5.5 – Taux de non divergence (%)

Scénario 1

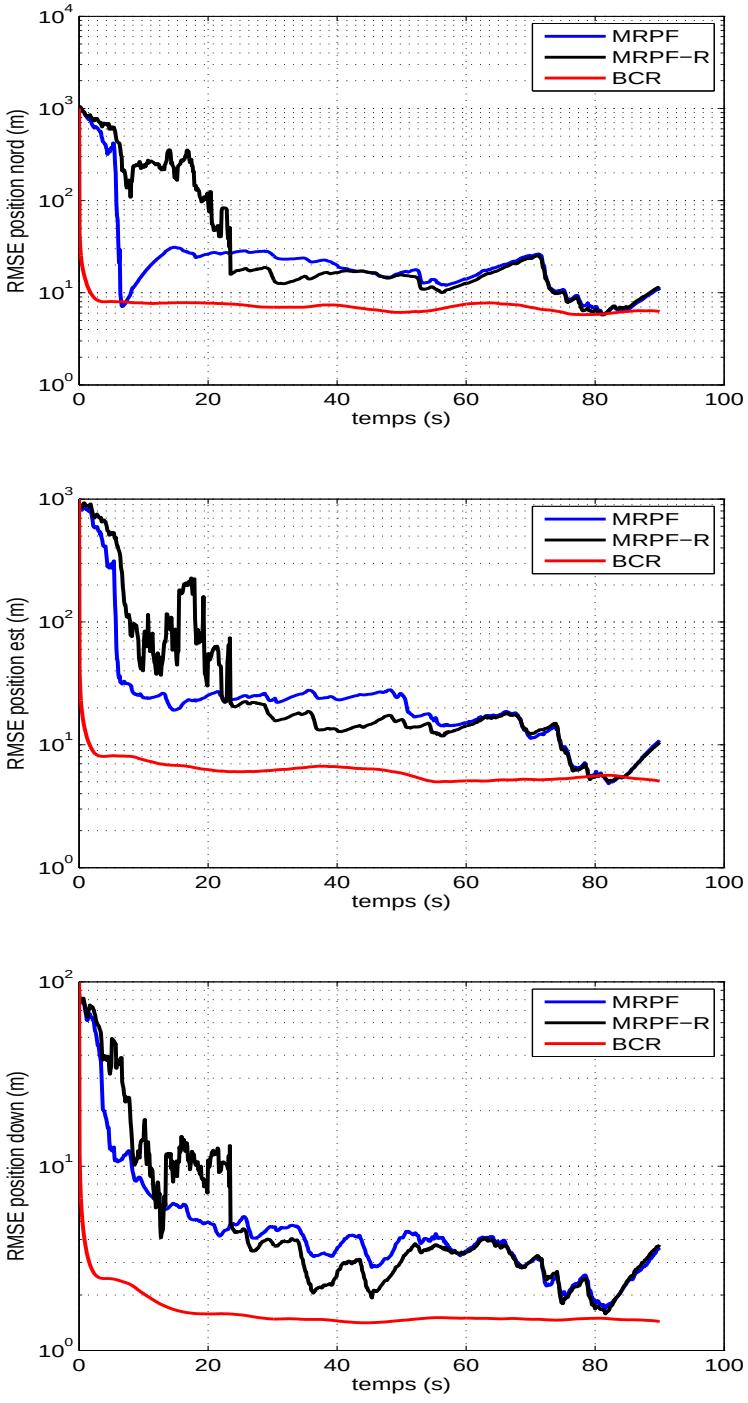


FIGURE 5.6 – RMSE en position ; noir : avec ré-initialisation, bleu : sans ré-initialisation, rouge : Borne de Cramer-Rao a posteriori

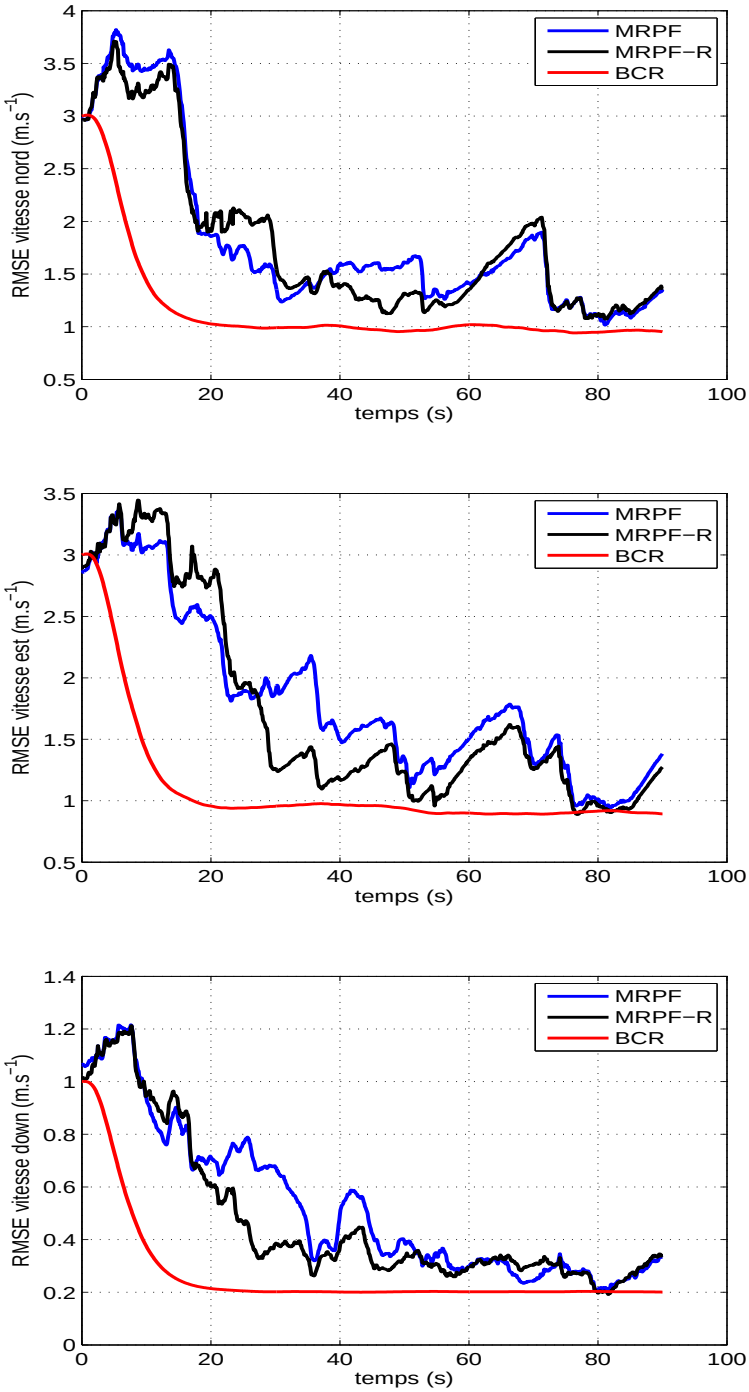


Figure 5.7 – RMSE en vitesse ; noir : avec ré-initialisation, bleu : sans ré-initialisation, rouge : Borne de Cramer-Rao a posteriori

Scénario 2

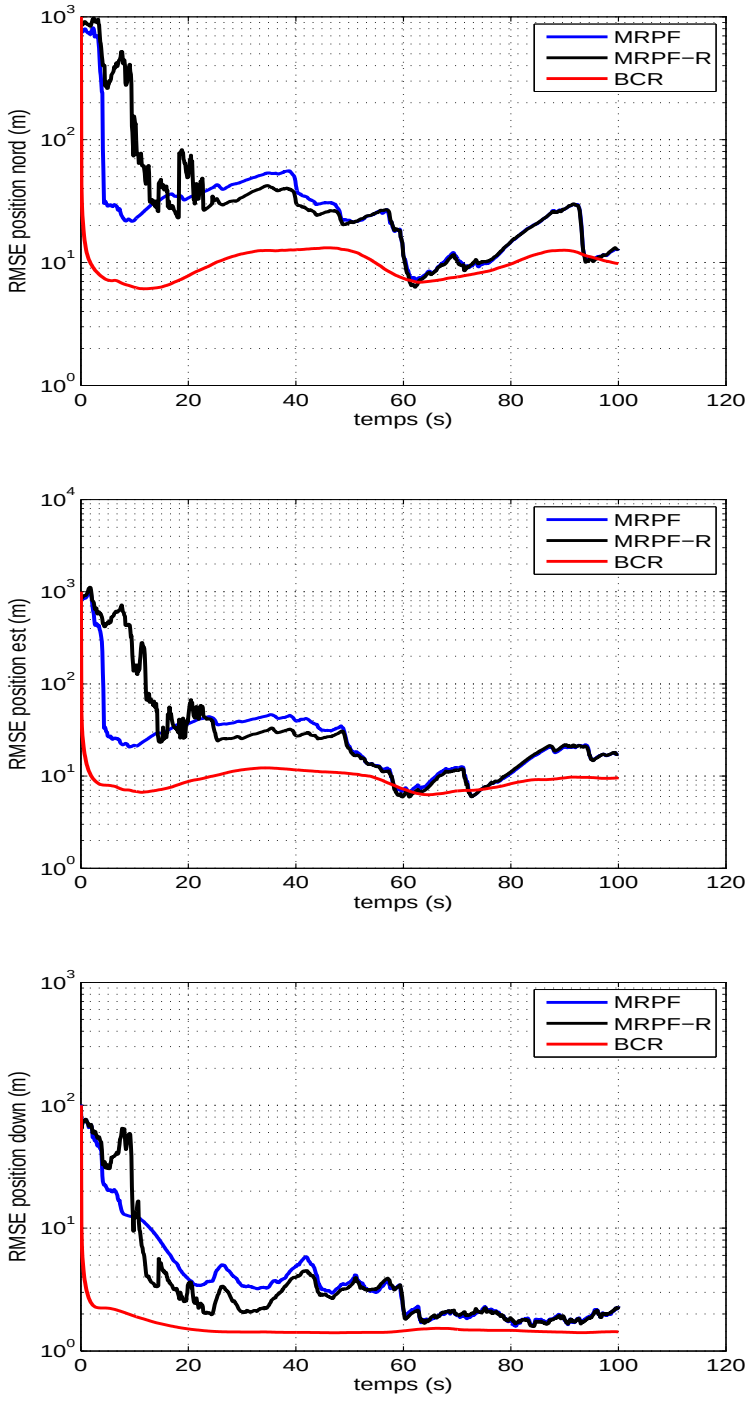


FIGURE 5.8 – RMSE en position ; noir : avec ré-initialisation, bleu : sans ré-initialisation, rouge : Borne de Cramer-Rao a posteriori

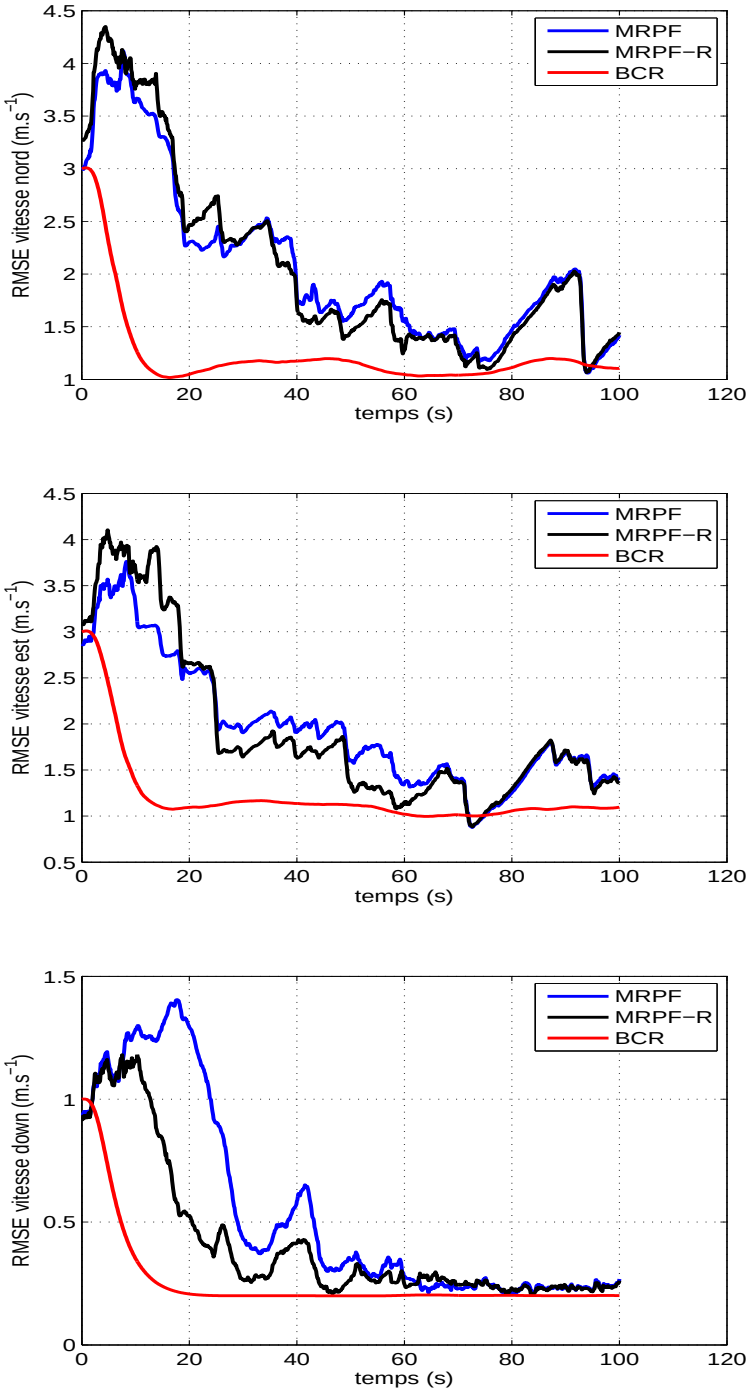


Figure 5.9 – RMSE en vitesse ; noir : avec ré-initialisation, bleu : sans ré-initialisation, rouge : Borne de Cramer-Rao a posteriori

Scénario 3

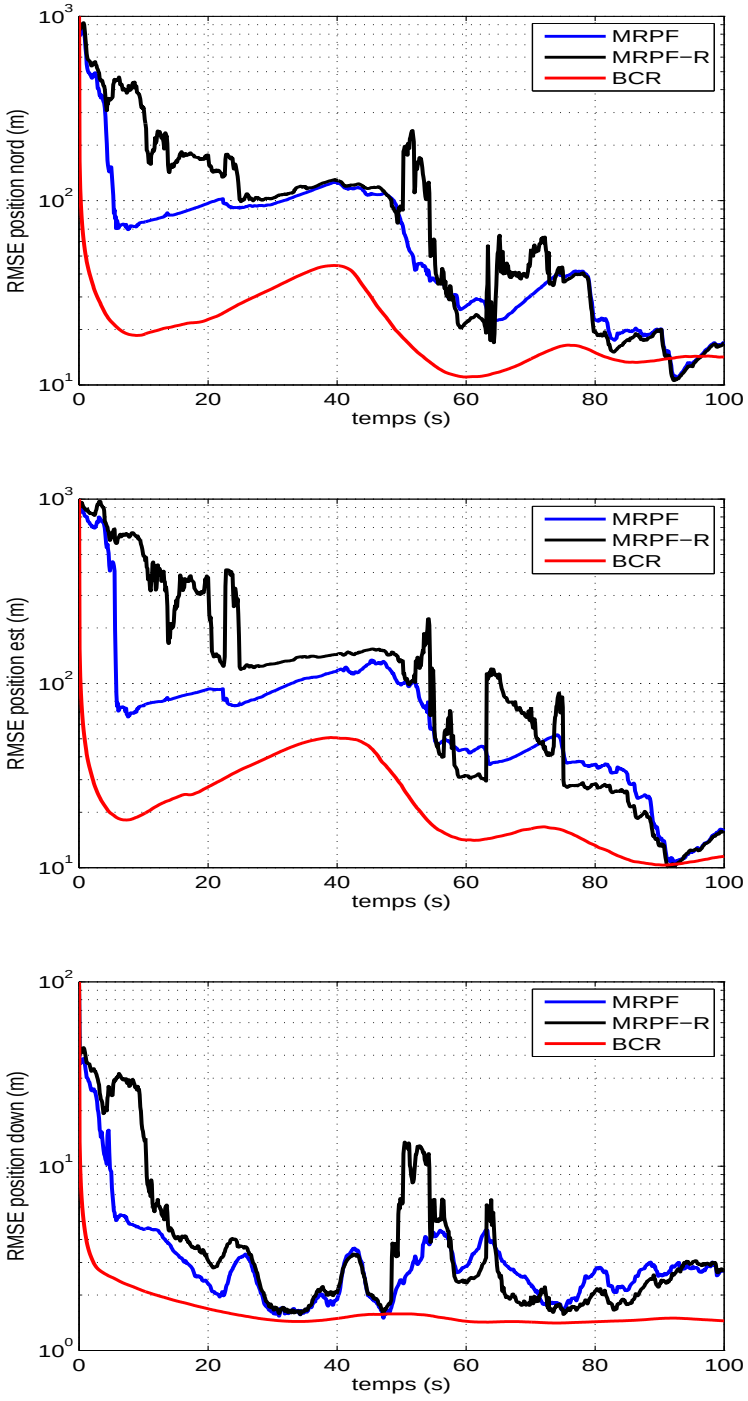


FIGURE 5.10 – RMSE en position ; noir : avec ré-initialisation, bleu : sans ré-initialisation, rouge : Borne de Cramer-Rao a posteriori

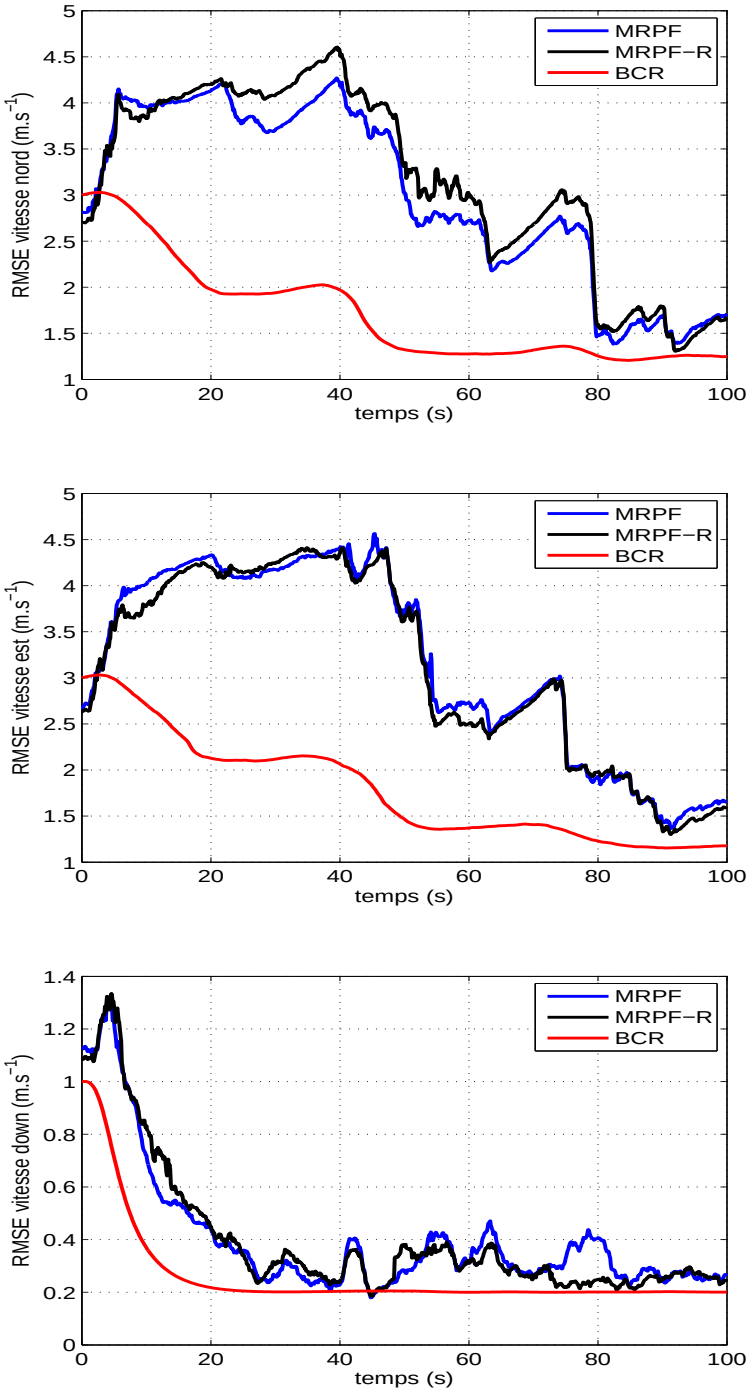


Figure 5.11 – RMSE en vitesse ; noir : avec ré-initialisation, bleu : sans ré-initialisation, rouge : Borne de Cramer-Rao a posteriori

Conclusion du Chapitre 5

Nous avons vu au cours des chapitres précédents que les algorithmes de filtrage particulaire d'hybridation pour le recalage peuvent diverger pour de multiples raisons liées à l'utilisation successives d'approximations Monte Carlo. Il est intéressant pour l'ingénieur de disposer d'un outil de diagnostic permettant d'estimer automatiquement et en ligne la fiabilité de la solution particulaire. Nous avons donc introduit et étudié une méthodologie pour la détection de divergence dans un filtre particulaire quelconque, avec l'hypothèse d'une observation scalaire. La méthode développée s'appuie sur la détection de changement de moyenne dans l'innovation du filtre particulaire grâce à l'algorithme CUSUM.
Le détecteur ainsi construit s'intègre facilement dans n'importe quelle routine de filtrage particulaire avec un coût additionnel relativement faible. Les performances de l'algorithme de détection sont étudiées sur la problématique de la navigation inertielle par mesures altimétriques, en considérant trois terrains de caractéristiques différentes. Sur la base de simulations, on montre qu'il est possible d'atteindre des taux de fausses alarmes inférieurs à 1%, avec un délai de détection acceptable devant la durée de la trajectoire. Le délai de détection moyen reste alors petit devant la durée de la trajectoire lorsque le niveau d'ambiguïté du relief survolé est faible à modéré. Pour des scénarios ambigus, la définition du test d'hypothèse s'avère moins efficace, puisque la divergence ne se traduit pas forcément par un changement significatif de la moyenne de l'innovation du filtre particulaire. En pratique cela signifie que les délais de détection sont bien plus importants et dépendent de la vitesse à laquelle l'ambiguïté est levée.
Enfin, nous avons illustré une application possible de l'algorithme de détection qui est celle de la ré-initialisation du filtre particulaire lorsqu'un dysfonctionnement a été détecté. Les résultats préliminaires suggèrent que l'on peut atteindre des taux de non-divergence plus élevés.

Conclusion générale

Le développement du filtrage particulaire dans les années 90 a ouvert des nouvelles possibilités dans le domaine du filtrage non-linéaire avec des applications diverses (pistage, surveillance vidéo, finance de marché, économétrie,...). En aéronautique, la mise au point de filtres particulaires pour le recalage de la navigation inertielle devrait permettre de s'affranchir de certaines hypothèses limitant l'utilisation d'algorithmes traditionnels (systèmes TERCOM, SITAN). Pour autant, ces méthodes soulèvent un certain nombre de difficultés. Dans cette thèse, nous nous sommes principalement intéressés à deux limitations rencontrées en recalage par mesures altimétriques :

- les similarités de terrain font que la densité a posteriori est multimodale en début de recalage. Les étapes de ré-échantillonnage successives peuvent provoquer la perte du mode correspondant à la position vraie.
- lorsque le terrain présente localement un gradient prononcé, ou bien lorsque la variance du bruit de mesure est faible, l'étape de prédiction des filtres standards (*bootstrap filter*, *RPF*) tend à échantillonner les particules dans des zones de faible vraisemblance ce qui augmente le risque de divergence du filtre.

Afin de limiter les divergences dues à une perte de mode prématurée, nous avons proposé une approche algorithmique formulant la densité conditionnelle cible sous forme de mélange fini. Les filtres développés au cours de cette thèse (MRPF, MRBPF) sont des extensions du MPF de Vermaak et al. [118]. Dans le MPF, chaque composante du mélange est supposée unimodale et est approchée par un cluster de particules qui peut être vu comme un sous-filtre dont la contribution est quantifée par le poids de mélange associé. Cette formulation permet de ré-échantillonner localement chaque cluster dans l'optique de maintenir la multimodalité autant que nécessaire. La première nouveauté dans le MRPF et le MRBPF par rapport au MPF est l'utilisation d'un algorithme de mean-shift clustering optimisé, pour l'identification automatique des clusters du nuage de particules. Une attention particulière a été accordée à la sélection de la fenêtre de lissage qui conditionne la bonne identification des clusters. Nous avons également suggéré une méthode permettant de supprimer les clusters devenus peu vraisemblables et de ré-allouer un nombre équivalent de particules aux clusters restants. Enfin, nous avons proposé une étape de régularisation locale pour conserver la forme de la distribution conditionnelle. Le MRPF a été comparé au RPF et s'est montré plus robuste aux divergences que ce dernier, notamment en cas de forte ambiguïté de terrain.

Dans un second temps, nous avons introduit un méthode d'échantillonnage d'importance permettant de générer les particules selon une densité différente de la densité de transition. Cette méthode est efficace lorsqu'il y a un faible recouvrement entre la densité prédite et la fonction

de vraisemblance. Cette situation survient par exemple lorsque la variance de bruit de mesure ou d'état est faible, ou lorsque la fonction d'observation varie brusquement. L'objectif de la méthode d'échantillonnage d'importance est de générer les particules dans les zones de forte probabilité. Dans cette thèse, l'échantillonnage des particules est effectuée selon une série de densités d'importances gaussiennes centrées sur les modes de la densité a posteriori. Ces modes ne sont pas connus explicitement mais sont calculés de manière approchée, en exploitant la structure partiellement linéaire de l'équation d'observation qui permet de se ramener à un problème d'optimisation en deux-dimensions. Nous avons également exploré différentes formulations de la matrice de covariance des densités d'importances gaussiennes afin d'approcher au mieux l'orientation et la dispersion locale de la densité a posteriori.
L'utilisation de cette méthode d'échantillonnage d'importance a donné lieu aux algorithmes MRPF-MAP et MRBPF-MAP. Le MRPF-MAP a été comparé au MRPF dans des contextes où les étapes de prédiction/correction classiques sont moins efficaces. Le MRPF-MAP améliore le taux de convergence par rapport au MRPF. De plus, on observe que l'erreur quadratique moyenne est plus faible sur une bonne portion des trajectoires tests. Cette méthode d'échantillonnage d'importance permet d'obtenir des améliorations encore plus significatives dans le MRBPF-MAP, notamment en terme de taux de non-divergence, lorsque la variance du bruit de mesure est faible.

Dans un troisième temps, nous avons abordé la problématique du contrôle d'intégrité du filtre particulaire. L'objectif est d'évaluer en ligne la cohérence de l'estimation particulaire afin de détecter une éventuelle divergence du filtre. En filtrage de Kalman standard, plusieurs tests fondés sur l'innovation ont été proposés afin de contrôler le bon fonctionnement de l'architecture de filtrage. Cependant, l'extension au filtrage particulaire est à notre connaissance très peu abordée. Nous avons formulé le problème de détection de la divergence du filtre particulaire sous forme d'un test d'hypothèse portant sur la moyenne de l'innovation normalisée du filtre, qui est définie de manière analogue à l'innovation du filtre de Kalman. Dans le cas d'une innovation scalaire, l'algorithme de détection adapté au problème de détection de changement de moyenne est le CUSUM bilatéral. Les paramètres influents de cet algorithme sont le saut de la moyenne de l'innovation normalisée sous l'hypothèse de divergence et le seuil sur la fonction de décision. Le détecteur de divergence a été intégré dans le MRPF et ses performances ont été établies sur trois scénarios de recalage de faible, moyenne et forte ambigüité de terrain. Sur chacun des scénarios les simulations ont fait apparaître un paramétrage du CUSUM permettant d'obtenir des taux de détection supérieur à 93% pour des taux de fausses alarmes inférieur à 1%. Lorsque le vecteur de mesure est multi-dimensionnel, nous avons proposé quelques formulations simples du problème de détection de divergence issues de la théorie de détection de changement et illustré la capacité des algorithmes de détection issus de la littérature à détecter la divergence du filtre particulaire. Enfin, nous avons réalisé une étude préliminaire sur l'apport de la ré-initialisation du filtre en cas de détection d'un faux recalage. L'algorithme de ré-initialisation est basée sur l'utilisation d'une loi a priori correspondant à la distribution de la dérivé inertielle à l'instant de détection. Les résultats expérimentaux montrent que cette méthode permet de diminuer le taux de faux recalage.

Perspectives

Au terme de ces travaux, nous avons envisagé plusieurs pistes d'études complémentaires.

- ▷ La stratégie de clustering pourrait être adaptée au KPKF avec l'avantage que ce dernier nécessite généralement moins de particules que les filtres particulaires standards pour des performances identiques.
- ▷ Dans le MRPF ou le MRBPF, lorsqu'il reste peu de modes, on peut étudier la possibilité de supprimer les clusters très peu vraisemblables sans ré-allouer un nombre équivalents de particules aux clusters restants. Cela permettrait de créer un algorithme dont le nombre de particules est adaptatif.
- ▷ Les algorithmes proposés mettent en jeu de nombreux paramètres qui ont été réglés empiriquement. Un apprentissage statistique peut permettre d'adapter ces paramètres en fonction du type de terrain survolé et des conditions initiales.
- ▷ Dans le test de divergence développé, l'hypothèse de non-divergence est caractérisée par une approximation de la moyenne de l'innovation normalisée (supposée nulle en fonctionnement normal). Il peut être intéressant de disposer d'un intervalle de confiance (ou d'une zone de confiance dans le cas non scalaire) sur cette approximation de manière à définir un test présentant moins de fausses alarmes.
- ▷ L'algorithme de ré-initialisation du filtre proposé ne prend pas en compte l'historique du filtre. La construction d'une loi a priori prenant en compte non seulement l'incertitude de la centrale inertielle, mais également les zones du relief incompatibles avec les mesures radio-altimétriques passées, doit permettre au filtre ré-initialisé de converger plus vite avec un meilleur taux de bon recalage.

Annexe A

Compléments au chapitre 1

A.1 Lemme 1

Soit A, B, C et D 4 matrices telles que les inverses $(A + BCD)^{-1}$, A^{-1} et C^{-1} existent. Alors,

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} \tag{A.1}$$

A.2 Lemme 2

Soit Q et R deux matrices symétriques définies positives, de dimensions respectives $d \times d$ et $m \times m$. Soit H une matrice $m \times d$. Alors,

$$(H^T R^{-1} H + Q^{-1})^{-1} = Q - QH^T(HQH^T + R)^{-1}HQ \tag{A.2}$$

et

$$(H^T R^{-1} H + Q^{-1})^{-1} H^T = QH^T(HQH^T + R)^{-1}R \tag{A.3}$$

Annexe B

Complément au chapitre 2

Variance du temps moyen d'absorption

L'objectif de cette section est d'établir une expression asymptotique de la variance du temps moyen d'absorption dans le modèle de pistage stationnaire présenté au chapitre 2, section 2.5. Rappelons que dans ce contexte, le temps moyen d'absorption correspond à la perte de l'un des deux modes de la densité a posteriori $p(x_k|y_{0:k}) = 0.5\delta_1(x_k) + 0.5\delta_{-1}(x_k)$. La démarche utilisée pour obtenir l'approximation asymptotique de la variance est largement inspirée de [39].

Rappelons d'abord quelques notations. On désigne par A_n le nombre de particules associées au premier mode ($x = 1$) où N la taille de l'échantillon. $(A_n)_{n\geq 0}$ est une chaîne de Markov homogène à valeur dans $\{0, 1, \ldots, N\}$, dont la matrice de transition $P = (p_{mj})_{0\leq m,j\leq N}$ est donnée par

$$p_{mj} = \mathbb{P}(A_k = j|A_{k-1} = m) = \binom{N}{j} \left(\frac{m}{N}\right)^j \left(1 - \frac{m}{N}\right)^{N-j} \tag{B.1}$$

Le temps d'absorption T est défini par

$$T = \inf\{n \geq 0 \mid A_n \in \{0, N\}\} \tag{B.2}$$

De même, pour $i = 1 \ldots N$, $t_i = \mathbb{E}(T|A_0 = i)$ est le temps moyen d'absorption, sachant que l'on dispose initialement de i particules sur le mode $x = 1$.
En notant $u = \frac{i}{N}$ la proportion initiale de particules associée au mode $x = 1$, on a l'équivalent suivant lorsque N est grand [39] :

$$t(u) \stackrel{\triangle}{=} \mathbb{E}(T \mid A_0 = \lfloor Nu \rfloor) \sim -2N(u \ln u + (1 - u) \ln(1 - u)) \tag{B.3}$$

Le calcul de la variance du temps moyen d'absorption peut se faire de manière similaire Ãă [39]. Notons $V(u)$ cette variance pour une proportion initiale de u particules associées au mode 1.

$$V(u) \stackrel{\triangle}{=} \mathrm{Var}(T \mid A_0 = \lfloor Nu \rfloor) \tag{B.4}$$

On a :

$$T = 1 + \inf\{n \geq 0|A_{n+1} = 0, N\} = 1 + T_1$$

avec $T_1 = \inf\{n \geq 0 | A_{n+1} = 0, N\}$. D'où

$$\mathbb{E}(T^2|A_0 = i) = 1 + 2\mathbb{E}(T_1|A_0 = i) + \mathbb{E}(T_1^2|A_0 = i) \tag{B.5}$$
$$= 1 + 2\mathbb{E}(T - 1|A_0 = i) + \mathbb{E}(T_1^2|A_0 = i) \tag{B.6}$$
$$= -1 + 2\mathbb{E}(T|A_0 = i) + \mathbb{E}(T_1^2|A_0 = i) \tag{B.7}$$

On peut alors exprimer (B.7) en fonction des v_i et t_i :

$$\mathbb{E}(T_1^2|A_0 = i) = \sum_{t=0}^{\infty} t^2 \mathbb{P}(T_1 = t|A_0 = i) \tag{B.8}$$
$$= \sum_{t=0}^{\infty} t^2 \sum_{j=0}^{N} \mathbb{P}(T_1 = t|A_1 = j, A_0 = i) p_{ij} \tag{B.9}$$
$$= \sum_{j=0}^{N} \sum_{t=0}^{\infty} t^2 \mathbb{P}(T_1 = t|A_1 = j) p_{ij} \tag{B.10}$$
$$= \sum_{j=0}^{N} \sum_{t=0}^{\infty} t^2 \mathbb{P}(T = t|A_0 = j) p_{ij} \tag{B.11}$$
$$= \sum_{j=0}^{N} \mathbb{E}(T^2|A_0 = j) p_{ij} \tag{B.12}$$

d'où :

$$\mathbb{E}(T^2|A_0 = i) = -1 + 2\mathbb{E}(T|A_0 = i) + \sum_{j=0}^{N} \mathbb{E}(T^2|A_0 = j) p_{ij}$$

Enfin,

$$\begin{aligned}
v_i &= \mathbb{E}(T^2|A_0 = i) - \mathbb{E}^2(T|A_0 = i) \\
&= -1 + 2\mathbb{E}(T|A_0 = i) + \sum_{j=0}^{N} \mathbb{E}(T^2|A_0 = j) p_{ij} - [\mathbb{E}(T|A_0 = i)]^2 \\
&= -1 + 2t_i + \sum_{j=0}^{N} \left[\mathbb{E}(T^2|A_0 = j) - \mathbb{E}^2(T|A_0 = j) + \mathbb{E}^2(T|A_0 = j)\right] p_{ij} - \mathbb{E}^2(T|A_0 = i) \\
&= -1 + 2t_i + \sum_{j=0}^{N} \left[\mathbb{E}(T^2|A_0 = j) - \mathbb{E}^2(T^2|A_0 = j)\right] p_{ij} + \sum_{j=0}^{N} \mathbb{E}^2(T^2|A_0 = j) p_{ij} - [\mathbb{E}(T|A_0 = i)]^2 \\
&= -1 + 2t_i + \sum_{j=0}^{N} v_j p_{ij} + \sum_{j=0}^{N} t_j^2 p_{ij} - t_i^2
\end{aligned} \tag{B.13}$$

En supposant que $t_0, \ldots, t_N$ soient donnés, on peut déterminer les variances conditionnelles $v_1, \ldots, v_N$ en résolvant le système d'équations

$$v_i - \sum_{j=0}^{N} v_j p_{ij} = -1 + 2t_i + \sum_{j=0}^{N} t_j^2 p_{ij} - t_i^2 \quad i = 0, \ldots, N \tag{B.14}$$

Les t_i sont eux-mêmes solutions d'une équation d'un système linéaire. En effet, un calcul similaire à (B.12) donne

$$t_i = \mathbb{E}(T|A_0 = i) = 1 + \mathbb{E}(T_1|A_0 = i) \tag{B.15}$$

$$= 1 + \sum_{j=0}^{N} \mathbb{E}(T|A_0 = j)p_{ij} \tag{B.16}$$

$$= 1 + \sum_{j=0}^{N} p_{ij}t_j \tag{B.17}$$

En posant $\underline{t} = \begin{bmatrix} t_0 & \dots & t_N \end{bmatrix}^T$, on a $\underline{t} = (I_{N+1} - P)^{-1}\mathbf{1}$, où $\mathbf{1} = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^T$. Malheureusement les formules (B.14) ne permettent pas d'étudier aisément les moments asymptotiques du temps d'absorption T.

On peut toutefois tenter d'utiliser une approximation en temps continu comme dans [39] pour trouver un équivalent de la variance asymptotique de T. Si on note $Z_n = A_n/N$, le temps d'absorption de $(A_n)_{n\geq 0}$ correspond au temps d'atteinte de Z dans $\{0, 1\}$. On a donc $t(u) = \mathbb{E}(T|Z_0 = u)$ et $v(u) = \mathrm{Var}(T \mid Z_0 = u)$.
De plus, si $Z_0 = u$ alors $Z_1 = u + U$ où $N(U - u) \sim \mathcal{B}(N, u)$. On peut réécrire (B.7) sous la forme suivante

$$\phi(u) = 1 + 2\mathbb{E}(t(u + U)) + \mathbb{E}(\phi(u + U))$$

avec $\phi(u) = \mathbb{E}(T^2|A_0 = \lfloor Nu \rfloor)$

En supposant ϕ continument dérivable 2 fois, on a :

$$\phi(u + U) = \phi(u) + U\phi'(u) + \frac{U^2}{2}\phi''(u) + \mathcal{O}(|U|^3)$$

De plus $\mathbb{E}(U) = 0$ et $\mathbb{E}(U^2) = \frac{u(1-u)}{N}$ car $N(U - u) \sim \mathcal{B}(N, u)$. Comme $\sqrt{N}U$ converge en loi vers $\mathcal{N}(0, u(1 - u))$ lorsque $N \to \infty$, on a $\mathbb{E}(U^{2p}) = \mathcal{O}(N^{-2})$.

$$\phi(u) = 1 + 2(t(u) - 1) + \phi(u) + \frac{u(1 - u)}{2N}\phi''(u) + \mathcal{O}(N^{-3/2})$$

ϕ est donc proche de la solution de l'équation différentielle

$$\phi''(u) = \frac{2N}{u(1 - u)} + 8N^2\left(\frac{\ln u}{1 - u} + \frac{\ln(1 - u)}{u}\right) \tag{B.18}$$

avec les conditions limites $\phi(0) = \phi(1) = 0$. La solution de (B.18) est

$$\phi(u) = 8N^2 u\,\mathrm{dilog}(u) + 8N^2 u \ln(u) - 8N^2\,\mathrm{dilog}(u) - 8N^2 u\,\mathrm{dilog}(1 - u)$$
$$+ 8N^2 \ln(1 - u) - 8N^2 u \ln(1 - u) - 8N^2 + (\frac{4}{3}\pi^2 + 8)N^2$$

où $u \mapsto \mathrm{dilog}(u) = \displaystyle\int_1^u \frac{\ln t}{1-t}\,dt$ est la fonction dilogarithme. Dans ce cas, la variance s'écrit :

$$\begin{aligned} v(u) =& \mathrm{Var}(T \mid A_0 = Nu) = \mathbb{E}(T^2 | A_0 = \lfloor Nu \rfloor) - \mathbb{E}^2(T | A_0 = \lfloor Nu \rfloor) \\ =& \phi(u) - t^2(u) \\ =& -2N\left[u \ln u + (1-u)\ln(1-u)\right] + 8N^2 u\,\mathrm{dilog}\,u + 8N^2 u \ln u \\ & - 8N^2\,\mathrm{dilog}\,u - 8N^2 u\,\mathrm{dilog}(1-u) + 8N^2 \ln(1-u) - 8N^2 u \ln(1-u) \\ & - 4N^2\left[u \ln u + (1-u)\ln(1-u)\right]^2 + \frac{4}{3}\pi^2 N^2 \end{aligned} \tag{B.19}$$

Annexe C

Complément au chapitre 4

Proposition 2. *Soit A et B deux matrices $d \times d$ symétriques avec B définie positive et A inversible. Alors $A^{-1} - B^{-1} < 0$ si et seulement si $A - B > 0$ (la notation $M > 0$ signifie que M est définie positive).*

Démonstration. Supposons $A - B > 0$. En utilisant le lemme d'inversion matricielle, on a

$$(B^{-1} - A^{-1})^{-1} = B + B(A - B)^{-1}B$$

et donc pour tout vecteur X de $\mathbb{R}^d$ non nul,

$$\begin{aligned} X^T(B^{-1} - A^{-1})^{-1}X &= X^TBX + X^TB(A - B)^{-1}BX \\ &= X^TBX + (BX)^T(A - B)^{-1}BX \end{aligned}$$

Le terme X^TBX est strictement positif car B est définie positive. Comme $A - B > 0$, alors $(A - B)^{-1} > 0$ et $X^TB(A - B)^{-1}BX > 0$ car le vecteur BX est non nul. Finalement, $X^T(B^{-1} - A^{-1})^{-1}X > 0$, pour tout vecteur X non nul d'où $(B^{-1} - A^{-1})^{-1} > 0$ ce qui implique que $B^{-1} - A^{-1} > 0$.
La réciproque se démontre en remplaçant A par A^{-1} et B par B^{-1}. □

Annexe D

Compléments au chapitre 5

D.1 Détection de divergence dans le cas d'un vecteur de mesure multidimensionnel

Dans ce complément, nous nous intéressons au problème de la détection de divergence lorsque la mesure est de dimension $m \geq 2$. A titre d'illustration, nous considérons l'application du filtrage particulaire au problème du pistage d'une cible dans le plan par mesures de distance et d'azimuth, abordée dans le chapitre 4, section 4.5.3.2.
Rappelons que le capteur est supposé fixe à la position $(0, 0)$ et la cible suit une trajectoire rectiligne uniforme. Le vecteur d'état contient les coordonnées de position et vitesse $x_k = [x_{k,1}, x_{k,2}, \dot{x}_{k,1}, \dot{x}_{k,2}]^T$. L'état initial de la cible est $x_{cible,0} = \begin{bmatrix} 10000 & 10000 & 40 & -40 \end{bmatrix}^T$.
Le vecteur de mesure y_k comprend la distance de la cible et l'azimuth entâchés d'un bruit additif gaussien.

$$y_k = \begin{bmatrix} \sqrt{x_{k,1}^2 + x_{k,2}^2} \\ \arctan(\frac{x_{k,2}}{x_{k,1}}) \end{bmatrix} + v_k \tag{D.1}$$

avec $v_k \sim \mathcal{N}\left(0_{2\times 1}, \mathrm{diag}(2^2, (\frac{\pi}{180})^2)\right)$.
Un filtre particulaire régularisé (RPF) avec $N = 10000$ particules, est utilisé pour le suivi de la cible. La précision du capteur fait que, si l'on utilise la densité de transition comme densité d'importance, le RPF diverge fréquemment : le taux de divergence est de 89% pour le choix de paramètres ci-dessus.
Dans la méthode de détection de divergence que nous avons suggérée dans le chapitre 5, la moyenne et la covariance de l'innovation sous l'hypothèse de non-divergence du filtre sont approximées de la même façon quelque soit la dimension du vecteur de mesure 5.39.
En revanche la définition d'un test de détection de changement de moyenne dans le cas multidimensionnel diffère du cas scalaire. En effet, dans ce cas, en supposant la suite d'innovation ϵ_k^N, $k \geq 1$ gaussienne de moyenne θ, le problème consiste à tester :

$$H_0 \ : \ \epsilon_k^N \sim \mathcal{N}(0, \Sigma_{\epsilon,k}) \text{ contre } H_1 \ : \ \epsilon_k^N \sim \mathcal{N}(\theta_1, \Sigma_{\epsilon,k})$$

où θ_1 désigne la moyenne de l'innovation sous l'hypothèse de divergence.
Les algorithmes de détection de changement de moyenne se formulent différemment selon le type d'information que l'on dispose sur la moyenne θ_1 [4] :

- θ_1 est entièrement connu,
- on connait la norme de θ_1 mais pas sa direction,
- on connait la direction de θ_1 mais pas sa norme,
- on connait une borne inférieure sur $\|\theta_1\|$ mais pas la direction de θ_1.

Dans le cas qui nous intéresse, quelques simulations nous indiquent que lorsque le RPF diverge, l'innovation n'a pas de direction privilégiée dans le plan cartésien. En revanche, selon [4], on peut tenter définir θ_1 à partir de la divergence de Kullback entre $\mathcal{N}(0, \Sigma_{\epsilon,k})$ et $\mathcal{N}(\theta_1, \Sigma_{\epsilon,k})$. Dans ce cas, θ_1 est défini comme appartenant à l'ensemble

$$\Theta_1 = \left\{\theta \mid \theta^T \Sigma_{\epsilon,k}^{-1} \theta = b^2\right\}$$

où $b > 0$.
La constante b peut être définie par exemple à partir de l'ellipsoïde de confiance correspondant à la distribution gaussienne sous l'hypothèse H_0 : on cherche b tel que

$$\mathbb{P}_{H_0}((\epsilon_k^N)^T \Sigma_{\epsilon,k}^{-1} \epsilon_k^N \leq b^2) = 0.99$$

Comme sous H_0, $\epsilon_k^N \sim \mathcal{N}(0, \Sigma_{\epsilon,k})$, $(\epsilon_k^N)^T \Sigma_{\epsilon,k}^{-1} \epsilon_k^N$ suit une loi du khi-deux à deux degrés de libertés et $b^2 = F_{\chi_2^2}^{-1}(0.99) \approx 9.2103$, $F_{\chi_2^2}$ étant la fonction de répartition du khi-deux à deux degrés de liberté.
Le problème de détection devient alors :

$$\theta = \begin{cases} 0 & \text{si } k < t_0 \\ \theta \ : \ \theta^T \Sigma_{\epsilon,k}^{-1} \theta = b^2 & \text{si } k \geq t_0 \end{cases} \tag{D.2}$$

L'algorithme de détection adapté est celui du χ^2-*CUSUM* [4, p.219]. Celui-ci ne s'écrit pas de manière récursive, ce qui rend son implémentation couteuse en temps de calcul. Cependant un algorithme proche, appelé χ^2-*CUSUM récursif* [91, 92] bien que n'étant pas strictement équivalent au χ^2-CUSUM, peut-être utilisé. Il est défini par son temps d'arrêt t_a et une statistique de test S_k selon :

$$t_a = \inf\{k \geq 1 \mid S_k \geq h\} \tag{D.3}$$

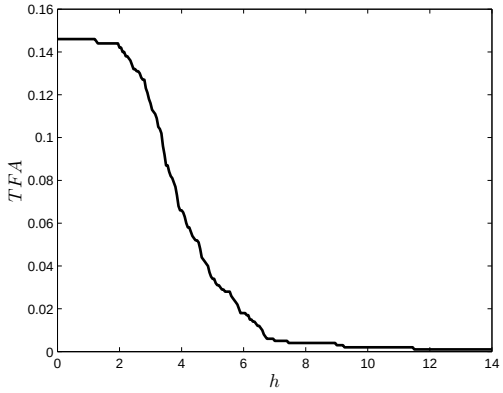
$$S_k = -n_k \frac{a^2}{2} + \log G\left(\frac{m}{2}, \frac{b^2 V_k^T \Sigma_{\epsilon,k}^{-1} V_k}{4}\right) \tag{D.4}$$

$$V_k = \mathbb{1}_{\{S_{k-1}>0\}} V_{k-1} + \epsilon_k^N \tag{D.5}$$

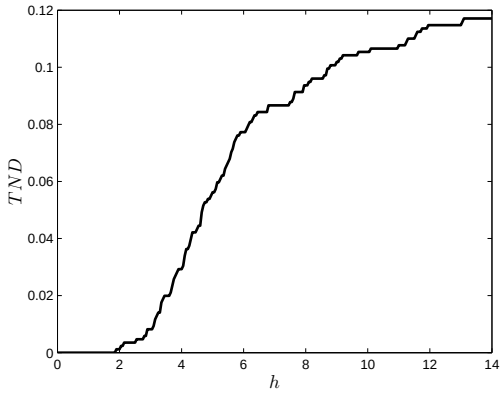
$$n_k = \mathbb{1}_{\{S_{k-1}>0\}} n_{k-1} + 1 \tag{D.6}$$

où $G(d, x) = \sum_{n=0}^{\infty} \frac{x^n}{d(d+1)\ldots(d+n-1)n!}$ est la fonction hypergéométrique ${}_0F_d$ et m désigne la dimension du vecteur de mesure (ici $m = 2$). Le test est initialisé avec $S_0 = 0$, $n_0 = 0$ et $V_0 = 0$.
Nous avons intégré cet algorithme dans l'algorithme du RPF afin de détecter une éventuelle divergence. L'innovation ϵ_k^N et sa covariance $\Sigma_{\epsilon,k}$ sont calculées selon (5.39). Les taux de fausses

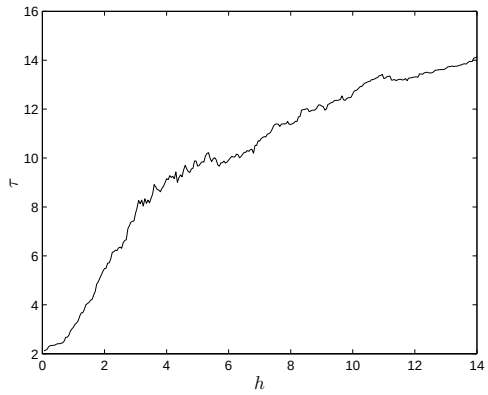
alarmes (TFA), de non-détection (TND) ainsi que le délai moyen de détection (τ) ont été établis sous la base de 1000 simulations Monte Carlo. Ces quantités sont définis plus précisément dans la section 5.6.3 et nécessitent d'estimer, s'il a lieu, l'instant réel de divergence t_0. Dans l'application du pistage, la détermination de t_0 s'est faite en considérant que le filtre a réellement divergé à l'instant t_0 si l'état vrai sort de l'ellipsoide de confiance associée à la Borne de Cramer-Rao a posteriori au seuil 99.99%, centrée sur l'état estimé, ceci pendant 5 instants successifs $t_0, \ldots, t_0 + 4$. Un grand nombre de simulations a été nécessaire pour établir ce critère et vérifier qu'il décrit précisément l'occurrence de la divergence. Les résultats sont présentés dans la figure D.1. Par exemple si l'on privilégie un faible taux de fausses alarmes ($< 1\%$), on peut constater que pour un seuil de détection $h = 8.15$, le taux de fausses alarmes est de 0.8% tandis que le taux de non-détection correspondant est de 8.4%.



(a) Taux de fausses alarmes (%)



(b) Taux de non-détection (%)



(c) Délai moyen de détection (nombre de mesures)

Figure D.1 – Performances de l'algorithme χ^2-*CUSUM récursif* pour la détection de divergence

D.2 Initialisation à partir d'une série de mesures

Le but de cette section est de rappeler une technique d'initialisation du filtre particulaire basée sur les m premières mesures. Cette technique est due à D.T. Pham (document technique non publié) et est partiellement rappelée dans [28]. Par souci de simplicité, on considère la dynamique suivante :

$$x_k = x_{k-1} + d_{k-1}, \qquad d_k = d_{k-1} + a_k \tag{D.7}$$

$$z_k = z_{k-1} + e_{k-1}, \qquad e_k = e_{k-1} + b_k \tag{D.8}$$

où x_k est la position horizontale et z_k la position verticale du mobile, d_k et e_k les vecteurs vitesse associés multipliés par période d'échantillonnage Δ et a_k et b_k les erreurs accélérométriques cumulées entre deux mesures (on suppose un mouvement uniforme). Ré-écrivons l'équation de mesure du radio-altimètre comme :

$$y_k = z_k - h(x_k) + v_k$$

On considère les m premières mesures concaténées dans le vecteur d'observation $Y = \begin{bmatrix} y_1 & \cdots, y_m \end{bmatrix}^T$.
On a

$$x_m = x_k + \sum_{j=k}^{m-1} d_j \quad d_m = d_k + \sum_{j=k+1}^{m} a_j$$

Puis,

$$x_k = x_m - \sum_{j=k}^{m-1} \left(d_m - \sum_{l=j+1}^{m} a_l \right) = x_m - (m-k)d_m + \sum_{l=k+1}^{m} (l-k)a_l$$

De même,

$$z_k = z_m - (m-k)e_m + \sum_{l=k+1}^{m} (l-k)b_l$$

D'où

$$y_k = z_m - (m-k)e_m - h\left[x_m - (m-k)d_m + \sum_{l=k+1}^{m} (l-k)a_l \right] + \sum_{l=k+1}^{m} (l-k)b_l + v_k$$

On suppose que le déplacement du mobile pendant les m premières mesures est faible, c'est-à-dire que $(m-k)d_m - \sum_{l=k+1}^{m}(l-k)a_l$ est petit. Cela permet de linéariser h autour de la m-ème position horizontale

$$h\left[x_m - (m-k)d_m + \sum_{l=k+1}^{m} (l-k)a_l \right] \approx h(x_m) - \nabla h^T(x_m) \left[m-k)d_m - \sum_{l=k+1}^{m} (l-k)a_l \right]$$

∇h étant le gradient de h. Il vient

$$y_k = z_m - h(x_m) + (m-k)\left[\nabla h^T(x_m)d_m - e_m \right] + \sum_{l=k+1}^{m} \left[b_m - \nabla h^T(x_m)a_l \right] + v_k$$

Ceci donne l'équation d'observation reliant le vecteur des m premières observations avec l'état x_m :

$$Y = H_1 \begin{bmatrix} z_m \\ d_m \\ e_m \end{bmatrix} - \begin{bmatrix} h(x_m) \\ \vdots \\ h(x_m) \end{bmatrix} + V \tag{D.9}$$

où H_1 est la matrice

$$\begin{bmatrix} -1 & (m-1)\nabla h^T(x_m) & (m-1) \\ \vdots & \vdots & \vdots \\ -1 & (m-1)\nabla h^T(x_m) & 1 \\ -1 & 0 & 0 \end{bmatrix}$$

et V le vecteur

$$V = \begin{bmatrix} v_1 + \sum_{l=2}^{m}(l-1)[b_l - a_l \nabla h^T(x_l)] \\ \vdots \\ v_{m-1} + [b_m - a_m \nabla h^T(x_m)] \\ v_m \end{bmatrix}$$

Bibliographie

[1] *Global Positioning System. Standard Positioning Service performance standard*, September 2008.

[2] A. Almagbile, J. Wang, and W. Ding, *Evaluating the performance of adaptive Kalman filter methods in GPS/INS integration*, CPGPS **9** (2010), 33Ű40.

[3] D. Alspach and H. Sorenson, *Nonlinear Bayesian estimation using Gaussian sum approximation*, IEEE Transactions on Automatic Control **17** (1972), 439–448.

[4] M. Basseville and I.V. Nikiforov, *Detection of abrupt changes : Theory and application*, Prentice Hall, 1993.

[5] D.O. Benson, *A comparison of two approaches to pure-inertial and doppler-inertial error analysis*, IEEE Transactions on Aerospace and Electronic Systems **AES-11** (1975), no. 4, 447–455.

[6] N. Bergman, *Recursive Bayesian Estimation : Navigation and Tracking Applications*, Thèse de doctorat, Linköping University, 1999.

[7] N. Bergman and L. Ljung, *Point-mass filter and cramer-rao bound for terrain-aided navigation*, Proceedings of the 36th IEEE Conference on Decision and Control, vol. 1, 1997, pp. 565–570.

[8] F.M. Boland and H. Nicholson, *Control of divergence in Kalman filters*, Electronics Letters **12** (1976), no. 15, 367–369.

[9] L. Breiman, W. Meisel, and E. Purcell, *Variable Kernel Estimates of Multivariate Densities*, Technometrics **19** (1977), no. 2, 135–144.

[10] C. G. Broyden, *The convergence of a class of double-rank minimization algorithms*, Journal of the Institute of Mathematics and Its Applications **6** (1970), 76–90.

[11] R.S. Bucy and K.D. Senne, *Digital synthesis of non-linear filters*, Automatica **7** (1971), no. 3, 287–298.

[12] A. Bugeau and P. Pérez, *Bandwidth selection for kernel estimation in mixed multi-dimensional spaces*, Tech. report, IRISA, 2007.

[13] K. P. Burnham and D. R. Anderson, *Model selection and multimodel inference : a practical information-theoretic approach*, 2 ed., Springer, 2002.

[14] Z. Cai, F. Le Gland, and H. Zhang, *An adaptive local grid refinement method for nonlinear filtering*, Rapport de recherche RR-2679, INRIA, 1995.

[15] C. P. Casella, G.and Robert, *Rao-Blackwellisation of Sampling Schemes*, Biometrika **83** (1996), no. 1, 81–94.

[16] G. Celeux, F. Forbes, and N. Peyrard, *EM procedures using mean field-like approximations for Markov model-based image segmentation*, Pattern Recognition **36** (2003), no. 1, 131–144.

[17] R. Chen and J.S. Liu, *Mixture Kalman filters*, J. R. Statist. Soc. B **62** (2000), 493–508.

[18] Z. Chen, *Bayesian Filtering : From Kalman Filters to Particle Filters, and Beyond*, Tech. report, McMaster University, 2003.

[19] Y. Cheng, *Mean shift, mode seeking, and clustering*, IEEE Transactions on Pattern Analysis and Machine Intelligence **17** (1995), no. 8, 790 –799.

[20] D. Comaniciu, *An algorithm for data-driven bandwidth selection*, IEEE Transactions on Pattern Analysis and Machine Intelligence **25** (2003), no. 2, 281–288.

[21] D. Comaniciu and P. Meer, *Mean shift : a robust approach toward feature space analysis*, IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (2002), no. 5, 603–619.

[22] D. Comaniciu, V. Ramesh, and P. Meer, *The variable bandwidth mean shift and data-driven scale selection*, in Proceedings of the 8th International Conference on Computer Vision, 2001, pp. 438–445.

[23] J. Cornebise, E. Moulines, and J. Olsson, *Adaptive methods for sequential importance sampling with application to state space models*, Statistics and Computing **18** (2008), no. 4, 461 – 480 (English).

[24] D. Crisan, *Exact rates of convergence for a branching particle approximation to the solution of the zakai equation*, Annals of Probability **32** (2003), 819 – 838.

[25] D. Crisan, P. Del Moral, and T. J. Lyons, *Interacting particle systems approximations of the Kushner Stratonovitch equation*, Advances in Applied Probability (1999), 819 – 838.

[26] D. Crisan and A. Doucet, *A survey of convergence results on particle filtering methods for practitioners*, IEEE Transactions on Signal Processing **50** (2002), no. 3, 736 – 746.

[27] E.P. Cunningham, *Single-parameter terrain classification for terrain following*, Journal of Aircraft **17** (1980), 909 – 914.

[28] K. Dahia, *Nouvelles méthodes en filtrage particulaire. Application au recalage de navigation inertielle par mesures altimétriques*, Thèse de doctorat, Université Joseph Fourier, 2005.

[29] M. H. DeGroot, *Optimal statistical decisions*, Wiley, 2005.

[30] P. Del Moral, *Non-linear filtering using random particles*, Theo. Prob. App. **40** (1995), 690–701.

[31] _____, *Feynman-Kac formulae : genealogical and interacting particle systems with applications*, Springer, 2004.

[32] P. Del Moral and L. Miclo, *Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering*, Séminaire de Probabilités XXXIV, Lecture Notes in Mathematics, vol. 1729, Springer, 2000, pp. 1–145.

[33] P. Del Moral, G. Rigal, and Salut G., *Estimation et commande optimale non-linéaire*, Tech. report, LAAS/CNRS, 1992.

[34] L. Devroye, *Non-uniform random variate generation*, Springer-Verlag, 1986.

[35] R. Douc and O. Cappe, *Comparison of resampling schemes for particle filtering*, Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis (ISPA 2005), 2005, pp. 64–69.

[36] A. Doucet, S. Godsill, and C. Andrieu, *On sequential Monte Carlo sampling methods for Bayesian filtering*, Statistics and Computing **10** (2000), no. 3, 197–208.

[37] A. Doucet, De Freitas N., and N. Gordon, *Sequential Monte Carlo methods in practice*, ch. Improving Regularized Particle Filters, pp. 247–271, Springer, 2001.

[38] R. Durrett, *Stochastic calculus : a practical introduction*, vol. 6, CRC press, 1996.

[39] W. J. Ewens, *Mathematical population genetics*, Springer-Verlag, 1979.

[40] R. Fitzgerald, *Divergence of the Kalman filter*, IEEE Transactions on Automatic Control **16** (1971), no. 6, 736 – 747.

[41] M. Flament, *Apport du filtrage particulaire au recalage altimétrique dans un contexte de navigation hybridée*, Thèse de doctorat, Université Paris-Sud 11, 2009.

[42] D. Fox, S. Thrun, F. Dellaert, and W. Burgard, *Particle filters for mobile robot localization*, Sequential Monte Carlo Methods in Practice, Springer Verlag, 2001.

[43] D. Freedman and P. Kisilev, *KDE paring and a faster mean shift algorithm*, SIAM J Imaging Sciences **3** (2010), no. 4, 878–903.

[44] K. Fukunaga, *Introduction to statistical pattern recognition (2nd ed.)*, Academic Press Professional, Inc., 1990.

[45] K. Fukunaga and L. Hostetler, *The estimation of the gradient of a density function, with applications in pattern recognition*, IEEE Transactions on Information Theory **21** (1975), no. 1, 32 – 40.

[46] B. Georgescu, I. Shimshoni, and P. Meer, *Mean shift based clustering in high dimensions : a texture classification example*, Proceedings of the 9th IEEE International Conference on Computer Vision **1** (2003), no. Iccv, 456–463.

[47] J. Georgy, A. Noureldin, and G.R. Mellema, *Clustered mixture particle filter for underwater multitarget tracking in multistatic active sonobuoy systems*, IEEE Transactions on Systems, Man, and Cybernetics, Part C : Applications and Reviews **42** (2012), no. 4, 547–560.

[48] J. Geweke, *Bayesian inference in econometric models using Monte Carlo integration*, Econometrica **57** (1989), no. 6, 1317–39.

[49] Audrey Giremus, *Apport des méthodes de filtrage particulaire pour la navigation GPS*, Thèse de doctorat, Ecole Nationale Supérieure de l'Aéronautique et de l'Espace, Toulouse, France, décembre 2005.

[50] N.J. Gordon, D.J. Salmond, and A.F.M. Smith, *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*, IEE Proceedings F Radar and Signal Processing **140** (1993), no. 2, 107 –113.

[51] F. Gustafsson, *Particle filter theory and practice with positioning applications*, IEEE Aerospace and Electronic Systems Magazine **25** (2010), no. 7, 53–82.

[52] J. E. Handschin and D. Q. Mayne, *Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering*, International Journal of Control **9** (1969), no. 5, 547– 559.

[53] J.E. Handschin, *Monte Carlo techniques for prediction and filtering of non-linear stochastic processes*, Automatica **6** (1970), no. 4, 555 – 563.

[54] M.R. Hestenes and E. Stiefel, *Methods of Conjugate Gradients for Solving Linear Systems*, Journal of Research of the National Bureau of Standards **49** (1952), no. 6, 409–436.

[55] N. J. Higham, *Computing a nearest symmetric positive semidefinite matrix*, Linear Algebra and Its Applications **103** (1988), 103–118.

[56] J. Hollowell, *Heli/SITAN : a terrain referenced navigation algorithm for helicopters*, Position Location and Navigation Symposium, 1990, pp. 616–625.

[57] L. D. Hostetler, *Optimal terrain-aided navigation systems*, AIAA Guidance and Control Conference, 1978.

[58] C. Hue, J.-P. Le Cadre, and P. Pérez, *Tracking multiple objects with particle filtering*, 2000.

[59] M. Isard and J. MacCormick, *Bramble : a Bayesian multiple-blob tracker*, Proceedings of the 8h IEEE International Conference on Computer Vision, vol. 2, 2001, pp. 34–41 vol.2.

[60] A. K. Jain, M. N. Murty, and P. J. Flynn, *Data clustering : a review*, ACM Computing Surveys **31** (1999), no. 3, 264–323.

[61] M. Joerger and B. Pervan, *Kalman filter-based integrity monitoring against sensor faults*, Journal of Guidance, Control and Dynamics **36** (2013), 349–361.

[62] S. Julier, J. Uhlmann, and H.F. Durrant-Whyte, *A new method for the nonlinear transformation of means and covariances in filters and estimators*, IEEE Transactions on Automatic Control **45** (2000), no. 3, 477–482.

[63] S.J. Julier and J.K. Uhlmann, *A new extension of the Kalman filter to nonlinear systems*, Proceedings of AeroSense : The 11th International Symposium on Aerospace/Defense Sensing, Simulations and Controls, 1997.

[64] A. Kallapur, S. Anavatti, and M. Garratt, *Extended and unscented Kalman filters for attitude estimation of an unmanned aerial vehicle*, Proceedings of the 27th IASTED International Conference on Modelling, Identification and Control, MIC '08, ACTA Press, 2008, pp. 498–502.

[65] R. E. Kalman, *A new approach to linear filtering and prediction problems*, Transactions of the ASME–Journal of Basic Engineering **82** (1960), no. Series D, 35–45.

[66] R. E. Kass, L. Tierney, and J. B. Kadane, *Asymptotics in Bayesian Computation*, Bayesian Statistics 3, Oxford University Press, 1988, pp. 261–278.

[67] O. King and D.A. Forsyth, *How does condensation behave with a finite number of samples ?*, Computer Vision - ECCV 2000, Lecture Notes in Computer Science, vol. 1842, Springer Berlin Heidelberg, 2000, pp. 695–709 (English).

[68] G. Kitagawa, *Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models*, Journal of Computational and Graphical Statistics **5** (1996), no. 1, 1–25.

[69] A. Kong, J. S. Liu, and W.H. Wong, *Sequential imputations and Bayesian missing data problems*, Journal of the American Statistical Association **89** (1994), 278–288.

[70] S. Kotz and S. Nadarajah, *Multivariate t distributions and their applications*, Cambridge University Press, 2004.

[71] H.J. Kushner and P. Dupuis, *Numerical methods for stochastic control problems in continuous time*, Springer Verlag, 1992.

[72] F. Le Gland, C. Musso, and N. Oudjane, *An analysis of regularized interacting particle methods for nonlinear filtering*, Proceedings of the 3rd IEEE European Workshop on Computer-Intensive Methods in Control and Signal Processing, 1998, pp. 167–174.

[73] E. L. Lehmann and George Casella, *Theory of Point Estimation (Springer Texts in Statistics)*, 2nd ed., Springer, August 1998.

[74] J.J. Leonard and H.F. Durrant-Whyte, *Mobile robot localization by tracking geometric beacons*, IEEE Transactions on Robotics and Automation **7** (1991), no. 3, 376–382.

[75] J. Lin, *Divergence measures based on the shannon entropy*, IEEE Transactions on Information Theory **37** (1991), no. 1, 145–151.

[76] J.S. Liu and R. Chen, *Sequential Monte Carlo methods for dynamic systems*, Journal of the American Statistical Association **93** (1998), 1032–1044.

[77] Z. Liu, Z.and Shi, M. Zhao, and W. Xu, *Adaptive dynamic clustered particle filtering for mobile robots global localization*, Journal of Intelligent and Robotic Systems **53** (2008), no. 1, 57–85 (English).

[78] G Lorden, *Procedures for reacting to a change in distribution*, The Annals of Mathematical Statistics **42** (1971), no. 6, 1897–1908.

[79] J. MacCormick and A. Blake, *A probabilistic exclusion principle for tracking multiple objects*, Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 1, 1999, pp. 572–578 vol.1.

[80] J. MacQueen, *Some methods for classification and analysis of multivariate observations*, Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability - Vol. 1 (L. M. Le Cam and J. Neyman, eds.), University of California Press, Berkeley, CA, USA, 1967, pp. 281–297.

[81] A. Maier and G. F. Trommer, *Comparison of centralized and decentralized Kalman filter for sar/trn/gps/ins integration*, Proceedings of the 2010 International Technical Meeting of The Institute of Navigation, 2010, pp. 74–80.

[82] P.W. McBurney, *A robust approach to reliable real-time Kalman filtering*, IEEE Position Location and Navigation Symposium - A Decade of Excellence in the Navigation Sciences, 1990, pp. 549–556.

[83] G. McLachlan and D. Peel, *Finite mixture models*, Wiley series in probability and statistics. Applied probability and statistics section, Wiley, 2004.

[84] R.K. Mehra and J. Peschon, *An innovations approach to fault detection and diagnosis in dynamic systems*, Automatica **7** (1971), no. 5, 637 – 640.

[85] A. Milstein, J. N. Sánchez, and E. T. Williamson, *Robust global localization using clustered particle filtering*, 18th National Conference on Artificial Intelligence (Menlo Park, CA, USA), American Association for Artificial Intelligence, 2002, pp. 581–586.

[86] A. H. Mohamed and K. P. Schwarz, *Adaptive Kalman filtering for INS/GPS*, Journal of Geodesy **73** (1999), 193–203.

[87] G. V. Moustakides, *Optimal stopping times for detecting changes in distributions*, The Annals of Statistics **14** (1986), 1379–1387.

[88] A. Murangira, C. Musso, K. Dahia, and J. Allard, *Robust regularized particle filter for terrain navigation*, Proceedings of the 14th International Conference on Information Fusion (FUSION), july 2011, pp. 1 –8.

[89] C. Musso and N. Oudjane, *Regularization schemes for branching particle systems as a numerical solving method of the nonlinear filtering problem*, Proceedings of the Irish Signals Systems Conference, 1998.

[90] C. Musso, P.B. Quang, and F. Le Gland, *Introducing the laplace approximation in particle filtering*, Proceedings of the 14th International Conference on Information Fusion, july 2011, pp. 1 –8.

[91] I. V. Nikiforov, *On the first-order optimality of an algorithm for detection of a fault in the vector case*, Automation and Remote Control **55** (1994), 66–82.

[92] _____, *A suboptimal quadratic change detection scheme*, IEEE Transactions on Information Theory **46** (2000), 2095–2107.

[93] P. J. Nordlund, *Efficient Estimation and Detection Methods for Airborne Applications*, Thèse de doctorat, Linköping University, 2009.

[94] P.-J. Nordlund and F. Gustafsson, *Marginalized particle filter for accurate and reliable terrain-aided navigation*, IEEE Transactions on Aerospace and Electronic Systems **45** (2009), no. 4, 1385 –1399.

[95] K. Okuma, A. Taleghani, N. Freitas, J.J. Little, and D.G. Lowe, *A boosted particle filter : Multitarget detection and tracking*, European Conference on Computer Vision **3021** (2004), 28 – 39.

[96] N. Oudjane, *Stabilité et approximations particulaires en filtrage non linéaire. Application au pistage*, Thèse de doctorat, Université de Rennes 1, 2000.

[97] N. Oudjane and C. Musso, *Progressive correction for regularized particle filters*, Proceedings of the Third International Conference on Information Fusion, vol. 2, 2000, pp. THB2/10–THB2/17 vol.2.

[98] E.S. Page, *Continuous inspection schemes*, Biometrika **41** (1954), no. 1/2, 100–115.

[99] E.J. Pauwels and G. Frederix, *Finding salient regions in images : Nonparametric clustering for image segmentation and grouping*, Computer Vision and Image Understanding **75** (1999), 73 – 85.

[100] N. Peach, *Bearings-only tracking using a set of range-parameterised extended Kalman filters*, IEE Proceedings on Control Theory and Applications **142** (1995), no. 1, 73–80.

[101] H. Permuter, J. Francos, and I. Jermyn, *A study of Gaussian mixture models of color and texture features for image classification and segmentation*, Pattern Recognition **39** (2006), no. 4, 695 – 706.

[102] D. T. Pham, *Stochastic methods for sequential data assimilation in strongly nonlinear systems*, Monthly Weather Review **129** (2001), no. 5, 1194–1207.

[103] D.-T. Pham, K. Dahia, and C. Musso, *A kalman-particle kernel filter and its application to terrain navigation*, Proceedings of the Sixth International Conference on Information Fusion, vol. 2, 2003, pp. 1172–1179.

[104] J. Picard, *Approximation of nonlinear filtering problems and order of convergence*, Filtering and Control of Random Processes, Lecture Notes in Control and Information Sciences, vol. 61, Springer, 1984, pp. 219–236.

[105] M. K. Pitt and N. Shephard, *Filtering via Simulation : Auxiliary Particle Filters*, Journal of the American Statistical Association **94** (1999), no. 446, 590–599.

[106] J.-C. Radix, *La navigation par inertie*, Presses Universitaires de France, 1967.

[107] D. B. Rubin, *Using the SIR algorithm to simulate posterior distributions*, Bayesian Statistics 3, Oxford University Press, 1988.

[108] M. Rudemo, *Empirical choice of histograms and kernel density estmators*, Scandinavian Journal of Statistics **9** (1982), 65 – 78.

[109] N. Ruixin, P.K. Varshney, M. Alford, A. Bubalo, E. Jones, and M. Scalzo, *Curvature nonlinearity measure and filter divergence detector for nonlinear tracking problems*, 11th International Conference on Information Fusion, 2008.

[110] T. Schon, F. Gustafsson, and P.-J. Nordlund, *Marginalized particle filters for mixed linear/nonlinear state-space models*, IEEE Transactions on Signal Processing **53** (2005), no. 7, 2279–2289.

[111] A.N. Shiryaev, *The problem of the most rapid detection of a disturbance in a stationary process*, Soviet Math. Dokl. **2** (1961), 795–799.

[112] B. W. Silverman, *Density estimation for statistics and data analysis*, Monographs on Statistics and Applied Probability, vol. 26, Chapman and Hall, 1986.

[113] I. Smal, K. Draegestein, N. Galjart, W. Niessen, and E. Meijering, *Particle filtering for multiple object tracking in dynamic fluorescence microscopy images : Application to microtubule growth analysis*, IEEE Transactions on Medical Imaging **27** (2008), no. 6, 789–804.

[114] P. Tichavsky, C.H. Muravchik, and A. Nehorai, *Posterior cramer-rao bounds for discrete-time nonlinear filtering*, IEEE Transactions on Signal Processing **46** (1998), no. 5, 1386–1396.

[115] L. Tierney, R. E. Kass, and J. B. Kadane, *Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions*, Journal of the American Statistical Association **84** (1989), no. 407, 710–716.

[116] B. Turgut and R.P. Martin, *Restarting particle filters : An approach to improve the performance of dynamic indoor localization*, IEEE Global Telecommunications Conference, 2009, pp. 5753–5759.

[117] R. Van der Merwe, N. de Freitas, and E. Doucet, A.and Wan, *The Unscented Particle Filter*, Advances in Neural Information Processing Systems 13, November 2000.

[118] J. Vermaak, A. Doucet, and P. Perez, *Maintaining multimodality through mixture tracking*, Proceedings of the Ninth IEEE International Conference on Computer Vision, oct. 2003, pp. 1110 –1116 vol.2.

[119] A. Wald, *Sequential Analysis*, Wiley, 1947.

[120] M. P. Wand and M. C. Jones, *Kernel Smoothing*, 1 ed., Monographs on Statistics & Applied Probability, Chapman and Hall/CRC, December 1994.

[121] N. Whiteley and A. M. Johansen, *Auxiliary particle filtering : recent developments*, Bayesian time series models, Cambridge University Press, Cambridge, 2011.

[122] M. Woodrofe, *On choosing a delta-sequence*, Ann. Math. Statist. **41** (1970), 1665 – 1671.

[123] Chunxia Xiao and Meng Liu, *Efficient mean-shift clustering using Gaussian kd-tree*, Computer Graphics Forum **29** (2010), no. 7, 2065–2073.

[124] R. Xu and D. Wunsch, *Survey of clustering algorithms*, IEEE Transactions on Neural Networks **16** (2005), no. 3, 645–678.

[125] M. Zakai, *On the optimal filtering of diffusion processes*, Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete **11** (1969), no. 3, 230–243.

[126] L. Zhibin, S. Zongying, Z. Mingguo, and X. Wenli, *Mobile robots global localization using adaptive dynamic clustered particle filters*, IEEE/RSJ International Conference on Intelligent Robots and Systems, 2007, pp. 1059–1064.

Achille MURANGIRA

Doctorat : Optimisation et Sûreté des Systèmes

Année 2014

Nouvelles approches en filtrage particulaire. Application au recalage de la navigation inertielle

Les travaux présentés dans ce mémoire de thèse concernent le développement et la mise en oeuvre d'un algorithme de filtrage particulaire pour le recalage de la navigation inertielle par mesures altimétriques. Le filtre développé, le MRPF (Mixture Regularized Particle Filter), s'appuie à la fois sur la modélisation de la densité a posteriori sous forme de mélange fini, sur le filtre particulaire régularisé ainsi que sur l'algorithme mean-shift clustering. Nous proposons également une extension du MRPF au filtre particulaire Rao-Blackwellisé appelée MRBPF (Mixture Rao-Blackwellized Particle Filter). L'objectif est de proposer un filtre adapté à la gestion des multimodalités dues aux ambiguïtés de terrain. L'utilisation des modèles de mélange fini permet d'introduire un algorithme d'échantillonnage d'importance afin de générer les particules dans les zones d'intérêt. Un second axe de recherche concerne la mise au point d'outils de contrôle d'intégrité de la solution particulaire. En nous appuyant sur la théorie de la détection de changement, nous proposons un algorithme de détection séquentielle de la divergence du filtre. Les performances du MRPF, MRBPF, et du test d'intégrité sont évaluées sur plusieurs scénarios de recalage altimétrique.

Mots clés : Monte-Carlo, méthode de - navigation par inertie, systèmes de - radioaltimètres – rupture (statistique).

New Particle Filtering Methods. Application to Inertial Navigation Update

This thesis deals with the development of a mixture particle filtering algorithm for inertial navigation update via radar-altimeter measurements. This particle filter, the so-called MRPF (Mixture Regularized Particle Filter), combines mixture modelling of the posterior density, the regularized particle filter and the mean-shift clustering algorithm. A version adapted to the Rao-Blackwellized particle filter, the MRBPF (Mixture Rao-Blackwellized Particle Filter), is also presented. The main goal is to design a filter well suited to multimodal densities caused by terrain amibiguity. The use of mixture models enables us to introduce an alternative importance sampling procedure aimed at proposing samples in the high likelihood regions of the state space. A second research axis is concerned with the development of particle filtering integrity monitoring tools. A novel particle filter divergence sequential detector, based on change detection theory, is presented. The performances of the MRPF, MRBPF and the divergence detector are reported on several terrain navigation scenarios.

Keywords: Monte Carlo method - inertial navigation systems – radioaltimeters - change-point problems.

Thèse réalisée en partenariat entre :

ONERA
THE FRENCH AEROSPACE LAB

utt
université de technologie
Troyes

Ecole Doctorale "Sciences et Technologies"