



HAL
open science

Apprentissage non-supervisé dans les modèles linéaires gaussiens. Application à la biométrie dynamique de l'iris

Valérian Nemesin

► To cite this version:

Valérian Nemesin. Apprentissage non-supervisé dans les modèles linéaires gaussiens. Application à la biométrie dynamique de l'iris. Optique / photonique. Ecole Centrale Marseille, 2014. Français. NNT : 2014ECDM0006 . tel-01137587

HAL Id: tel-01137587

<https://theses.hal.science/tel-01137587>

Submitted on 31 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ÉCOLE CENTRALE MARSEILLE

École doctorale - Physique et Sciences de la Matière (ED 352)

Institut Fresnel - Équipe GSM

THÈSE DE DOCTORAT

pour obtenir le grade de
DOCTEUR de l'ÉCOLE CENTRALE MARSEILLE

Discipline: **Optique, Photonique et Traitement d'image**

**Apprentissage non-supervisé
dans les modèles linéaires gaussiens.
Application à la biométrie dynamique de l'iris.**

par

Valérian Nêmesin

Directeur de thèse: Stéphane Derrode

soutenue le 30 septembre 2014
devant le jury composé de

L. Chen	Professeur, École Centrale de Lyon	Examineur
W. Pieczynski	Professeur, Télécom SudParis	Rapporteur
E. Duflos	Professeur, École Centrale de Lille	Rapporteur
M.V. Serfaty	Direction générale à l'armement	Examineur
J. Blanc-Talon	Direction générale à l'armement	Examineur
S. Bourennane	Professeur, École Centrale Marseille	Examineur
S. Derrode	MCF/HDR, École Centrale Marseille	Directeur de thèse

Résumé

La famille de modèles dite des filtres de Kalman permet d'estimer les états d'un système dynamique à partir d'une série de mesures incomplètes ou bruitées. Malgré leur relative simplicité de modélisation, ces filtres sont utilisés dans un large spectre scientifique dont le radar, la vision, et les communications. Ce succès repose, pour l'essentiel, sur l'existence d'algorithmes de filtrage et de lissage exacts et rapides, *i.e.* linéaires au nombre d'observations, qui minimisent l'erreur quadratique moyenne.

Dans cette thèse, nous nous sommes intéressés au filtre de Kalman couple. Celui-ci intègre, par rapport au modèle original, de nouvelles possibilités d'interactions entre états cachés et observations, tout en conservant des algorithmes exacts et rapides dans le cas linéaire et gaussien. Nous étudions plus particulièrement le problème de l'estimation non supervisée et robuste des paramètres d'un filtre de Kalman couple à partir d'observations en nombre limité. Le manuscrit décrit ainsi plusieurs algorithmes d'apprentissage par estimation du maximum de vraisemblance selon le principe EM (*Expectation-Maximization*). Ces algorithmes originaux permettent d'intégrer des contraintes a priori sur les paramètres du système étudié, comme expressions de connaissances partielles sur la physique de l'application ou sur le capteur. Ces systèmes contraints réduisent l'ambiguïté liée au problème d'identifiabilité du filtre de Kalman couple lors de l'estimation des paramètres. Ils permettent également de limiter le nombre de maxima locaux de la fonction de vraisemblance en réduisant la dimension de l'espace de recherche, et ainsi évitent parfois le piègeage de l'algorithme EM.

Il est important de noter que l'ensemble des algorithmes proposés dans ce manuscrit s'applique directement au filtre de Kalman original, comme cas particulier du filtre de Kalman couple. Tous les algorithmes sont rendus robustes par la propagation systématique de racines-carrés des matrices de covariance au lieu des matrices de covariance elles-mêmes, permettant ainsi d'éviter les difficultés numériques bien connues liées à la perte de positivité et de symétrie de ces matrices. Ces algorithmes robustes sont finalement évalués et comparés dans le cadre d'une application de biométrie de l'iris à partir de vidéos. Le suivi de la pupille est exploitée pour enrôler et identifier en temps-réel une personne grâce à son iris-code.

Abstract

The family of Kalman filter model allows to estimate the states of a dynamical system from a set of observations. Despite a simple model, these filters are used in a large field of applications : Radar, vision and communications. The success is mainly based on the existence of exact smoothing or filtering algorithms, *i.e.* linear to the number of observations and which minimize the mean square error.

In this thesis, we are concerned about the pairwise Kalman filter. This filter adds from the original model, new interactions between hidden states and observations while keeping exact algorithms in the case of linear and Gaussian models. We studied particularly the problem of the unsupervised and robust estimation of a pairwise Kalman filter parameters from a limited set of observations. The manuscript describes several learning algorithms by the estimation of the likelihood maximum according to EM (*Expectation-Maximization*) principle. These original algorithms allow to embed a-priori constraints on studied system parameters, like a knowledge about physical or sensors. These constrained systems reduce the ambiguity, linked to identifiability issue of the pairwise Kalman filter during the parameter estimation. They allow also to limit the number of local maxima of likelihood function with the reduction of the dimension of search space and avoid sometime the trapping of EM algorithm.

It is important to note that all proposed algorithm of this manuscript can be applied to the original Kalman filter, as a particular pairwise Kalman filter. All algorithm are made robust by the propagation of square root matrices instead of the covariance matrices, which allows to limit the numerical issues, linked to the loses of symmetry or positivity of these matrices. These algorithm are finally evaluated and compared in the case of an iris biometry application from video sequences. Pupil tracking is used to enroll and recognize in real-time a person thanks to its iris-code.

Table des matières

Résumé	iii
Abstract	iv
Table des matières	v
Introduction	1
1 Autours du filtre de Kalman	2
2 Filtre de Kalman non-supervisé	7
3 Contributions de cette thèse	8
1 Filtre de Kalman couple robuste	13
1 Algorithme EM	14
1.1 Étape E	14
1.2 Étape M	17
2 Implémentation du filtre robuste	18
2.1 Étape E	19
2.2 Étape M	23
3 Illustration de la robustesse	24
4 Conclusion	25
2 Filtres de Kalman couple équivalents	27
1 Propriété	29
2 Paramètres équivalents : 3 cas intéressants	30
2.1 Cas α : contraintes sur la matrice de covariance du bruit	31
2.2 Cas β : contraintes sur les matrices d'états	32
2.3 Cas γ : contraintes sur les matrices d'observations	32
3 Évaluations	34
3.1 Filtre équivalent de type α	34
3.2 Filtre équivalent de type β	37
3.3 Filtre équivalent de type γ	39
4 Conclusion	42
3 Filtre de Kalman couple contraint	43
1 Expression des contraintes	44
1.1 Décorrélacion du bruit	44
1.2 Blocs de la matrice de transition	45

1.3	Blocs de la matrice de covariance	47
2	Ré-estimation robuste	49
2.1	Calcul de la matrice de transition	50
2.2	Calcul de la matrice de covariance du bruit	51
3	Évaluations	54
3.1	Comparaison avec le filtre équivalent de type α	54
3.2	Comparaison avec le filtre équivalent de type β	58
3.3	Comparaison avec le filtre équivalent de type γ	61
3.4	Comparaison entre filtres contraints	64
4	Conclusion	69
4	Discussion	71
1	Données théoriques	71
1.1	Données suivant le modèle du filtre de Kalman	72
1.2	Données suivant le modèle du filtre de Kalman couple	73
2	Données simulées	76
2.1	Signal sinusoïdal	78
2.2	Signal asservi	82
3	Suivi de la pupille dans une séquence vidéo	84
4	Conclusion	90
5	Biométrie dynamique de l'iris	91
1	Principe	91
1.1	Identification et critères d'évaluation	92
1.2	Modalité vidéo	93
2	Enrôlement et fusion des iris-codes	94
2.1	Enrôlement	94
2.2	Fusion	97
3	Enrôlement de l'iris temps-réel	98
3.1	Suivi de la pupille	98
3.2	Sélection à la volée des meilleures images	100
4	Évaluation des performances sur la base MBGC	102
4.1	Influence du nombre d'iris-codes fusionnés	102
4.2	Influence des bits fragiles	105
4.3	Comparaison avec d'autres méthodes de fusion	107
5	Conclusion	108
	Conclusion	109
	Bibliographie	113
	Annexe	119
1	Preuves du chapitre 2	119
1.1	Démonstration 1	119

1.2	Démonstration 2	119
2	Preuves du chapitre 3	120
2.1	Démonstration 3	120
2.2	Démonstration 4	121
2.3	Démonstration 5	121
2.4	Démonstration 6	122
2.5	Démonstration 7	123
2.6	Démonstration 8	123
2.7	Démonstration 9	124
3	Filtres de Kalman particuliers	124
3.1	Filtre de Kalman équivalent	125
3.2	Filtre de Kalman contraint	125
Liste des symboles		127
Table des figures		129
Liste des tableaux		133

Introduction

Dans cette thèse, nous considérons le problème, inclus dans le filtrage statistique bayésien, qui consiste à estimer des états cachés $\mathbf{x} = \{\mathbf{x}_n\}_{n \in \mathbb{N}}$ à partir d'un ensemble d'observations $\mathbf{y} = \{\mathbf{y}_n\}_{n=0:N}$. La famille de modèles dite des filtres de Kalman permet d'estimer les états d'un système dynamique linéaire gaussien à partir d'une série de mesures incomplètes ou bruitées. Malgré leur relative simplicité de modélisation, ces filtres sont utilisés dans un large spectre scientifique dont

- la poursuite d'objets mobiles [58],
- la restauration de signaux bruités [61],
- l'analyse de marchés financiers [56] *etc*

Ce succès repose, pour l'essentiel, sur l'existence d'algorithmes de filtrage et de lissage exacts et rapides, *i.e* linéaires au nombre d'observations, qui minimisent l'erreur quadratique moyenne.

Pour s'affranchir des hypothèses parfois trop contraignantes de linéarité et/ou de gaussiannité sous-jacentes au modèle original, des méthodes par approximations déterministes ou stochastiques ont ensuite été proposées, parmi lesquelles on peut citer le filtre de Kalman étendu, le filtre de Kalman sans parfum ou bien encore le filtrage particulaire. La première section de cette introduction propose une synthèse de ces différents modèles.

Le modèle de Kalman est un modèle paramétrique et la connaissance de ses paramètres est requise pour restaurer les données. Or, il est parfois délicat de régler ces paramètres manuellement pour un problème ou une application donnée. La seconde section de cette introduction décrit certains algorithmes d'apprentissage non-supervisés des paramètres pour le filtre de Kalman.

Dans cette thèse, nous nous sommes intéressés au modèle linéaire du filtre de Kalman couple qui conserve des équations de filtrage et de lissage exacts et pour lequel il est possible d'implémenter un algorithme d'apprentissage non-supervisé. La troisième section de cette introduction expose la problématique de la thèse et décrit nos principales contributions méthodologiques et numériques sur ce filtre.

1 Autours du filtre de Kalman

Dans le cadre de la modélisation d'un signal naturel, le plus simple est de considérer un bruit aléatoire additif entre observations \mathbf{y}_n (de dimension n_y) et états cachés \mathbf{x}_n (de dimension n_x). Dès lors, chaque observation \mathbf{y}_n peut être notée

$$\mathbf{y}_n = \mathbf{F}_n^{y,x} \mathbf{x}_n + \boldsymbol{\omega}_n^y, \quad (1)$$

où $\mathbf{F}_n^{y,x}$ est la matrice d'observation et $\boldsymbol{\omega}_n^y$ un bruit. Dans le contexte probabiliste de cette thèse, les états et observations sont modélisés par des variables aléatoires (v.a). Dans le cas simple où le bruit est négligeable et où la matrice d'observation $\mathbf{F}_n^{y,x}$ est connue, le filtre inverse [52] ou le filtre régularisé s'appliquent. Sinon, le filtre de Wiener [73] permet de restaurer les états cachés quand les densités spectrales de puissance du bruit et des états cachés sont connues.

Or, ce modèle (1) ne tient pas compte des éventuelles dépendances entre états cachés successifs, pouvant améliorer significativement les résultats du filtrage. Le système, connu sous le nom de filtre de Kalman [37], ajoute à (1) une équation d'évolution markovienne des états cachés, aboutissant au système suivant

$$\begin{cases} \mathbf{x}_n = \mathbf{F}_n^{x,x} \mathbf{x}_{n-1} + \boldsymbol{\omega}_n^x \\ \mathbf{y}_n = \mathbf{F}_n^{y,x} \mathbf{x}_n + \boldsymbol{\omega}_n^y \end{cases}, \quad (2)$$

où

- $\boldsymbol{\omega}_n^x$ est le bruit de modèle, gaussien, centré, de matrice de covariance $\mathbf{Q}_n^{x,x}$.
- $\boldsymbol{\omega}_n^y$ le bruit de mesure, gaussien, centré, de matrice de covariance $\mathbf{Q}_n^{y,y}$.
- $\mathbf{F}_n^{x,x}$ la matrice d'état de dimensions $(n_x \times n_x)$.
- les v.a $\boldsymbol{\omega}_n^x$ sont supposées mutuellement indépendantes et indépendantes de la v.a \mathbf{x}_0 qui suit une loi gaussienne de moyenne $\hat{\mathbf{x}}_0$ et de matrice de covariance $\mathbf{Q}_0^{x,x}$.
- les v.a $\boldsymbol{\omega}_n^y$ sont supposées elles-mêmes mutuellement indépendantes et indépendantes des v.a $\boldsymbol{\omega}_n^x$ et de la v.a \mathbf{x}_0 .

Remarque: Dans un tel modèle, le processus conjoint $\mathbf{t} = \{\mathbf{t}_n\}_{n=0:N}$, avec $\mathbf{t}_n = [\mathbf{x}_n^T, \mathbf{y}_{n-1}^T]^T$ constitue une chaîne de Markov cachée [21], *c.a.d* la loi conjointe du couple (\mathbf{x}, \mathbf{y}) s'écrit :

$$p(\mathbf{t}) = p(\mathbf{x}_0) \prod_{n=1}^N p(\mathbf{x}_n | \mathbf{x}_{n-1}) \prod_{n=0}^N p(\mathbf{y}_n | \mathbf{x}_n). \quad (3)$$

Sous ces hypothèses, toutes les densités de probabilité (d.d.p) a posteriori des états cachés connaissant les observations sont gaussiennes. Nous noterons dans le suite du manuscrit $\mathbf{x}_{n|p}$ la v.a $\mathbf{x}_n | \{\mathbf{y}_i\}_{i=0:p}$. Depuis les travaux précurseurs de R. Kalman datant de 1960, la restauration optimale des états, au sens de l'erreur quadratique moyenne (EQM), se fait de manière exacte et rapide à partir des

équations de filtrage et de lissage et détermine à chaque instant n , la loi de $\mathbf{x}_{n|p} \sim \mathcal{N}(\hat{\mathbf{x}}_{n|p}, \mathbf{P}_{n|p})$. Le calcul de la densité de probabilité de $\mathbf{x}_{n|N}$ est effectué en deux passes séquentielles :

- **Passé avant** : Cette étape itérative propage, de proche en proche, les d.d.p a posteriori des états cachés $\mathbf{x}_{n|n}$. Elle est composée d'une étape de prédiction, suivie d'une étape de filtrage :

1. Prédiction : la d.d.p de $\mathbf{x}_{n+1|n}$ se calcule à partir de la d.d.p de $\mathbf{x}_{n|n}$ selon

$$\begin{aligned}\hat{\mathbf{x}}_{n+1|n} &= \mathbf{F}_{n+1}^{x,x} \hat{\mathbf{x}}_{n|n}, \\ \mathbf{P}_{n+1|n} &= \mathbf{Q}_{n+1}^{x,x} + \mathbf{F}_{n+1}^{x,x} \mathbf{P}_{n|n} [\mathbf{F}_{n+1}^{x,x}]^T.\end{aligned}\quad (4)$$

2. Filtrage : la d.d.p de $\mathbf{x}_{n+1|n+1}$ se calcule à partir de la d.d.p de celle de $\mathbf{x}_{n+1|n}$.

D'abord, la d.d.p de $\mathbf{y}_{n+1|n}$ est déterminée par

$$\begin{aligned}\hat{\mathbf{y}}_{n+1|n} &= \mathbf{F}_{n+1}^{y,x} \hat{\mathbf{x}}_{n+1|n}, \\ \mathbf{S}_{n+1|n} &= \mathbf{Q}_{n+1}^{y,y} + \mathbf{F}_{n+1}^{y,x} \mathbf{P}_{n+1|n} [\mathbf{F}_{n+1}^{y,x}]^T,\end{aligned}\quad (5)$$

où $\hat{\mathbf{y}}_{n+1|n}$ et $\mathbf{S}_{n+1|n}$ représentent respectivement l'espérance et la covariance de $\mathbf{y}_{n+1|n}$.

Puis, la d.d.p de $\mathbf{x}_{n+1|n+1}$ est obtenue grâce à

$$\begin{aligned}\tilde{\mathbf{y}}_{n+1|n} &= \mathbf{y}_{n+1} - \hat{\mathbf{y}}_{n+1|n}, \\ \mathbf{K}_{n+1|n+1} &= \mathbf{Q}_{n+1|n}^{x,x} [\mathbf{F}_{n+1}^{y,x}]^T \mathbf{S}_{n+1|n}^{-1}, \\ \hat{\mathbf{x}}_{n+1|n+1} &= \hat{\mathbf{x}}_{n+1|n} + \mathbf{K}_{n+1|n+1} \tilde{\mathbf{y}}_{n+1|n}, \\ \mathbf{P}_{n+1|n+1} &= \mathbf{P}_{n+1|n} - \mathbf{K}_{n+1|n+1} \mathbf{Q}_{n+1|n}^{y,y} \mathbf{K}_{n+1|n+1}^T,\end{aligned}\quad (6)$$

où $\tilde{\mathbf{y}}_{n+1|n}$ est l'innovation et $\mathbf{K}_{n+1|n+1}$ le gain de filtrage.

- **Passé arrière** : Cette étape, aussi dénommée lissage, propage de proche en proche de manière rétrograde la d.d.p a posteriori de $\mathbf{x}_{n|N}$. La d.d.p de $\mathbf{x}_{n|N}$ se calcule en fonction de la d.d.p de $\mathbf{x}_{n+1|N}$ selon

$$\begin{aligned}\mathbf{K}_{n|N} &= \mathbf{P}_{n|n} [\mathbf{F}_n^{x,x}]^T \mathbf{P}_{n+1|N}^{-1}, \\ \hat{\mathbf{x}}_{n|N} &= \hat{\mathbf{x}}_{n+1|N} - \mathbf{K}_{n|N} [\hat{\mathbf{x}}_{n+1|N} - \hat{\mathbf{x}}_{n+1|n}], \\ \mathbf{P}_{n|N} &= \mathbf{P}_{n|n} + \mathbf{K}_{n|N} [\mathbf{P}_{n+1|N} - \mathbf{P}_{n+1|n}] \mathbf{K}_{n|N}^T,\end{aligned}\quad (7)$$

où $\mathbf{K}_{n|N}$ est le gain de lissage.

Remarque: La restauration des états cachés par $\hat{\mathbf{x}} = \{\hat{\mathbf{x}}_{n|N}\}_{n \in \mathbb{N}}$ est optimale au sens de l'EQM,

$$EQM(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{P} \sum_{i=0}^{P-1} [\mathbf{x}_i - \hat{\mathbf{x}}_i]^T [\mathbf{x}_i - \hat{\mathbf{x}}_i] \quad (8)$$

et maximise la fonction de vraisemblance $p(\mathbf{y} | \mathbf{x})$ du modèle.

Bien que le filtre de Kalman permette le calcul exact des d.d.p a posteriori des états cachés, il est bien connu que la propagation des matrices de covariance peut provoquer une divergence numérique du filtre [24,55,63] : les matrices de covariance, théoriquement positives et symétriques, peuvent perdre l'une ou/et l'autre des propriétés au cours du déroulement des algorithmes séquentiels de filtrage ou de lissage. Ces erreurs proviennent de la précision limitée des ordinateurs. La divergence numérique apparaît principalement dans les cas où

1. les bruits de modèle et de mesure ne sont pas du même ordre de grandeur,
2. le signal à restaurer comporte un grand nombre N d'échantillons.

Deux techniques permettent de retarder la divergence numérique du filtre : l'augmentation du nombre de digits employés pour les calculs numériques (de 32 à 64 bits par exemple) ou la symétrisation forcée des matrices de covariance. Cependant, aucune de ces deux méthodes ne garantit formellement la positivité des matrices de covariance, et n'empêche donc la divergence numérique. Afin de remédier à ce problème, deux familles de méthodes sont utilisées :

1. La première famille ajoute un bruit fictif aux observations ou aux états cachés. Cela permet de faire face au mauvais conditionnement des matrices de covariance du bruit de mesure ou de modèle, et de garantir la positivité des matrices de covariance. Celle-ci suffit à garantir la stabilité du filtre, mais l'ajout de bruit biaise l'estimation des états cachés.
2. La seconde famille propose de remplacer la propagation des matrices de covariance dans l'algorithme séquentiel, par la propagation d'une de leur racine-carrée [7], ou d'une matrice dérivée. Ainsi, si on garde les matrices de covariance sous cette forme, on impose d'une part, la symétrie et d'autre part, la positivité. Dès lors, la divergence numérique due à la perte d'une de ces deux propriétés devient impossible. Deux types d'algorithmes rapides ont été proposés pour propager ces matrices :
 - a. Les algorithmes basés sur des décompositions UD [6]¹ propagent une matrice unitaire ainsi que les valeurs propres de la matrice de covariance.
 - b. Les algorithmes basés sur des décompositions QR [47]² propagent une racine carrée triangulaire supérieure de la matrice de covariance.

Le filtre de Kalman, tel qu'il a été défini dans les équations (4 – 7) n'est optimal que dans le cadre de modèles linéaires et gaussiens, ce qui n'est pas le cas de la plupart des modèles physiques. Dans un cas non-linéaire, l'équation d'état du filtre de Kalman est remplacée par une équation fonctionnelle de la forme

$$\mathbf{x}_{n+1} = \mathbf{G}^{\mathbf{x},\mathbf{x}}(\mathbf{x}_n) + \boldsymbol{\omega}_{n+1}^{\mathbf{x}}, \quad (9)$$

1. Toute matrice symétrique positive \mathbf{S} s'écrit $\mathbf{S} = \mathbf{U}\mathbf{D}\mathbf{U}^T$, avec \mathbf{U} , une matrice unitaire et \mathbf{D} , une matrice de valeurs propres positives.

2. Toute matrice \mathbf{M} peut s'écrire $\mathbf{M} = \mathbf{U}\mathbf{R}$, où \mathbf{U} est une matrice unitaire et \mathbf{R} , une matrice triangulaire supérieure.

où $\mathbf{G}^{x,x}$ est une fonction matricielle. Dès lors, le filtre de Kalman ne peut plus s'appliquer. Une première approche, le filtre de Kalman non-linéaire ou filtre de Kalman étendu (EKF) [36], propose de réaliser une linéarisation de l'équation (9) sous la forme

$$\mathbf{x}_{n+1} = \underbrace{\frac{\partial \mathbf{G}^{x,x}}{\partial \mathbf{x}}(\mathbf{x}_n)}_{\mathbf{F}_{n+1}^{x,x}} \mathbf{x}_n + \boldsymbol{\omega}_{n+1}^x. \quad (10)$$

Les équations du filtre de Kalman sont alors ré-utilisées en tenant compte de cette approximation. Cependant, l'EKF n'est pas optimal selon les critères de la vraisemblance et de l'erreur quadratique moyenne [8,44]. Une autre conséquence est la tendance du filtre à sous-estimer les matrices de covariance, et donc à sur-estimer la confiance dans les estimations produites. La qualité de filtrage peut être améliorée de deux façons :

1. Il a été proposé d'ajouter un bruit fictif aux mesures ou aux états cachés pour corriger les covariances des états cachés estimés. Bien que cette solution puisse dans certains cas résoudre le problème, elle biaise l'estimation et ne garantit pas la convergence.
2. Une autre approche [19] consiste à intégrer l'erreur d'approximation lors de la linéarisation de (9) dans les matrices de covariance estimées. Cette méthode est plus fiable que la précédente mais ne garantit pas non-plus théoriquement la convergence du filtre de Kalman étendu vers la solution optimale.

À l'instar du filtre de Kalman, le filtre de Kalman étendu souffre lui aussi de problème de stabilité numérique. Les algorithmes robustes proposés pour le filtre de Kalman sont eux-aussi applicables au filtre de Kalman étendu.

Au lieu d'utiliser des méthodes de linéarisation de la matrice d'état, Julier *et al* ont proposé de propager un ensemble de points pour approximer les densités de probabilité des états cachés. Cette méthode est employée dans le filtre de Kalman sans-saveur (UKF) [71] qui utilise un ensemble déterministe de $2n_x + 1$ points sigma sélectionnés autour de l'espérance de l'état caché $\hat{\mathbf{x}}_{n|n}^{UKF}$ selon

$$\begin{aligned} \mathbf{x}_{n,0}^\sigma &= \hat{\mathbf{x}}_{n|n}^{UKF}, \\ \forall i \in [1, n_x], \mathbf{x}_{n,i}^\sigma &= \hat{\mathbf{x}}_{n|n}^{UKF} + \left[\frac{w_i}{2}\right]^{-1} [\mathbf{P}_{n|n}^{UKF}]_i^{\frac{1}{2}}, \\ \forall i \in [1, n_x], \mathbf{x}_{n,i+n_x}^\sigma &= \hat{\mathbf{x}}_{n|n}^{UKF} - \left[\frac{w_i}{2}\right]^{-1} [\mathbf{P}_{n|n}^{UKF}]_i^{\frac{1}{2}}, \\ \forall i \in [1, 2n_x], w^i &= \frac{1 - w^0}{2n_x}, \end{aligned} \quad (11)$$

où $[\mathbf{P}_{n|n}^{UKF}]_i^{\frac{1}{2}}$ est la i -ième colonne d'une racine-carrée de la matrice de covariance estimée $\mathbf{P}_{n|n}^{UKF}$, et où w^0 est un poids pris entre 0 et 1. Les points sont ensuite propagés via l'équation d'état (9) :

$$\hat{\mathbf{x}}_{n+1,i}^{\sigma,p} = \mathbf{G}^{x,x}(\mathbf{x}_{n,i}^\sigma). \quad (12)$$

L'espérance et la covariance de l'état prédit $\mathbf{x}_{n+1|n}$ sont alors déterminées de manière empirique grâce aux formules suivantes

$$\begin{aligned}\hat{\mathbf{x}}_{n+1|n}^{UKF} &= \sum_{i=0}^{2n_x} w^i \mathbf{x}_{n+1,i}^{\sigma,p}, \\ \mathbf{P}_{n+1|n}^{UKF} &= \sum_{i=0}^{2n_x} w^i [\mathbf{x}_{n+1,i}^{\sigma,p} - \hat{\mathbf{x}}_{n+1|n}^{UKF}] [\mathbf{x}_{n+1,i}^{\sigma,p} - \hat{\mathbf{x}}_{n+1|n}^{UKF}]^T + \mathbf{Q}^{x,x}.\end{aligned}\quad (13)$$

L'état filtré $\mathbf{x}_{n+1|n+1}$ est ensuite estimé grâce aux équations de filtrage du filtre de Kalman étendu, remplaçant $\hat{\mathbf{x}}_{n+1|n}$ et $\mathbf{P}_{n+1|n}$ par $\hat{\mathbf{x}}_{n+1|n}^{UKF}$ et $\mathbf{P}_{n+1|n}^{UKF}$. Pour le lissage, les auteurs ré-utilisent également les équations présentées dans le filtre de Kalman étendu. À l'instar du filtre de Kalman, l'UKF souffre lui aussi de problème de stabilité numérique. Dans [62, 70], les auteurs proposent des méthodes qui propagent les racines carrées des matrices de covariance. Cependant, il faut souligner que l'UKF reste non optimal selon le critère de la vraisemblance et selon le critère de l'erreur quadratique moyenne.

Les méthodes dites de Monte-Carlo [46] constituent une alternative à ces filtres dans le cas non-linéaire et/ou non-gaussien. Ces méthodes reposent sur l'estimation empirique des d.d.p des variables aléatoires par un grand nombre de tirages aléatoires. Elles sont connues pour être asymptotiquement efficaces selon le critère de la vraisemblance, *c.a.d* que quand le nombre de tirages tend vers l'infini, la méthode tend à maximiser la fonction de vraisemblance $p(\mathbf{y} | \mathbf{x})$. Elles ont été adaptées au filtre de Kalman sous le nom de filtre particulière [27]. Les états cachés sont tirés de proche en proche selon la d.d.p de transition $\Pi(\mathbf{x}_n | \mathbf{x}_{n-1}^j)$, où j représente le numéro de la particule. À chaque instant n , toutes les particules sont pondérées par un poids proportionnel à la vraisemblance du chemin suivi par la particule soit

$$\omega_{n|p}^j \propto p(\mathbf{y}_{0:p} | \mathbf{x}_{0:n}^j). \quad (14)$$

Les d.d.p des états cachés sont ensuite déterminées de manière empirique. Cependant, il a été constaté par de nombreux auteurs que, après un grand nombre d'itérations, seulement quelques particules pouvaient être représentatives (*i.e.* les poids d'un grand nombre de particules devenait négligeable). Ce problème est celui de la dégénérescence des méthodes de Monte-Carlo. L'article [18] propose d'introduire un ré-échantillonnage des particules, nommé SIR (*Sequential Importance Resampling*) : à chaque itération n , la dispersion des particules est mesurée et si elle est trop importante, les particules sont ré-échantillonnées, *c.a.d* que les particules de poids les plus faibles sont alors supprimées et les particules représentatives sont dupliquées proportionnellement à leur poids. Il est à noter que les méthodes de ré-échantillonnage combinent les particules entre elles et empêchent donc toute parallélisation³ pour le tirage des particules.

3. La parallélisation des particules dans les méthodes particulières sans ré-échantillonnage est réalisée par le tirage simultané de plusieurs particules.

Notons finalement, que les méthodes de filtre de Kalman sans-saveur, ou de filtre de Kalman à particules augmentent la complexité algorithmique, comparé à l'algorithme original, d'un facteur proportionnel au nombre de particules ou de points sigma. Aucun de ces modèles n'est optimal selon le critère de la vraisemblance, même si le filtre particulière l'est asymptotiquement.

2 Filtre de Kalman non-supervisé

Tous les modèles présentés ci-dessus requièrent la connaissance parfaite de leurs paramètres, dont l'ensemble sera noté θ , pour restaurer ou prédire les états cachés. Ces paramètres sont soit donnés par la connaissance physique du problème, soit réglés manuellement, ce qui s'avère souvent délicat. Il est donc nécessaire d'apprendre les paramètres du modèle, *c.a.d* maximiser un critère en fonction de ces paramètres. Dans la littérature, très souvent, on cherche à maximiser la fonction de vraisemblance, dont on nomme l'argument du maximum, estimateur du maximum de vraisemblance. Cet estimateur permet d'estimer les paramètres dans le cadre de modèles simples, mais n'a plus nécessairement d'expression analytique quand le modèle devient plus complexe.

Soit le logarithme népérien de la fonction de vraisemblance d'un filtre de Kalman

$$\begin{aligned} \log \mathcal{L}(\mathbf{y}; \theta) &= \log p(\mathbf{y} | \theta), \\ &= \sum_{n=0}^N \log |\mathbf{S}_{n|n-1}| + \tilde{\mathbf{y}}_{n|n-1}^T \mathbf{S}_{n|n-1}^{-1} \tilde{\mathbf{y}}_{n|n-1}. \end{aligned} \quad (15)$$

Il n'est pas possible de maximiser directement la vraisemblance. L'algorithme d'espérance-maximisation (EM) [14] offre une solution itérative pour approcher le maximum de vraisemblance. Chaque itération j de l'algorithme EM se décompose en deux étapes :

1. l'étape d'espérance (E) :

Les d.d.p de $\mathbf{x}_{n|N}^{(j)} = \mathbf{x}_n | \mathbf{y}_{0:N}, \theta^{(j)}$ sont estimées via les algorithmes de filtrage et de lissage pour déterminer la fonction auxiliaire de vraisemblance

$$g^{(j)}(\theta) = E \left(\log(\mathcal{L}(\mathbf{x}, \mathbf{y}; \theta)) \middle| \mathbf{y}, \theta^{(j)} \right) \quad (16)$$

qui a le même sens de variation que la fonction de vraisemblance.

2. l'étape de maximisation (M) :

Les paramètres du modèle sont mis à jour en maximisant la fonction auxiliaire de vraisemblance par rapport à ceux-ci :

$$\theta^{(j+1)} = \underset{\theta}{\operatorname{argmax}} [g^{(j)}(\theta)]. \quad (17)$$

Quand il n'est pas possible de trouver une expression analytique pour l'argument du maximum de la fonction auxiliaire de vraisemblance, on peut se contenter d'en accroître la valeur. Un tel algorithme converge lui-aussi vers un maximum de la fonction de vraisemblance et se dénomme algorithme EM généralisé (GEM).

Remarques:

- L'algorithme EM converge vers un maximum local de la fonction de vraisemblance et non global. Donc, si l'algorithme est initialisé trop loin de la solution optimale, il peut rester piégé par l'un de ces maxima locaux.
- Une variante de l'algorithme EM permet de réduire les risques de convergence vers les maxima locaux : l'algorithme EM stochastique (SEM) [10]. Celui-ci ajoute une étape stochastique (S) entre les étapes d'espérance et de maximisation. L'étape S repose sur le tirage des états cachés \mathbf{x} selon la loi a posteriori $p(\mathbf{x}_n | \mathbf{y}_{0:N})$. Dans sa version initiale, l'algorithme EM est arrêté quand la variation des états cachés est négligeable entre deux itérations. Cependant, la garantie de l'arrêt de l'algorithme SEM est impossible à cause de sa nature stochastique.
- Le principe de l'algorithme EM a été appliqué dans le cadre du filtre de Kalman [66] pour l'apprentissage non-supervisé de ses paramètres.

Les algorithmes d'apprentissage non-supervisés sont aussi confrontés au problème de l'identifiabilité des modèles que l'on rencontre quand plusieurs jeux de paramètres ont la même fonction de vraisemblance. On peut illustrer simplement ce problème d'identifiabilité dans les modèles de mélange finis, où il est possible d'invertir deux classes sans modifier la fonction de vraisemblance. Dans le contexte de la classification de données où il est utilisé, l'ordre des classes n'a généralement pas d'importance. Par contre, le problème d'identifiabilité devient nettement plus rédhibitoire dans le cadre du modèle de Kalman où, comme nous le verrons en détails dans les prochains chapitres, toute homothétie des états cachés donne exactement la même fonction de vraisemblance. Dès lors, ce problème peut être dramatique pour l'algorithme EM car il existe ainsi plusieurs jeux de paramètres optimaux aboutissant à des restaurations différentes ! Réduire cette ambiguïté lors de l'estimation EM est l'un des principaux objectifs de cette thèse que nous détaillons maintenant.

3 Contributions de cette thèse

Cette thèse propose plusieurs contributions concernant une extension du filtre de Kalman, appelée filtre de Kalman couple [17, 43]. Celui-ci intègre des dépendances supplémentaires entre les observations et les états cachés et offre donc une plus grande richesse de modélisation que le filtre de Kalman, et par conséquent, devrait permettre de traiter des signaux plus complexes.

Notre étude soulève cinq problématiques distinctes :

1. Tout d'abord, les problèmes de stabilité numérique du filtre de Kalman se reportent-ils sur le filtre de Kalman couple ?
2. Le modèle du filtre de Kalman couple est-il identifiable ? Et si non, cela a-t-il une influence sur la convergence de l'algorithme EM ?
3. L'ajout d'informations a priori sur les paramètres provenant de connaissances partielles sur la physique du problème ou le capteur améliore-t-il leur estimation ?
4. Le filtre de Kalman couple est-il performant pour filtrer des données sortant de son cadre théorique ?
5. En particulier, dans le cadre d'une application de biométrie dynamique à partir de l'iris, permet-il d'améliorer le suivi du mouvement de l'oeil, et donc d'améliorer sa reconnaissance ?

Notre travail se décompose donc en cinq chapitres visant à apporter des éléments de réponse aux différents points évoqués ci-dessus.

Chapitre 1

Ce chapitre est dédié à la présentation du filtre de Kalman couple, de ses algorithmes de filtrage et lissage, ainsi que de son algorithme EM pour l'estimation de l'ensemble des paramètres. Nous présentons à la fois son implémentation originale, détaillée dans [4] et notre variante robuste, publiée dans [48]. Nous proposons une méthode qui adapte les algorithmes de type QR du filtre de Kalman au filtre de Kalman couple. Ceux-ci ne calculent plus les matrices de covariance ou de corrélation directement mais une de leur racine-carrée triangulaire supérieure.

Le chapitre 1 est donc organisé de la manière suivante :

- La section 1 est dédiée à la présentation de l'algorithme EM original du filtre de Kalman couple.
- La section 2 est dédiée à la présentation de sa version robuste.
- Dans la section 3, l'algorithme EM robuste est mis en valeur par une expérience mal-conditionnée, qui met en valeur sa plus grande stabilité numérique.

Chapitre 2

Ce chapitre présente le problème d'identifiabilité du filtre de Kalman couple, du à l'existence de jeux de paramètres possédant exactement la même fonction de vraisemblance. Il introduit l'utilisation de jeux de paramètres équivalents pour contraindre la structure des paramètres du modèle du filtre couple et restreindre l'ambiguïté lors de l'estimation par l'algorithme EM.

Nous établissons 3 types de contraintes qui, à partir de l'estimation du filtre original, permettent d'estimer un jeu de paramètres avec exactement la même fonction de vraisemblance et une structure particulière. De tels filtres seront ap-

pelés, dans la suite du manuscrit, filtres équivalents. Dans le but d'illustrer l'influence du problème d'identifiabilité sur l'algorithme EM et de mettre en valeur l'intérêt de ces contraintes, nous avons mené plusieurs séries d'expériences qui comparent systématiquement (i) le filtre supervisé avec les paramètres optimaux, (ii) le filtre de Kalman couple original ou libre et (iii) un filtre de Kalman couple équivalent.

Le chapitre 2 est donc organisé de la manière suivante :

- La section 1 présente la principale propriété des jeux de paramètres équivalents : si deux algorithmes EM sont initialisés, l'un sur le jeu de paramètres θ et l'autre sur le jeu de paramètres équivalents θ^M , alors le chemin parcouru sur la fonction de vraisemblance par les deux algorithmes EM sont identiques.
- La section 2 propose de sélectionner des jeux de paramètres équivalents qui imposent une structure particulière aux paramètres. Elle développe les contraintes imposant une restauration unique des données dégradées par un bruit additif.
- La dernière section montre le résultat de plusieurs séries d'expériences qui mettent en valeur l'utilité du filtre de Kalman couple équivalent vis-à-vis du filtre de Kalman couple libre.

Chapitre 3

Ce chapitre propose une approche différente de celle du chapitre 2 pour contraindre le jeu de paramètres du modèle du filtre de Kalman couple lors de l'apprentissage par un algorithme EM. Il est proposé de contraindre directement la forme du jeu de paramètres lors de l'étape de maximisation de l'algorithme EM. Il faut noter que l'algorithme EM requiert des formules de ré-estimation exactes des paramètres, ce qui ne peut être garanti pour tout type de contraintes. Aussi, nous avons établi des contraintes peu restrictives qui préservent l'existence d'un algorithme EM. Nous proposons également une variante robuste de cet algorithme, compatible avec l'étape d'espérance robuste présentée dans le chapitre 1.

Le chapitre 3 est donc organisé de la manière suivante :

- La section 1 présente les différentes contraintes que nous avons réussies à imposer lors de l'étape de maximisation de l'algorithme EM.
- La section suivante propose une variante robuste de l'étape M contrainte, qui est compatible avec l'étape d'espérance robuste présentée dans la section 2 du chapitre 1.
- La section 3 montre le résultat de plusieurs séries d'expériences qui comparent les filtres de Kalman couple libre, équivalent, contraint et optimal sur des données suivant le modèle du filtre de Kalman couple.

Chapitre 4

Ce chapitre propose une analyse expérimentale qui compare les différents algorithmes introduits dans les 3 premiers chapitres et le filtre de Kalman non-supervisé. Dans ces expériences, sélectionnées pour leur diversité, nous testons les capacités du filtre de Kalman couple (i) à restaurer des données adaptées au filtre de Kalman, (ii) à restaurer des données simulées dégradées par un bruit additif, (iii) à estimer des états dans une problématique d'asservissement et, (iv) finalement, à poursuivre le mouvement d'une pupille humaine dans une vidéo avec l'aide d'un algorithme de segmentation.

Le chapitre 4 est donc organisé de la manière suivante :

- La première section montre les résultats d'une série d'expériences qui compare les performances des filtres de Kalman couples non-supervisés aux filtres de Kalman non-supervisés sur des données théoriques.
- La section 2 présente plusieurs séries d'expériences qui comparent les différents filtres de Kalman et Kalman couple sur des données ne suivant pas le modèle du filtre de Kalman couple.
- La dernière section compare différents filtres de Kalman non-supervisés contraints pour suivre le mouvement de la pupille.

Chapitre 5

Le dernier chapitre propose d'intégrer les résultats sur le suivi de la pupille obtenus dans le chapitre précédent dans une application complète de reconnaissance d'iris. En plus du modèle de mouvement estimé, nous proposons une méthode de segmentation robuste de la pupille, une sélection intelligente des images à traiter, ainsi qu'une décomposition de la chaîne de traitements en plusieurs tâches parallèles. Ces améliorations ont permis d'une part de traiter les vidéos en temps-réel et, d'autre part, d'obtenir des résultats d'identification légèrement supérieurs à ceux de l'état de l'art sur la base MBGC [53].

Le chapitre 5 est organisé de la manière suivante :

- La première section 1 décrit les principes de la biométrie dynamique de l'iris.
- La section 2 présente le processus qui transforme une image d'oeil en un iris-code (*i.e.* code binaire caractéristique de l'individu).
- La troisième section décrit la méthode de suivi de la pupille, la sélection intelligente des meilleures images dans la vidéo, ainsi que la parallélisation permettant d'obtenir un traitement temps-réel.
- La dernière section montre les résultats de segmentation sur la base MBGC et conclut sur les stratégies à adopter pour la haute-sécurité ou les applications tout-public de reconnaissance d'iris.

Publications de l'auteur durant la thèse

Publication dans des revues

V. Nêmesin and S. Derrode. Robust blind pairwise Kalman algorithms using QR decompositions. *IEEE Transactions on Signal Processing*, 61(1) :5–9, Jan 2013.

Conférences internationales avec comité de lecture

V. Nêmesin et al. The first ICB competition on iris recognition. In *IEEE/IAPR International Joint Conference on Biometrics*, 2014.

V. Nêmesin, S. Derrode, and A. Benazza-Benyahia. Gradual iris code construction from close-up eye video. In J. Blanc-Talon, W. Philips, D. C. Popescu, P. Scheunders, and P. Zemečik, editors, *14th Int. Conf. on Advanced Concepts for Intelligent Vision Systems (ACIVS'12)*, volume 7517 of *Lecture Notes in Computer Science*, pages 12–23. Springer, 2012.

Chapitre 1

Filtre de Kalman couple robuste

Ce chapitre décrit un algorithme d'apprentissage robuste des paramètres pour le système linéaire gaussien homogène [17, 43] suivant :

$$\underbrace{\begin{bmatrix} \mathbf{x}_{n+1} \\ \mathbf{y}_n \end{bmatrix}}_{\mathbf{t}_{n+1}} = \underbrace{\begin{bmatrix} \mathbf{F}^{x,x} & \mathbf{F}^{x,y} \\ \mathbf{F}^{y,x} & \mathbf{F}^{y,y} \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} \mathbf{x}_n \\ \mathbf{y}_{n-1} \end{bmatrix}}_{\mathbf{t}_n} + \underbrace{\begin{bmatrix} \boldsymbol{\omega}_{n+1}^x \\ \boldsymbol{\omega}_{n+1}^y \end{bmatrix}}_{\boldsymbol{\omega}_{n+1}}, \quad (1.1)$$

dans lequel $\boldsymbol{\omega}_n$ est un bruit gaussien centré de matrice de covariance

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}^{x,x} & \mathbf{Q}^{x,y} \\ [\mathbf{Q}^{x,y}]^T & \mathbf{Q}^{y,y} \end{bmatrix},$$

et où les $\boldsymbol{\omega}_n$ sont mutuellement indépendants et indépendant de \mathbf{t}_0 qui suit la loi gaussienne $\mathcal{N}(\hat{\mathbf{t}}_0, \mathbf{Q}_0)$.

Le processus $\mathbf{t} = \{\mathbf{t}_n\}_{n \in \mathbb{N}}$ est un processus de Markov couple [4], par analogie aux processus couples à temps discret introduits dans [54] et évalués dans [16] en segmentation d'images. Dans ce modèle, le processus \mathbf{x} n'est plus markovien ; cependant, il reste possible de construire des algorithmes de filtrage et de lissage exacts [2, proposition 3] sans avoir recours aux méthodes particulières [35].

Remarques:

- Le filtre de Kalman est un filtre de Kalman couple simplifié où $\mathbf{F}^{x,y} = \mathbf{O}$ et $\mathbf{F}^{y,y} = \mathbf{O}$.
- Le filtre de Kalman couple peut s'écrire sous la forme d'un filtre de Kalman dégénéré¹ pour l'état augmenté $\mathbf{z}_n = [\mathbf{t}_n^T, \mathbf{y}_n^T]^T$, pour la matrice de transition

$$\mathbf{F}^{z,z} = \begin{bmatrix} \mathbf{F}^{x,x} & \mathbf{F}^{x,y} & \mathbf{O}_{n_x, n_y} \\ \mathbf{F}^{y,x} & \mathbf{F}^{y,y} & \mathbf{O}_{n_y, n_y} \\ \mathbf{F}^{y,x} & \mathbf{F}^{y,y} & \mathbf{O}_{n_y, n_y} \end{bmatrix},$$

1. Le filtre est appelé dégénéré puisque la matrice de covariance du bruit $\mathbf{Q}^{z,z}$ est singulière.

et pour la matrice de covariance

$$Q^{z,z} = \begin{bmatrix} Q^{x,x} & Q^{x,y} & Q^{x,y} \\ Q^{y,x} & Q^{y,y} & Q^{y,y} \\ Q^{y,x} & Q^{y,y} & Q^{y,y} \end{bmatrix}.$$

Cependant, le filtre de Kalman couple est décrit par un nombre plus grand de paramètres que le filtre de Kalman, il est donc plus difficile à régler. Pour répondre à ce problème, un algorithme EM a été proposé pour apprendre de manière non-supervisée les paramètres [3]. Cependant, les équations proposées souffrent d'instabilités numériques, dues à la perte de la symétrie ou de la positivité des matrices de covariance, ce qui peut entraîner la divergence de la vraisemblance.

La principale contribution de ce chapitre concerne donc le développement d'un algorithme robuste qui adapte la méthode de factorisation [38] au cas du filtre de Kalman couple, *c.a.d* qui propage des racines carrées des matrices de covariance à la place de celles-ci. Les travaux présentés dans ce chapitre sont publiés dans [48].

Ce chapitre est divisé en trois sections. La première présente les étapes de prédiction, filtrage et de lissage ainsi que l'algorithme EM pour le filtre de Kalman couple. La deuxième section décrit une implémentation numériquement robuste de cet algorithme EM. La dernière section présente une illustration du bénéfice obtenu par l'algorithme robuste par rapport à l'algorithme original.

1 Algorithme EM

La figure 1.1 schématise le principe de fonctionnement de l'algorithme EM pour le FKC dans lequel se succède l'étape d'espérance et l'étape de maximisation jusqu'à la convergence. Nous décrivons dans la suite le détail de chaque étape. Nous omettons volontairement l'indice j de l'itération EM pour simplifier les écritures.

1.1 Étape E

L'étape E du filtre de Kalman couple se décompose en trois phases successives, la passe-avant, la passe-arrière et le calcul de la fonction auxiliaire de vraisemblance. Pour les calculs, nous avons recours aux constantes suivantes :

$$\begin{aligned} \mathbf{F}^{y,t} &= [\mathbf{F}^{y,x}, \mathbf{F}^{y,y}], \\ \mathbf{F}_2^{x,x} &= \mathbf{F}^{x,x} - \mathbf{Q}^{x,y} [\mathbf{Q}^{y,y}]^{-1} \mathbf{F}^{y,x}, \\ \mathbf{F}_2^{x,y} &= \mathbf{F}^{x,y} - \mathbf{Q}^{x,y} [\mathbf{Q}^{y,y}]^{-1} \mathbf{F}^{y,y}, \\ \mathbf{F}_2^{x,t} &= [\mathbf{F}_2^{x,x}, \mathbf{F}_2^{x,y}], \\ \mathbf{Q}_2^{x,x} &= \mathbf{Q}^{x,x} - \mathbf{Q}^{x,y} [\mathbf{Q}^{y,y}]^{-1} [\mathbf{Q}^{x,y}]^T. \end{aligned} \tag{1.2}$$

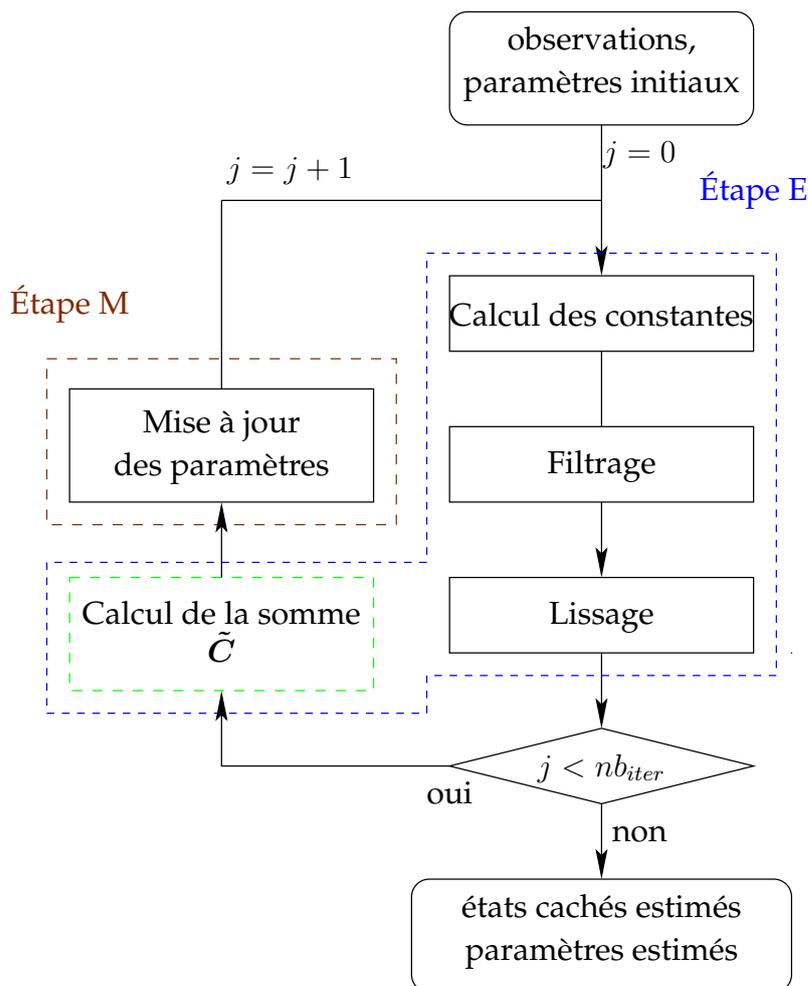


FIGURE 1.1: Algorithme EM pour le filtre de Kalman couple.

Passé avant :

La passe-avant est analogue à celle du filtre de Kalman. Il s'agit d'une étape itérative qui propage de proche en proche les d.d.p a posteriori des états cachés $\mathbf{x}_{n|n}$. Chaque itération est composée d'une phase de prédiction et d'une phase de filtrage :

1. Prédiction :

La d.d.p de $\mathbf{x}_{n+1|n}$ se calcule à partir de la d.d.p de $\mathbf{x}_{n|n}$:

$$\begin{aligned}\hat{\mathbf{x}}_{n+1|n} &= \mathbf{F}_2^{x,x} \hat{\mathbf{x}}_{n|n} + \mathbf{Q}^{x,y} [\mathbf{Q}^{y,y}]^{-1} \mathbf{y}_n + \mathbf{F}_2^{x,y} \mathbf{y}_{n-1}, \\ \mathbf{P}_{n+1|n} &= \mathbf{Q}_2^{x,x} + \mathbf{F}_2^{x,x} \mathbf{P}_{n|n} [\mathbf{F}_2^{x,x}]^T.\end{aligned}\quad (1.3)$$

Remarque: La d.d.p de $\mathbf{x}_{1|0}$ est obtenue par

$$\begin{aligned}\hat{\mathbf{x}}_{1|0} &= \mathbf{F}_2^{x,t} \hat{\mathbf{t}}_{0|0} + \mathbf{Q}^{x,y} [\mathbf{Q}^{y,y}]^{-1} \mathbf{y}_0, \\ \mathbf{P}_{1|0} &= \mathbf{Q}_2^{x,x} + \mathbf{F}_2^{x,t} \mathbf{Q}_{0|0} [\mathbf{F}_2^{x,t}]^T,\end{aligned}\quad (1.4)$$

où $\hat{\mathbf{t}}_{0|0}$ et $\mathbf{Q}_{0|0}$ sont respectivement l'espérance et la matrice de covariance de la v.a. $\mathbf{t}_{0|0} = \mathbf{t}_0 | \mathbf{y}_0$.

2. Filtrage :

La d.d.p de $\mathbf{x}_{n+1|n+1}$ se calcule à partir de la d.d.p de $\mathbf{x}_{n+1|n}$.

– La d.d.p de $\mathbf{y}_{n+1|n}$ est requise et se calcule selon

$$\begin{aligned}\hat{\mathbf{y}}_{n+1|n} &= \mathbf{F}^{y,x} \hat{\mathbf{x}}_{n+1|n} - \mathbf{F}^{y,y} \mathbf{y}_n, \\ \mathbf{S}_{n+1|n} &= \mathbf{Q}^{y,y} + \mathbf{F}^{y,x} \mathbf{P}_{n+1|n} [\mathbf{F}^{y,x}]^T,\end{aligned}\quad (1.5)$$

où $\hat{\mathbf{y}}_{n+1|n}$, $\mathbf{S}_{n+1|n}$ désignent respectivement l'espérance et la covariance de $\mathbf{y}_{n+1|n}$.

Remarque: La d.d.p de \mathbf{y}_0 est obtenue par

$$\begin{aligned}\hat{\mathbf{y}}_0 &= \mathbf{F}^{y,t} \hat{\mathbf{t}}_0, \\ \mathbf{S}_0 &= \mathbf{Q}^{y,y} + \mathbf{F}^{y,t} \mathbf{Q}_0 [\mathbf{F}^{y,t}]^T.\end{aligned}\quad (1.6)$$

– la d.d.p de $\mathbf{x}_{n+1|n+1}$ se calcule à partir des d.d.p de $\mathbf{y}_{n+1|n}$ et $\mathbf{x}_{n+1|n}$

$$\begin{aligned}\tilde{\mathbf{y}}_{n+1|n} &= \mathbf{y}_{n+1} - \hat{\mathbf{y}}_{n+1|n}, \\ \mathbf{K}_{n+1|n+1} &= \mathbf{P}_{n+1|n} [\mathbf{F}^{y,x}]^T \mathbf{S}_{n+1|n}^{-1}, \\ \hat{\mathbf{x}}_{n+1|n+1} &= \hat{\mathbf{x}}_{n+1|n} + \mathbf{K}_{n+1|n+1} \tilde{\mathbf{y}}_{n+1|n}, \\ \mathbf{P}_{n+1|n+1} &= \mathbf{P}_{n+1|n} - \mathbf{K}_{n+1|n+1} \mathbf{S}_{n+1|n} \mathbf{K}_{n+1|n+1}^T,\end{aligned}\quad (1.7)$$

où $\tilde{\mathbf{y}}_{n+1|n}$ est l'innovation et $\mathbf{K}_{n+1|n+1}$ le gain de filtrage.

Remarque: Pour $n = 0$, la d.d.p de $\mathbf{t}_{0|0}$ se calcule selon

$$\begin{aligned}\mathbf{K}_{0|0} &= \mathbf{Q}_0 [\mathbf{F}^{y,t}]^T \mathbf{S}_0^{-1}, \\ \hat{\mathbf{t}}_{0|0} &= \hat{\mathbf{t}}_0 + \mathbf{K}_{0|0} \tilde{\mathbf{y}}_0, \\ \mathbf{Q}_{0|0} &= \mathbf{Q}_0 - \mathbf{K}_{0|0} \mathbf{S}_0 \mathbf{K}_{0|0}^T.\end{aligned}\quad (1.8)$$

Passé arrière :

L'étape de passé arrière est quasiment identique à celle décrite pour le filtre de Kalman (7). Les d.d.p de $\mathbf{x}_{n|N}$ se calculent de proche à proche selon

$$\begin{aligned}\mathbf{K}_{n|N} &= \mathbf{P}_{n|n} [\mathbf{F}_2^{x,x}]^T \mathbf{P}_{n+1|n}^{-1}, \\ \hat{\mathbf{x}}_{n|N} &= \hat{\mathbf{x}}_{n|n} + \mathbf{K}_{n|N} [\hat{\mathbf{x}}_{n+1|N} - \hat{\mathbf{x}}_{n+1|n}], \\ \mathbf{P}_{n|N} &= \mathbf{P}_{n|n} + \mathbf{K}_{n|N} [\mathbf{P}_{n+1|N} - \mathbf{P}_{n+1|n}] \mathbf{K}_{n|N}^T,\end{aligned}\quad (1.9)$$

où $\mathbf{K}_{n|N}$ est le gain de lissage.

Remarques:

– Pour $n = 0$, la d.d.p de $\mathbf{t}_{0|N}$ se calcule selon

$$\begin{aligned}\mathbf{K}_{0|N} &= \mathbf{Q}_{0|0} [\mathbf{F}_2^{x,t}]^T \mathbf{P}_{1|0}^{-1}, \\ \hat{\mathbf{t}}_{0|N} &= \hat{\mathbf{x}}_{0|0} + \mathbf{K}_{0|N} [\hat{\mathbf{x}}_{1|N} - \hat{\mathbf{x}}_{1|0}], \\ \mathbf{Q}_{0|N} &= \mathbf{Q}_{0|0} + \mathbf{K}_{0|N} [\mathbf{P}_{1|N} - \mathbf{P}_{1|0}] \mathbf{K}_{0|N}^T.\end{aligned}\quad (1.10)$$

- Pour l’algorithme EM, la covariance entre les états lissés successifs $\mathbf{x}_{n|N}$ et $\mathbf{x}_{n+1|N}$ est requise et est calculée grace à

$$\mathbf{C}_{n,n+1|N} = \mathbf{K}_{n|N} \mathbf{P}_{n+1|N}. \quad (1.11)$$

Fonction auxiliaire de vraisemblance :

Dans le modèle du filtre de Kalman couple, la fonction auxiliaire de vraisemblance [3] s’écrit

$$\begin{aligned} g(\hat{\mathbf{t}}_0, \mathbf{Q}_0, \mathbf{F}, \mathbf{Q}) &= E \left(\log(\mathcal{L}(\boldsymbol{\theta}; \mathbf{x}, \mathbf{y})) \mid \mathbf{y}, \boldsymbol{\theta}^{(j)} \right) \\ &= \underbrace{\frac{1}{2} \log |\mathbf{Q}_0^{-1}| - \frac{1}{2} \text{tr} [\mathbf{Q}_0^{-1} \mathbf{Q}_{0|N}] - [\hat{\mathbf{t}}_{0|N} - \hat{\mathbf{t}}_0]^T \mathbf{Q}_0^{-1} [\hat{\mathbf{t}}_{0|N} - \hat{\mathbf{t}}_0]}_{g_0(\hat{\mathbf{t}}_0, \mathbf{Q}_0)} \\ &\quad - \underbrace{\frac{1}{2} [N+1] \log |\mathbf{Q}| - \frac{1}{2} \text{tr} \left[\mathbf{Q}^{-1} \begin{bmatrix} \mathbf{F}^T \\ -\mathbf{I}_{n_t} \end{bmatrix}^T \tilde{\mathbf{C}} \begin{bmatrix} \mathbf{F}^T \\ -\mathbf{I}_{n_t} \end{bmatrix} \right]}_{g_{\tilde{\mathbf{C}}}(\mathbf{Q}, \mathbf{F})} \\ &\quad - \frac{n_t}{2} [N+1] \log [2\pi], \end{aligned} \quad (1.12)$$

où

$$\tilde{\mathbf{C}} = \begin{bmatrix} \tilde{\mathbf{C}}_{0,0} & \tilde{\mathbf{C}}_{0,1} \\ \tilde{\mathbf{C}}_{0,1}^T & \tilde{\mathbf{C}}_{1,1} \end{bmatrix} \quad (1.13)$$

est la somme des corrélations des vecteurs $[\mathbf{t}_{n|N}^T, \mathbf{t}_{n+1|N}^T]$ avec

$$\begin{aligned} \tilde{\mathbf{C}}_{0,0} &= \mathbf{Q}_{0|N} + \hat{\mathbf{t}}_{0|N} \hat{\mathbf{t}}_{0|N}^T + \sum_{n=1}^N \begin{bmatrix} \mathbf{P}_{n|N} & \mathbf{O}_{n_x, n_y} \\ \mathbf{O}_{n_y, n_x} & \mathbf{O}_{n_y, n_y} \end{bmatrix} + \hat{\mathbf{t}}_{n|N} \hat{\mathbf{t}}_{n|N}^T, \\ \tilde{\mathbf{C}}_{0,1} &= [\mathbf{C}_{0,1|N} \quad \mathbf{O}_{n_t, n_y}] + \hat{\mathbf{t}}_{0|N} \hat{\mathbf{t}}_{1|N}^T + \sum_{n=1}^N \begin{bmatrix} \mathbf{C}_{n,n+1|N} & \mathbf{O}_{n_x, n_y} \\ \mathbf{O}_{n_y, n_x} & \mathbf{O}_{n_y, n_y} \end{bmatrix} + \hat{\mathbf{t}}_{n|N} \hat{\mathbf{t}}_{n+1|N}^T, \\ \tilde{\mathbf{C}}_{1,1} &= \sum_{n=0}^N \begin{bmatrix} \mathbf{P}_{n+1|N} & \mathbf{O}_{n_x, n_y} \\ \mathbf{O}_{n_y, n_x} & \mathbf{O}_{n_y, n_y} \end{bmatrix} + \hat{\mathbf{t}}_{n+1|N} \hat{\mathbf{t}}_{n+1|N}^T. \end{aligned} \quad (1.14)$$

Cette fonction (1.12) se décompose donc en une somme de deux fonctions indépendantes et d’une constante. La première sous-fonction $g_0(\hat{\mathbf{t}}_0, \mathbf{Q}_0)$ dépend uniquement de l’état initial et la seconde $g_{\tilde{\mathbf{C}}}(\mathbf{Q}, \mathbf{F})$ dépend des matrices de covariance du bruit \mathbf{Q} et de transition \mathbf{F} .

1.2 Étape M

Les paramètres estimés à l’itération j sont obtenus en maximisant la fonction auxiliaire de vraisemblance (1.12). Le jeu de paramètres $\boldsymbol{\theta}^{(j)} = [\hat{\mathbf{t}}_0^{(j)}, \mathbf{Q}_0^{(j)}, \mathbf{F}^{(j)}, \mathbf{Q}^{(j)}]$

qui la maximise est donné par (cf. [3])

$$\begin{aligned}
 \hat{\mathbf{t}}_0^{(j)} &= \hat{\mathbf{t}}_{0|N}, \\
 \mathbf{Q}_0^{(j)} &= \mathbf{Q}_{0|N}, \\
 \mathbf{F}^{(j)} &= \tilde{\mathbf{C}}_{0,1}^T \tilde{\mathbf{C}}_{0,0}^{-1}, \\
 \mathbf{Q}^{(j)} &= \frac{1}{N+1} \left[\tilde{\mathbf{C}}_{1,1} - \tilde{\mathbf{C}}_{0,1}^T \tilde{\mathbf{C}}_{0,0}^{-1} \tilde{\mathbf{C}}_{0,1} \right].
 \end{aligned} \tag{1.15}$$

Cependant, à l'instar du filtre de Kalman, cette implémentation conduit à une instabilité numérique à laquelle nous apporterons une solution dans la section 2.

2 Implémentation du filtre robuste

Cette section est dédiée à la description de l'implémentation robuste du filtre de Kalman couple. Nous reprenons les algorithmes robustes développés pour le filtre de Kalman [47] et les étendons au cas du filtre de Kalman couple. Ces méthodes garantissent la symétrie et la positivité des matrices de covariance ou de corrélation par la propagation dans l'algorithme itératif, de leurs racines carrées. En effet, toute matrice symétrique positive peut s'écrire $\mathbf{S} = \mathbf{N}^T \mathbf{N}$. Gardées sous cette forme, les matrices de covariance ou de corrélation ne peuvent pas devenir asymétriques ou non-positives. Parmi toutes les méthodes de factorisation, nous avons retenu la méthode basée sur des décompositions QR qui a l'avantage d'être rapide, et de doubler la précision des calculs.

Toutes les décompositions QR qui sont utilisées dans ce manuscrit sont basées sur le calcul d'une racine carrée des matrices symétriques et positives de la forme générale suivante :

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} + \mathbf{M}_1^T \mathbf{B} \mathbf{M}_1 & \delta_1 \mathbf{M}_1^T \mathbf{B} & \mathbf{O} \\ \delta_1 \mathbf{B} \mathbf{M}_1 & \delta_1 [\mathbf{B} + \delta_2 \mathbf{M}_2^T \mathbf{C} \mathbf{M}_2] & \delta_1 \delta_2 \delta_3 \mathbf{M}_2^T \mathbf{C} \\ \mathbf{O} & \delta_1 \delta_2 \delta_3 \mathbf{C} \mathbf{M}_2 & \delta_1 \delta_2 \delta_3 \mathbf{C} \end{bmatrix}, \tag{1.16}$$

où \mathbf{A} , \mathbf{B} et \mathbf{C} sont des matrices symétriques positives et δ_1 , δ_2 et δ_3 des valeurs booléennes. Une racine carrée triviale de la matrice \mathbf{M} est donnée par

$$\mathbf{N}_{input} = \begin{bmatrix} \mathbf{A}^{\frac{1}{2}} & \mathbf{O} & \mathbf{O} \\ \mathbf{B}^{\frac{1}{2}} \mathbf{M}_1 & \delta_1 \mathbf{B}^{\frac{1}{2}} & \mathbf{O} \\ \mathbf{O} & \delta_1 \delta_2 \mathbf{C}^{\frac{1}{2}} \mathbf{M}_2 & \delta_1 \delta_2 \delta_3 \mathbf{C}^{\frac{1}{2}} \end{bmatrix}.$$

Avec l'aide d'une décomposition QR de la matrice \mathbf{N}_{input} , nous obtenons une racine triangulaire supérieure par blocs \mathbf{N}_{output} de \mathbf{M} de la forme

$$\mathbf{N}_{output} = \begin{bmatrix} \mathbf{T}_{1,1} & \mathbf{T}_{1,2} & \mathbf{O} \\ \mathbf{O} & \mathbf{T}_{2,2} & \mathbf{T}_{2,3} \\ \mathbf{O} & \mathbf{O} & \mathbf{T}_{3,3} \end{bmatrix}.$$

Par identification des termes de \mathbf{M} (1.16) et de $\mathbf{N}_{output}^T \mathbf{N}_{output}$, nous obtenons

- $T_{1,1} = [A + M_1^T B M_1]^{\frac{1}{2}}$,
- $T_{1,2} = \delta_1 T_{1,1}^{-T} M_1^T B$,
- $T_{2,2} = \delta_1 [B - B M_1 T_{1,1}^{-1} T_{1,1}^{-T} M_1^T B + \delta_2 M_2^T C M_2]^{\frac{1}{2}}$,
- $T_{2,3} = \delta_1 \delta_2 \delta_3 T_{2,2}^{-T} M_2^T C$ et
- $T_{3,3} = \delta_1 \delta_2 \delta_3 [C - C M_2 T_{2,2}^{-1} T_{2,2}^{-T} M_2^T C]^{\frac{1}{2}}$.

Remarques:

- Si une colonne ou une ligne est nulle dans la matrice N_{input} , alors il est équivalent d'utiliser une matrice N_{input}^* où la ligne ou colonne nulle est supprimée.
- L'algorithme robuste du filtre de Kalman couple modifie les paramètres requis pour l'estimation des états cachés : les matrices de covariance du bruit Q_0 et Q sont remplacées par une de leurs racines carrées respectives notées $Q_0^{\frac{1}{2}}$ et $Q^{\frac{1}{2}}$.

Dans la suite, nous décrivons les modifications à apporter à l'algorithme original de la section 1 pour propager les racines carrées des matrices de covariance.

2.1 Étape E

Le calculs des espérances des états cachés $\hat{x}_{n|p}$, $\hat{t}_{n|p}$ et des gains de Kalman couple $K_{n|p}$ restent ici identiques à ceux de l'algorithme original. Par contre, les calculs des matrices de covariance sont remplacés par ceux de leurs racines carrées respectives.

Remarque: Toutes matrices unitaires Z^* introduites par les décompositions QR ne sont jamais ré-utilisées.

Constantes :

Lors de la phase de calcul des constantes, les racines carrées des matrices de covariance $Q_2^{x,x}$ et $Q^{y,y}$, utilisées dans (1.2), sont calculées à partir de la décomposition QR suivante :

$$\underbrace{\begin{bmatrix} [Q^{y,y}]^{\frac{1}{2}} & [Q^{y,y}]^{-\frac{1}{2}} \\ O_{n_x, n_y} & [Q_2^{x,x}]^{\frac{1}{2}} \end{bmatrix}}_{N_{output}^c} = Z^* \underbrace{\begin{bmatrix} \underbrace{\left[Q^{y,y} - [Q^{x,y}]^T [Q^{x,x}]^{-1} Q^{x,y} \right]^{\frac{1}{2}}}_{E^{y,y}} & O_{n_y, n_x} \\ \underbrace{\left[Q^{x,x} \right]^{-\frac{1}{2}} Q^{x,y}}_{E^{x,y}} & \underbrace{\left[Q^{x,x} \right]^{\frac{1}{2}}}_{E^{x,x}} \end{bmatrix}}_{N_{input}^c} \quad (1.17)$$

et où $Q^{\frac{1}{2}} = \begin{bmatrix} E^{x,x} & E^{x,y} \\ O_{n_y, n_x} & E^{y,y} \end{bmatrix}$. Ce résultat est obtenu en posant $A = Q^{y,y} - [Q^{x,y}]^T [Q^{x,x}]^{-1} Q^{x,y}$, $B = Q^{x,x}$, $M_1 = Q^{x,y}$, $\delta_1 = 1$ et $\delta_2 = \delta_3 = 0$ dans l'équation (1.16).

Prédiction :

Lors des phases de prédiction de l'état caché, les racines carrées de $P_{n+1|n}$ (1.3) se déduisent de la décomposition QR suivante :

$$\underbrace{\begin{bmatrix} P_{n+1|n}^{\frac{1}{2}} \\ O \end{bmatrix}}_{N_{output}^p} = Z^* \underbrace{\begin{bmatrix} [Q_2^{x,x}]^{\frac{1}{2}} \\ P_{n|n}^{\frac{1}{2}} [F_2^{x,x}]^T \end{bmatrix}}_{N_{input}^p}. \quad (1.18)$$

Ce résultat est obtenu en posant $A = Q_2^{x,x}$, $B = P_{n|n}$, $M_1 = [F_2^{x,x}]^T$, et $\delta_1 = \delta_2 = \delta_3 = 0$ dans l'équation (1.16).

Remarque: Pour $n = 1$ (1.4), on remplace la seconde ligne de N_{input}^p par $Q_{0|0}^{\frac{1}{2}} [F_2^{x,t}]^T$.

Filtrage :

Lors de la phase de filtrage, les racines carrées de $P_{n+1|n+1}$ et $S_{n+1|n}$, présentées dans (1.5) et (1.7), se calculent à l'aide de la décomposition QR suivante :

$$\underbrace{\begin{bmatrix} S_{n+1|n}^{\frac{1}{2}} & S_{n+1|n}^{\frac{1}{2}} K_{n+1|n+1}^T \\ O_{n_x, n_y} & P_{n+1|n+1}^{\frac{1}{2}} \end{bmatrix}}_{N_{output}^f} = Z^* \underbrace{\begin{bmatrix} [Q^{y,y}]^{\frac{1}{2}} & O_{n_y, n_x} \\ P_{n|n-1}^{\frac{1}{2}} [F^{y,x}]^T & P_{n|n-1}^{\frac{1}{2}} \end{bmatrix}}_{N_{input}^f}. \quad (1.19)$$

Ce résultat est obtenu en posant $A = Q^{y,y}$, $B = P_{n+1|n}$, $M_1 = [F^{y,x}]^T$, $\delta_1 = 1$ et $\delta_2 = \delta_3 = 0$ dans l'équation (1.16).

Remarque: Pour $n = 0$ (1.6), (1.8), nous remplaçons

- dans N_{input}^f , $P_{n+1|n}^{\frac{1}{2}}$, $[F^{y,x}]^T$ respectivement par $Q_0^{\frac{1}{2}}$, $[F^{y,t}]^T$ et
- dans N_{output}^f , $P_{n|n}^{\frac{1}{2}}$ par $Q_{0|0}^{\frac{1}{2}}$.

Lissage :

Lors de la phase de lissage, une racine des matrices de covariance $P_{n|N}$ (1.9) se calcule à l'aide de la décomposition QR suivante :

$$\underbrace{\begin{bmatrix} P_{n+1|n}^{\frac{1}{2}} & P_{n+1|n}^{\frac{1}{2}} K_{n|N}^T \\ O_{n_x, n_x} & P_{n|N}^{\frac{1}{2}} \\ O_{n_x, n_x} & O_{n_x, n_x} \end{bmatrix}}_{N_{output}^s} = Z^* \underbrace{\begin{bmatrix} [Q_2^{x,x}]^{\frac{1}{2}} & O_{n_x, n_x} \\ P_{n|n}^{\frac{1}{2}} [F_2^{x,x}]^T & P_{n|n}^{\frac{1}{2}} \\ O_{n_x, n_x} & P_{n+1|N}^{\frac{1}{2}} K_{n|N}^T \end{bmatrix}}_{N_{input}^s} \dots \quad (1.20)$$

Ce résultat est obtenu en posant $A = Q_2^{x,x}$, $B = P_{n|n}$, $C = P_{n+1|N}$, $M_1 = [F_2^{x,x}]^T$, $M_2 = K_{n|N}^T$, $\delta_1 = \delta_2 = 1$ et $\delta_3 = 0$ dans l'équation (1.16).

Remarque: Pour $n = 0$ (1.10), nous remplaçons

- dans N_{input}^s , $P_{n|n}^{\frac{1}{2}}$, $[F_2^{x,x}]^T$ respectivement par $Q_{0|0}^{\frac{1}{2}}$, $[F_2^{x,t}]^T$ et
- dans N_{output}^s , $P_{n|N}^{\frac{1}{2}}$ par $Q_{0|N}^{\frac{1}{2}}$.

Dans le cadre d'un algorithme EM, le calcul de $C_{n+1,n|N}$ (1.11) n'est plus requis et est remplacé par la quantité suivante

$$C_{n+1,n|N}^* = P_{n+1|n}^{\frac{1}{2}} K_{n|N}^T. \quad (1.21)$$

Somme des corrélations :

Le calcul de $\tilde{C}^{\frac{1}{2}}$ (1.13) s'effectue en quatre étapes, synthétisées par la figure 1.2 :

1. La racine de la matrice de covariance du vecteur $[\mathbf{x}_{n|N}^T, \mathbf{x}_{n+1|N}^T]^T$, dénotée $C_{n,n+1|N}^x$, est requise et se calcule à l'aide de la décomposition QR suivante :

$$\underbrace{\begin{bmatrix} P_{n+1|n}^{\frac{1}{2}} & \begin{bmatrix} P_{n+1|n}^{\frac{1}{2}} K_{n|N}^T & O_{n_x, n_x} \end{bmatrix} \\ O_{2n_x, n_x} & \underbrace{[C_{n,n+1|N}^x]^{\frac{1}{2}}} \end{bmatrix}}_{N_{output}^{cov}} = Z^* \underbrace{\begin{bmatrix} [Q_2^{x,x}]^{\frac{1}{2}} & O_{n_x, n_x} & O_{n_x, n_x} \\ P_{n|n}^{\frac{1}{2}} [F_2^{x,x}]^T & P_{n|n}^{\frac{1}{2}} & O_{n_x, n_x} \\ O_{n_x, n_x} & C_{n+1,n|N}^* & P_{n+1|N}^{\frac{1}{2}} \end{bmatrix}}_{N_{input}^{cov}}. \quad (1.22)$$

Remarque: Pour $n = 0$, nous devons calculer la racine carrée de la matrice de covariance $C_{n,n+1|N}^{tx}$ du vecteur aléatoire $[\mathbf{t}_{0|N}^T, \mathbf{x}_{1|N}^T]^T$. Dans ce cas, il faut remplacer

$P_{n|n}^{\frac{1}{2}}$, $[F_2^{x,x}]^T$ et $C_{n+1,n|N}^*$ respectivement par $Q_{0|0}^{\frac{1}{2}}$, $[F_2^{x,t}]^T$ et $C_{1,0|N}^*$.

Ce résultat est obtenu en posant $A = Q_2^{x,x}$, $B = P_{n|n}$, $C = P_{n+1|N}$, $M_1 = [F_2^{x,x}]^T$, $M_2 = K_{n|N}^T$, $\delta_1 = \delta_2 = 1$ et $\delta_3 = 1$ dans l'équation (1.16).

Pour la suite, nous décomposons en blocs la matrice

$$[C_{n,n+1|N}^x]^{\frac{1}{2}} = \begin{bmatrix} C_{0,0}^x & C_{0,1}^x \\ O_{n_x, n_x} & C_{1,1}^x \end{bmatrix},$$

où chaque sous-matrice $C_{i,j}^x$ est de dimension $(n_x \times n_x)$.

2. La racine carrée de la matrice de covariance du vecteur $[\mathbf{t}_{n|N}^T, \mathbf{t}_{n+1|N}^T]^T$, $C_{n,n+1|N}^t$ est requise et se calcule à partir de $[C_{n,n+1|N}^x]^{\frac{1}{2}}$ selon

$$[C_{n,n+1|N}^t]^{\frac{1}{2}} = \begin{bmatrix} C_{0,0}^x & O_{n_x, n_y} & C_{0,1}^x & O_{n_x, n_y} \\ O_{n_y, n_x} & O_{n_y, n_y} & O_{n_y, n_x} & O_{n_y, n_y} \\ O_{n_x, n_x} & O_{n_x, n_y} & C_{1,1}^x & O_{n_x, n_y} \\ O_{n_y, n_x} & O_{n_y, n_y} & O_{n_y, n_x} & O_{n_y, n_y} \end{bmatrix}. \quad (1.23)$$

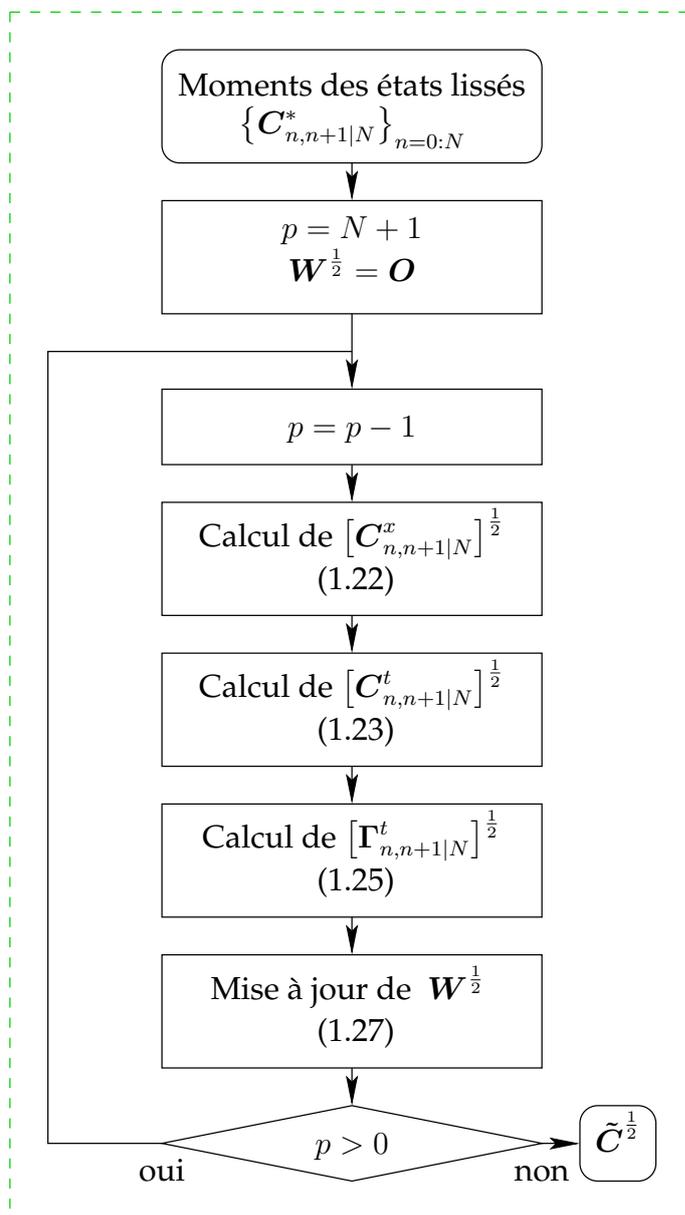


FIGURE 1.2: Calcul de $\tilde{C}^{\frac{1}{2}}$.

Remarque: Pour $n = 0$, la racine de la matrice de covariance $[\mathbf{C}_{0,1|N}^t]^{\frac{1}{2}}$ se calcule différemment

$$[\mathbf{C}_{0,1|N}^t]^{\frac{1}{2}} = \begin{bmatrix} \mathbf{C}_{0,0}^{tx} & \mathbf{C}_{0,1}^{tx} & \mathbf{O}_{n_t, n_y} \\ \mathbf{O}_{n_x, n_t} & \mathbf{C}_{1,1}^{tx} & \mathbf{O}_{n_x, n_y} \\ \mathbf{O}_{n_y, n_t} & \mathbf{O}_{n_y, n_x} & \mathbf{O}_{n_y, n_y} \end{bmatrix}, \quad (1.24)$$

où les sous-matrices $\mathbf{C}_{0,0}^{tx}$, $\mathbf{C}_{0,1}^{tx}$ et $\mathbf{C}_{1,1}^{tx}$ ont comme dimensions respectives $(n_t \times n_t)$, $(n_t \times n_x)$ et $(n_x \times n_x)$.

3. La racine carrée de la matrice d'auto-corrélation du vecteur $[\mathbf{t}_{n|N}^T, \mathbf{t}_{n+1|N}^T]^T$, notée $\mathbf{\Gamma}_{n,n+1|N}^t$, se calcule à l'aide de la décomposition QR suivante :

$$\underbrace{\begin{bmatrix} [\mathbf{\Gamma}_{n,n+1|N}^t]^{\frac{1}{2}} \\ \mathbf{O}_{1, n_t} \end{bmatrix}}_{N_{output}^{\Gamma}} = \mathbf{Z}^* \underbrace{\begin{bmatrix} [\mathbf{C}_{n,n+1|N}^t]^{\frac{1}{2}} \\ [\hat{\mathbf{t}}_{n|N}^T, \hat{\mathbf{t}}_{n+1|N}^T] \end{bmatrix}}_{N_{input}^{\Gamma}}. \quad (1.25)$$

Remarque: Pour tout n dans $[1, N]$, l'espérance $\hat{\mathbf{t}}_{n|N}$ est donnée par

$$\hat{\mathbf{t}}_{n|N} = \begin{bmatrix} \hat{\mathbf{x}}_{n|N} \\ \mathbf{y}_{n-1} \end{bmatrix}. \quad (1.26)$$

4. Finalement, les racines carrées des matrices $\mathbf{\Gamma}_{n,n+1|N}^t$ sont réutilisées pour calculer une racine carrée de la somme des corrélations $\tilde{\mathbf{C}} = \sum_{n=0}^N \mathbf{\Gamma}_{n,n+1|N}^t$. Dans ce but, un algorithme itératif qui calcule récursivement la racine de $\mathbf{W}_p = \sum_{n=p}^N \mathbf{\Gamma}_{n,n+1|N}^t$ est utilisé :

- Pour $p = N$, $\mathbf{W}_{\frac{1}{N}} = \mathbf{O}_{2n_t, 2n_t}$.
- Pour p allant de $N - 1$ à 0 , nous utilisons la décomposition QR suivante :

$$\underbrace{\begin{bmatrix} \mathbf{W}_{\frac{1}{p}} \\ \mathbf{O}_{2n_t, 2n_t} \end{bmatrix}}_{N_{output}^W} = \mathbf{Z}^* \underbrace{\begin{bmatrix} \mathbf{W}_{\frac{1}{p+1}} \\ [\mathbf{\Gamma}_{p,p+1|N}^t]^{\frac{1}{2}} \end{bmatrix}}_{N_{input}^W}. \quad (1.27)$$

- Finalement, pour $p = 0$, on obtient $\tilde{\mathbf{C}}^{\frac{1}{2}} = \mathbf{W}_0^{\frac{1}{2}}$.

2.2 Étape M

La ré-estimation robuste des matrices $\mathbf{Q}^{(j)}$ et $\mathbf{F}^{(j)}$ à l'itération j de l'algorithme EM est obtenue directement à partir de $\tilde{\mathbf{C}}^{\frac{1}{2}}$ car nous avons

$$\tilde{\mathbf{C}}^{\frac{1}{2}} = \begin{bmatrix} \tilde{\mathbf{C}}_{0,0}^{\frac{1}{2}} & \tilde{\mathbf{C}}_{0,0}^{\frac{1}{2}} [\mathbf{F}^{(i)}]^T \\ \mathbf{O}_{n_t, n_t} & \sqrt{N+1} [\mathbf{Q}^{(i)}]^{\frac{1}{2}} \end{bmatrix}. \quad (1.28)$$

Dans cette section, nous avons présenté l'implémentation robuste du filtre de Kalman couple. Dans la section 3, nous présentons une expérience qui illustre sa robustesse vis-à-vis du filtre original.

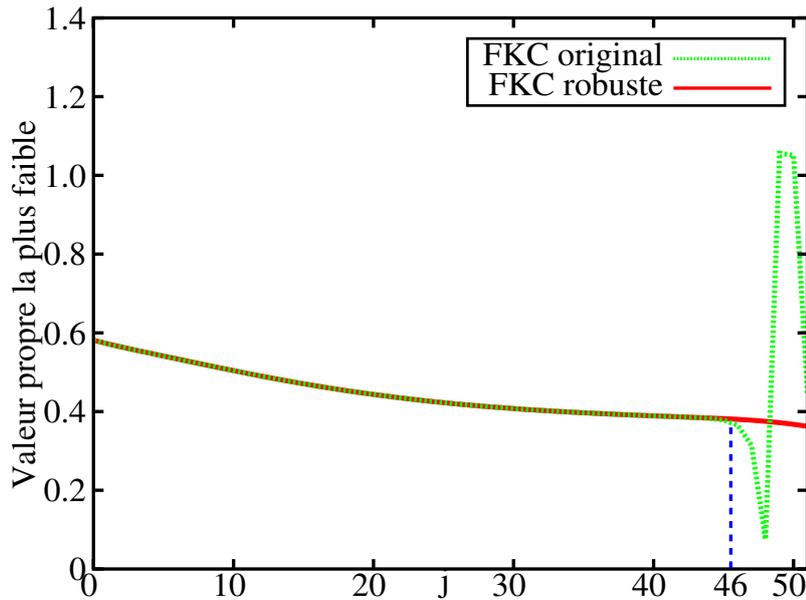


FIGURE 1.3: Évolution de la plus petite valeur propre de Q au cours des itérations de l'algorithme EM.

3 Illustration de la robustesse

Dans le but d'illustrer à la fois la perte de positivité de la matrice de covariance Q de l'algorithme original et le bon comportement de l'algorithme robuste, l'expérience suivante a été réalisée : un signal de 100 points a été simulé selon le modèle du filtre de Kalman couple suivant :

$$\begin{aligned} \hat{t}_0 &= \mathbf{O}_{4,1}, \\ \mathbf{Q}_0 &= \begin{bmatrix} \mathbf{I}_2 & \mathbf{O}_{2,2} \\ \mathbf{O}_{2,2} & \mathbf{O}_{2,2} \end{bmatrix}, \\ \mathbf{F} &= \begin{bmatrix} 0.5 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.2 & 0.1 & 0.1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \\ \mathbf{Q} &= 0.5\mathbf{I}_4. \end{aligned}$$

Les paramètres ont ensuite été appris séparément par l'algorithme EM original et par l'algorithme EM robuste, tous deux initialisés avec les paramètres suivants :

$$\begin{aligned} \mathbf{F}^{(0)} &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \\ \mathbf{Q}^{(0)} &= \mathbf{I}_4. \end{aligned}$$

Pour mettre en évidence la perte de la positivité de Q dans l'algorithme original du filtre de Kalman couple, la plus petite valeur propre de Q a été calculée . La

figure 1.3 met en valeur de graves instabilités de l'algorithme original après 50 itérations, contrairement à l'algorithme robuste.

4 Conclusion

Les instabilités numériques du filtre de Kalman lors de la propagation des matrices de covariance sont bien connues dans la littérature. La principale contribution de ce chapitre a été de proposer un algorithme EM robuste (incluant les algorithmes de prédiction, filtrage et lissage) en étendant au modèle du filtre de Kalman couple, une méthode basée sur des décompositions QR, proposée dans le cadre du filtre de Kalman classique.

L'évaluation des performances de l'algorithme EM du filtre de Kalman couple sera réalisée dans les chapitres 2, 3 et 4 par comparaison avec d'autres algorithmes que nous présenterons. En effet, nous montrons dans le chapitre suivant que le modèle du filtre de Kalman couple souffre d'un problème d'identifiabilité qui doit être pris en compte pour restaurer correctement dans le cadre non-supervisé, des données bruitées.

Chapitre 2

Filtres de Kalman couple équivalents

Nous avons présenté dans le chapitre 1, un algorithme EM pour l'estimation non-supervisée des paramètres d'un filtre de Kalman couple quelconque. Nous nous intéressons ici au problème de l'identifiabilité du modèle du filtre de Kalman couple. En effet, pour un jeu d'observations donné \mathbf{y} , il existe une infinité de filtres de Kalman couples qui partagent exactement la même fonction de vraisemblance : si le filtre original est paramétré par le jeu $\boldsymbol{\theta} = \{\hat{\mathbf{t}}_0, \mathbf{Q}_0, \mathbf{F}, \mathbf{Q}\}$, alors comme nous le démontrons ci-dessous, les paramètres de l'ensemble des filtres équivalents se déduisent exactement par

$$\begin{aligned}\hat{\mathbf{t}}_0^M &= M\hat{\mathbf{t}}_0, \\ \mathbf{Q}_0^M &= M\mathbf{Q}_0M^T, \\ \mathbf{F}^M &= M\mathbf{F}M^{-1}, \\ \mathbf{Q}^M &= M\mathbf{Q}M^T,\end{aligned}\tag{2.1}$$

avec

$$M = \begin{bmatrix} M^{x,x} & M^{x,y} \\ \mathbf{O}_{n_y,n_x} & \mathbf{I}_{n_y} \end{bmatrix},\tag{2.2}$$

une matrice inversible. De tels filtres seront appelé filtres de Kalman couple équivalents (FKCE).

Remarques:

- L'inversibilité de $M^{x,x}$ est suffisante pour prouver l'inversibilité de M .
- Les termes \mathbf{O}_{n_y,n_x} et \mathbf{I}_{n_y} dans M permettent de conserver exactement le même jeu d'observations avant et après transformation.
- Nous pouvons obtenir directement les racines carrées des matrices \mathbf{Q}^M et \mathbf{Q}_0^M avec l'aide des décompositions QR suivantes

$$\begin{aligned}[\mathbf{Q}^M]^{\frac{1}{2}} &= \mathbf{Z}^* \mathbf{Q}^{\frac{1}{2}} M^T, \\ [\mathbf{Q}_0^M]^{\frac{1}{2}} &= \mathbf{Z}^* \mathbf{Q}_0^{\frac{1}{2}} M^T.\end{aligned}$$

Démonstration:

- Nous cherchons à montrer d’abord que le processus t^M suit le modèle du filtre de Kalman couple paramétré par θ^M .

Le processus t suit un modèle de Kalman couple, donc

$$\begin{aligned} t_0 &= \hat{t}_0 + \omega_0, \\ Mt_0 &= M\hat{t}_0 + M\omega_0, \\ t_0^M &= \hat{t}_0^M + \omega_0^M. \end{aligned}$$

Donc, t_0^M suit la loi Gaussienne $\mathcal{N}(t_0^M, MQ_0M^T)$. De même,

$$\begin{aligned} t_{n+1} &= Ft_n + \omega_{n+1}, \\ Mt_{n+1} &= MFt_n + M\omega_{n+1}. \end{aligned}$$

La matrice M étant, par hypothèse, inversible, nous pouvons écrire :

$$\begin{aligned} t_{n+1}^M &= MF M^{-1} Mt_n + \omega_{n+1}^M, \\ t_{n+1}^M &= MF M^{-1} t_n^M + \omega_{n+1}^M, \\ t_{n+1}^M &= F^M t_n^M + \omega_{n+1}^M, \end{aligned}$$

avec ω_{n+1}^M une v.a. de loi gaussienne $\mathcal{N}(O, MQM^T)$.

- Nous cherchons à montrer que la transformation par M n’affecte pas la vraisemblance.

Le logarithme de la vraisemblance s’écrit :

$$\log \mathcal{L}(\mathbf{y} | \theta) = \sum_{n=0}^N \log |S_{n|n-1}| + \tilde{\mathbf{y}}_{n|n-1}^T S_{n|n-1}^{-1} \tilde{\mathbf{y}}_{n|n-1}.$$

Pour prouver l’invariance de la vraisemblance, il suffit de prouver l’invariance des moments de l’innovation : à partir de la relation (2.1), nous en déduisons l’espérance et la covariance de la v.a. $t_{n|p}$:

$$\begin{aligned} \hat{t}_{n|p}^M &= M\hat{t}_{n|p}, \\ Q_{n|p}^M &= MQ_{n|p}M^T. \end{aligned}$$

La forme particulière de M permet de déduire que

$$\begin{aligned} \hat{\mathbf{y}}_{n|p}^M &= \hat{\mathbf{y}}_{n|p}, \\ S_{n|p}^M &= S_{n|p}, \end{aligned}$$

et donc

$$\begin{aligned} \tilde{\mathbf{y}}_{n+1|n}^M &= \tilde{\mathbf{y}}_{n+1|n}, \\ S_{n+1|n}^M &= S_{n+1|n}. \end{aligned}$$

Donc la fonction de vraisemblance est invariante pour les jeux de paramètres équivalents.

La prédiction, le filtrage et le lissage pour le modèle équivalent se déduisent de ceux du modèle original car nous avons les relations suivantes :

$$\begin{aligned}\hat{\mathbf{t}}_{n|p}^M &= M\hat{\mathbf{t}}_{n|p}, \\ \mathbf{Q}_{n|p}^M &= M\mathbf{Q}_{n|p}M^T.\end{aligned}\tag{2.3}$$

Tous les jeux de paramètres équivalents se déduisent de la relation (2.1). Ainsi, le seul critère de la vraisemblance ne permet pas de distinguer ces différents jeux de paramètres. Par exemple, si nous prenons le jeu de paramètres équivalent donné par la matrice de transformation

$$M^{opp} = \begin{bmatrix} -I_{n_x} & \mathbf{O} \\ \mathbf{O} & I_{n_y} \end{bmatrix},\tag{2.4}$$

alors les états cachés équivalents restaurés seront exactement les opposés des originaux ! Cela pose problème pour l'estimation non-supervisée des paramètres et des états cachés.

Nous étudions ce problème en trois étapes. Dans un premier temps, nous démontrons les propriétés des jeux de paramètres équivalents dans l'algorithme EM. Puis, dans un second temps, nous proposons de choisir un jeu de paramètres équivalent qui conserve une structure particulière pour les matrices \mathbf{F} et \mathbf{Q} . Finalement, dans un dernier temps, nous évaluons l'utilisation d'un jeu de paramètres équivalent particulier à la place du jeu de paramètres estimé par l'algorithme EM original, et comparons les performances vis-à-vis du filtre optimal.

1 Propriété

Une propriété importante des jeux de paramètres équivalents, analogue à une propriété de l'estimateur du maximum de vraisemblance, est que si l'on initialise deux algorithmes EM, l'un sur le jeu de paramètres $\boldsymbol{\theta}^{(0)}$ et l'autre sur un jeu de paramètres équivalent $[\boldsymbol{\theta}^{(0)}]^M$, alors si le premier algorithme estime $\boldsymbol{\theta}^{(j)}$, alors le second estime exactement $[\boldsymbol{\theta}^{(j)}]^M$ à chaque itération j de l'algorithme EM.

Démonstration: Montrons cette propriété par récurrence.

- Pour $j = 0$, ce sont les hypothèses de départ.
- Si on suppose que cette propriété est vérifiée à l'itération $j - 1$, alors la relation (2.3) permet de déduire que à l'itération j , si le filtre original calcule la somme $\tilde{\mathbf{C}}$, le filtre équivalent calculera la somme

$$\tilde{\mathbf{C}}^M = \begin{bmatrix} \mathbf{M} & \mathbf{O}_{n_t, n_t} \\ \mathbf{O}_{n_t, n_t} & \mathbf{M} \end{bmatrix} \tilde{\mathbf{C}} \begin{bmatrix} \mathbf{M} & \mathbf{O}_{n_t, n_t} \\ \mathbf{O}_{n_t, n_t} & \mathbf{M} \end{bmatrix}^T.$$

Alors, nous pouvons calculer, à l'itération j , les paramètres estimés par le

second filtre en fonction de ceux estimés par le premier filtre :

$$\begin{aligned} [\hat{\mathbf{t}}_0^M]^{(j)} &= M\hat{\mathbf{t}}_0^{(j)}, \\ [\mathbf{Q}_0^M]^{(j)} &= M\mathbf{Q}_0^{(j)}M^T, \\ [\mathbf{F}^M]^{(j)} &= M\mathbf{F}^{(j)}M^{-1}, \\ [\mathbf{Q}^M]^{(j)} &= M\mathbf{Q}^{(j)}M^T. \end{aligned}$$

Les paramètres estimés par le second filtre sont donc les équivalents par la matrice de transformation M de ceux estimés par le premier filtre.

Cette propriété montre que si nous souhaitons remplacer le jeu de paramètres estimé dans l'étape M de l'algorithme EM par un jeu de paramètres équivalent, alors

1. d'une part, le chemin parcouru par l'algorithme EM dans la fonction de vraisemblance est exactement le même pour tous les filtres équivalents ;
2. et si, de plus, nous imposons une forme particulière au jeu de paramètres, le faire à chaque étape de maximisation ou après l'estimation finale de l'algorithme original, donnera exactement le même résultat.

La dernière remarque est très intéressante dans la pratique puisque si nous souhaitons imposer une forme particulière aux paramètres, nous pouvons le faire après la fin de l'algorithme EM original.

Nous proposons dans la section suivante d'étudier les contraintes possibles sur F et Q qui préservent la fonction de vraisemblance.

2 Paramètres équivalents : 3 cas intéressants

Cette section est dédiée à l'utilisation des jeux de paramètres équivalents pour imposer une structure particulière soit à la matrice de transition F , soit à la matrice de covariance Q . Cependant, très peu de contraintes permettent d'imposer à la fois les formes souhaitées aux matrices F et Q et de préserver la vraisemblance. En effet, pour toute matrice de transformation M de la forme

$$M = \begin{bmatrix} M^{x,x} & M^{x,y} \\ \mathbf{O}_{n_y,n_x} & \mathbf{I}_{n_y} \end{bmatrix} = \begin{bmatrix} M^{x,t} \\ M^{y,t} \end{bmatrix}, \quad (2.5)$$

le jeu de paramètres équivalent associé s'écrit, à partir de (2.1),

$$\begin{aligned}
 [\mathbf{F}^M]^{x,x} &= \mathbf{M}^{x,x} \mathbf{F}^{x,x} [\mathbf{M}^{x,x}]^{-1} + \mathbf{M}^{x,y} \mathbf{F}^{y,x} [\mathbf{M}^{x,x}]^{-1}, \\
 [\mathbf{F}^M]^{x,y} &= \mathbf{M}^{x,x} \mathbf{F}^{x,y} + \mathbf{M}^{x,y} \mathbf{F}^{y,y}, \\
 &\quad - \mathbf{M}^{x,x} \mathbf{F}^{x,x} [\mathbf{M}^{x,x}]^{-1} \mathbf{M}^{x,y} - \mathbf{M}^{x,y} \mathbf{F}^{y,x} [\mathbf{M}^{x,x}]^{-1} \mathbf{M}^{x,y}, \\
 [\mathbf{F}^M]^{y,x} &= \mathbf{F}^{y,x} [\mathbf{M}^{x,x}]^{-1}, \\
 [\mathbf{F}^M]^{y,y} &= \mathbf{F}^{y,y} - \mathbf{F}^{y,x} [\mathbf{M}^{x,x}]^{-1} \mathbf{M}^{x,y}, \\
 [\mathbf{Q}^M]^{x,x} &= \mathbf{M}^{x,x} \mathbf{Q}^{x,x} [\mathbf{M}^{x,x}]^T + \mathbf{M}^{x,x} \mathbf{Q}^{x,y} [\mathbf{M}^{x,y}]^T \\
 &\quad + \mathbf{M}^{x,y} [\mathbf{Q}^{x,y}]^T [\mathbf{M}^{x,x}]^T + \mathbf{M}^{x,y} \mathbf{Q}^{y,y} [\mathbf{M}^{x,y}]^T, \\
 [\mathbf{Q}^M]^{x,y} &= \mathbf{M}^{x,x} \mathbf{Q}^{x,y} + \mathbf{M}^{x,y} \mathbf{Q}^{y,y}, \\
 [\mathbf{Q}^M]^{y,y} &= \mathbf{Q}^{y,y}.
 \end{aligned} \tag{2.6}$$

Remarque: Il est impossible d'imposer tout type de contraintes puisque le nombre de variables à déterminer est inférieur au nombre d'équations. Par exemple, dans le cadre scalaire, il y a 7 équations pour 2 inconnues. De plus, la matrice $\mathbf{Q}^{y,y}$ est indépendante du filtre équivalent.

Nous avons réussi à caractériser la matrice \mathbf{M} dans trois cas particuliers. Toutes les preuves relatives à ces contraintes sont reportées dans la partie annexe du manuscrit section 1.

2.1 Cas α : contraintes sur la matrice de covariance du bruit

Il est possible d'imposer la covariance du bruit de modèle $[\mathbf{Q}^M]^{x,x}$, notée alors $[\mathbf{Q}_0^M]^{x,x}$, ainsi que sa décorrélation avec le bruit de mesure, *c.a.d* que \mathbf{Q}^M a la forme suivante

$$\mathbf{Q}^M = \begin{bmatrix} [\mathbf{Q}_0^M]^{x,x} & \mathbf{O}_{n_x, n_y} \\ \mathbf{O}_{n_y, n_x} & \mathbf{Q}^{y,y} \end{bmatrix}. \tag{2.7}$$

On montre que les sous-matrices de la matrice \mathbf{M} qui permettent de transformer \mathbf{Q} en \mathbf{Q}^M sont données par (Annexes : démonstration 1)

$$\begin{aligned}
 \mathbf{M}^{x,y} &= -\mathbf{M}^{x,x} \mathbf{Q}^{x,y} [\mathbf{Q}^{y,y}]^{-1}, \\
 \mathbf{M}^{x,x} &= \left[[\mathbf{Q}_0^M]^{x,x} \right]^{\frac{T}{2}} [\mathbf{Q}_*^{x,x}]^{-\frac{T}{2}},
 \end{aligned} \tag{2.8}$$

avec

$$\mathbf{Q}_*^{x,x} = \mathbf{Q}^{x,x} - \mathbf{Q}^{x,y} [\mathbf{Q}^{y,y}]^{-1} [\mathbf{Q}^{x,y}]^T.$$

Remarques:

- Le jeu de paramètres trouvé est unique.

- Dans le cadre de l’algorithme EM robuste, $[Q_*^{x,x}]^{\frac{1}{2}}$ peut être calculé directement à partir de $Q^{\frac{1}{2}}$:

$$\begin{bmatrix} [Q_*^{x,x}]^{\frac{1}{2}} & * \\ O & * \end{bmatrix} = Z^* \begin{bmatrix} I & -Q^{x,y} [Q^{y,y}]^{-1} \\ O & I \end{bmatrix} Q^{\frac{1}{2}}.$$

2.2 Cas β : contraintes sur les matrices d’états

Il est possible, sous certaines conditions, d’imposer la forme des équations d’états, par le biais des matrices $[F^M]^{x,x}$ et $[F^M]^{x,y}$. Pour trouver la matrice M associée, nous devons résoudre le système linéaire (démonstration triviale) :

$$vec([O_{n_x, n_x}, [F^M]^{x,y}]) = K vec(M^{x,t}), \quad (2.9)$$

avec

$$K = [F^T \otimes I_{n_x}] - [I_{n_t} \otimes [F^M]^{x,x}], \quad (2.10)$$

où la notation $vec(A)$ représente le dé-plierement vectoriel de A et la notation $[A \otimes B]$ représente le produit de Kronecker de A et B .

Sous condition que la matrice K soit inversible, nous obtenons

$$vec(M^{x,t}) = K^{-1} vec([O_{n_x, n_x}, [F^M]^{x,y}]). \quad (2.11)$$

Remarque: Le jeu de paramètres trouvé est unique.

2.3 Cas γ : contraintes sur les matrices d’observations

Quand la dimension des états cachés n_x est supérieure ou égale à la dimension des observations n_y , il est possible d’imposer les valeurs des matrices d’observations $[F^M]^{y,x}$ et $[F^M]^{y,y}$, dénotées $[F_0^M]^{y,x}$ et $[F_0^M]^{y,y}$, avec seule condition que les matrices $[F_0^M]^{y,x}$ et $F^{y,x}$ soient de rang plein. La matrice M qui permet de transformer F en F^M est calculée en deux étapes (Annexe : démonstration 2) :

A. Le bloc $M^{x,x}$ de la matrice M est déterminé par l’algorithme suivant :

1. Nous effectuons les décompositions QR des matrices $[F^{y,x}]^T$ et $[[F_0^M]^{y,x}]^T$ (les matrices unitaires sont requises ici) :

$$\begin{aligned} [F^{y,x}]^T &= Q^o \begin{bmatrix} T^o \\ O_{n_x - n_y, n_y} \end{bmatrix}, \\ [[F_0^M]^{y,x}]^T &= Q^+ \begin{bmatrix} T^+ \\ O_{n_x - n_y, n_y} \end{bmatrix}. \end{aligned} \quad (2.12)$$

2. Ensuite, nous calculons

$$\begin{aligned} Q_M^o &= Q^o \begin{bmatrix} T^o & O_{n_y, n_x - n_y} \\ O_{n_x - n_y, n_y} & I_{n_x - n_y} \end{bmatrix}, \\ Q_M^+ &= Q^+ \begin{bmatrix} T^+ & O_{n_y, n_x - n_y} \\ O_{n_x - n_y, n_y} & I_{n_x - n_y} \end{bmatrix}. \end{aligned} \quad (2.13)$$

3. Nous obtenons

$$M^{x,x} = [Q_M^+]^{-T} [Q_M^o]^T. \quad (2.14)$$

B. Ensuite, le bloc $M^{x,y}$ se calcule selon

$$M^{x,y} = [Q_M^+]^{-T} \begin{bmatrix} F^{y,y} - [F_0^M]^{y,y} \\ O_{n_x-n_y, n_y} \end{bmatrix}. \quad (2.15)$$

Remarque: Le jeu de paramètres trouvé n'est unique que lorsque les dimensions des états cachés et des observations sont égales. Les blocs de la matrice de transformation M sont alors donnés par

$$M^{x,x} = [[F^M]^{y,x}]^{-1} F^{y,x}, \quad (2.16)$$

$$M^{x,y} = [[F^M]^{y,x}]^{-1} [F^{y,y} - [F^M]^{y,y}]. \quad (2.17)$$

Cas de la restauration de signaux dégradés par un bruit additif

Ce dernier type de contraintes devient particulièrement utile dans le cadre de la restauration de signaux dégradés par un bruit additif gaussien. En effet, sous ces hypothèses, nous pouvons noter l'observation comme

$$y_n = F^{y,z} z_n + \omega_{n+1}^y, \quad (2.18)$$

où z est le signal utile et ω^y le bruit additif. Si l'on construit l'état $x_n = [z_n^T, r_n^T]^T$ avec r un processus latent à valeur réelles, alors il est possible de construire un filtre de Kalman couple qui respecte cette hypothèse. En effet, il suffit d'imposer à la matrice F^M la forme suivante

$$F^M = \begin{bmatrix} F^{z,z} & F^{z,r} & F^{z,y} \\ F^{r,z} & F^{r,r} & F^{r,y} \\ F^{y,z} & O_{n_y, n_r} & O_{n_y, n_y} \end{bmatrix}. \quad (2.19)$$

Nous remarquons que la forme de F est compatible avec les contraintes de type γ et donc, permet de trouver la matrice M qui transforme F en F^M . Une autre propriété extrêmement importante est que le processus z ainsi trouvé est unique.

Démonstration:

- Les matrices M qui permettent de conserver la structure de F du type (2.19) sont de forme

$$M = \begin{bmatrix} I_{n_y} & O_{n_y, n_r} & O_{n_y, n_y} \\ M^{r,z} & M^{r,r} & M^{r,y} \\ O_{n_y, n_y} & O_{n_y, n_r} & I_{n_y} \end{bmatrix}.$$

- Les processus z et r équivalents sont donnés par

$$\begin{aligned} z_n^M &= z_n \\ r_n^M &= M^{r,z} z_n + M^{r,r} r_n + M^{r,y} y_{n-1} \end{aligned}$$

- Cela montre que le processus z est indépendant du processus latent r choisi.

Dans cette section, nous avons proposé 3 types de contraintes qui préservent la fonction de vraisemblance utilisée par l’algorithme EM. Dans la section suivante, nous proposons plusieurs séries d’expériences qui évaluent les filtres équivalents vis-à-vis du filtre de Kalman couple original.

3 Évaluations

Cette section est dédiée aux tests des différents filtres de Kalman couples équivalents. Nous nous sommes placés dans les trois cas décrits dans la section 2. Nous avons sélectionné l’erreur quadratique moyenne (EQM) comme mesure de la qualité de la restauration du signal. Les résultats sont les moyennes de 1000 expériences indépendantes. Les tailles de signaux varient de 50 échantillons à 10 000 échantillons. Pour éviter des confusions, nous nommerons filtre de Kalman couple libre (FKCL), le filtre sans contrainte sur ses paramètres, tel qu’il a été défini dans le chapitre 1.

3.1 Filtre équivalent de type α

Pour tester l’apport des filtres de Kalman couples équivalents de type α par rapport au filtre de Kalman couple original dans l’estimation des paramètres, nous avons simulé des signaux suivant le modèle du filtre de Kalman couple, avec une décorrélation entre bruits de mesure et de modèle (les paramètres sont donnés dans la ligne 1 du tableau 2.1). Les signaux ont ensuite été restaurés par

0. le filtre de Kalman couple optimal (FKCO - filtre supervisé avec les paramètres de simulation),
1. un filtre de Kalman couple libre (FKCL),
2. un filtre de Kalman couple équivalent de type α (FKCE), avec $[Q_0^M]^{x,x} = 0.5$.

Les paramètres d’initialisation des filtres non-supervisés sont donnés dans la ligne 2 du tableau 2.1. Les estimations moyennes par les deux filtres non-supervisés sont reportées dans les lignes 3 et 4. Un exemple de signal généré est représenté sur la figure 2.1a et ses restaurations par les différents filtres sur les figures 2.1b, 2.1c et 2.1d. La vraisemblance moyenne pour les signaux de 1000 points est reportée sur la figure 2.2. Finalement, l’EQM des états lissés est représentée sur la figure 2.3 pour des signaux variant de 50 échantillons à 10 000 échantillons.

Il convient d’analyser séparément les données expérimentales

- Concernant la restauration du signal par les différents filtres, nous remarquons que le FKCE dégrade le signal, que le FKCL l’amplifie légèrement et que les performances de ces deux filtres restent éloignées de celles du filtre optimal.

TABLE 2.1: Paramètres de l'expérience de la section 3.1 ; Ligne 1 : paramètres de simulation ; Ligne 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 3 et 4 : estimation moyenne des paramètres par les différents filtres pour un signal de 1 000 échantillons (100 itérations EM).

	\hat{t}_0	Q_0	F	Q
θ	$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.5 & -0.5 \\ 1 & -0.5 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$
$\theta^{(0)}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
$\theta_{FKCL}^{(100)}$	$\begin{bmatrix} 1.433 \\ -0.983 \end{bmatrix}$	$\begin{bmatrix} 0.013 & 0.025 \\ 0.025 & 0.141 \end{bmatrix}$	$\begin{bmatrix} 0.110 & -0.296 \\ 0.904 & -0.119 \end{bmatrix}$	$\begin{bmatrix} 0.452 & -0.181 \\ -0.181 & 0.699 \end{bmatrix}$
$\theta_{FKCE}^{(100)}$	$\begin{bmatrix} 0.247 \\ -0.983 \end{bmatrix}$	$\begin{bmatrix} 0.047 & 0.011 \\ 0.011 & 0.141 \end{bmatrix}$	$\begin{bmatrix} 0.341 & -0.071 \\ 0.104 & -0.349 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0 \\ 0 & 0.699 \end{bmatrix}$

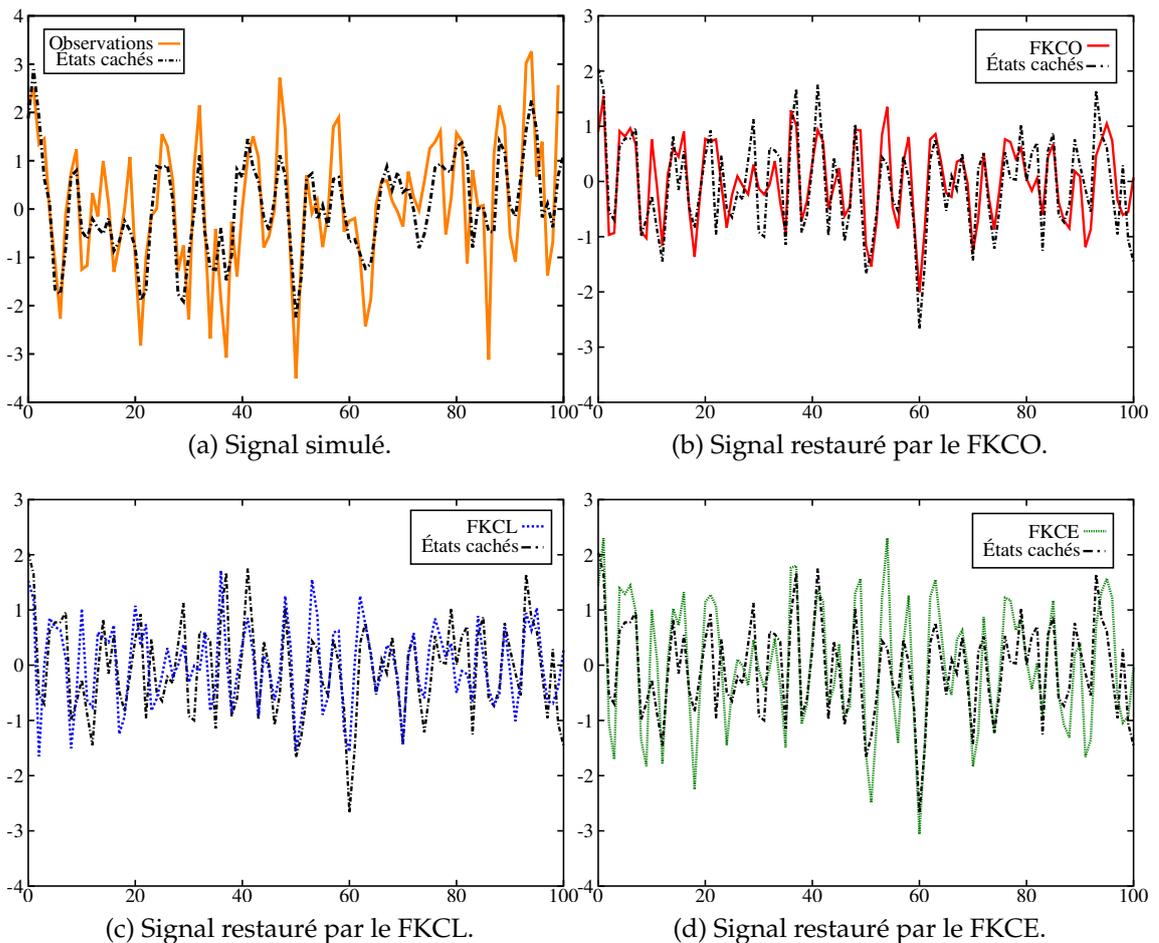


FIGURE 2.1: Exemple d'un signal de 100 échantillons simulé et de ses restaurations par les différents filtres pour l'expérience de la section 3.1.

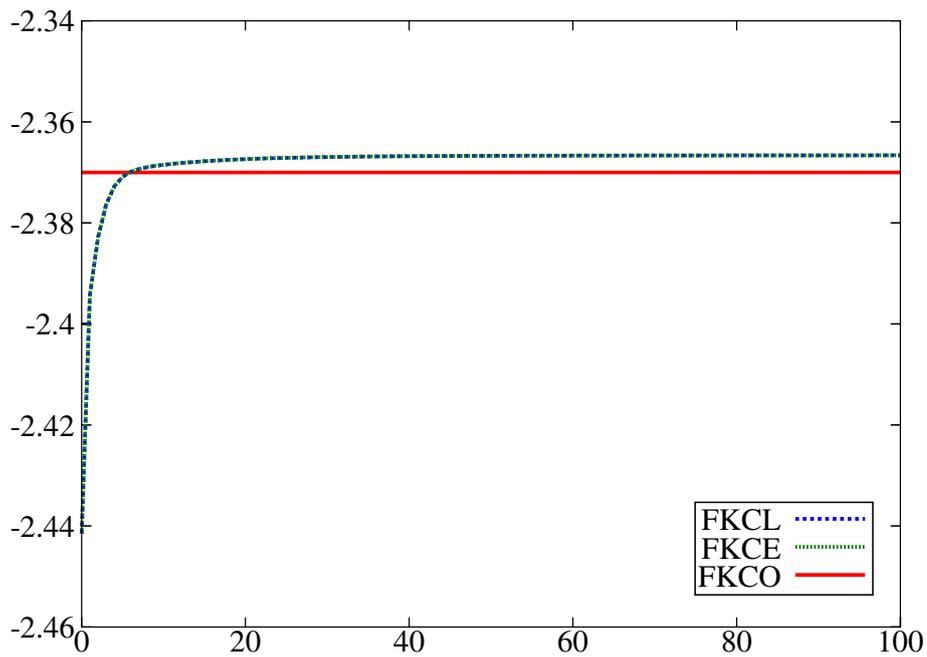


FIGURE 2.2: Vraisemblances moyennes des différents filtres pour un signal de 1 000 échantillons (section 3.1).

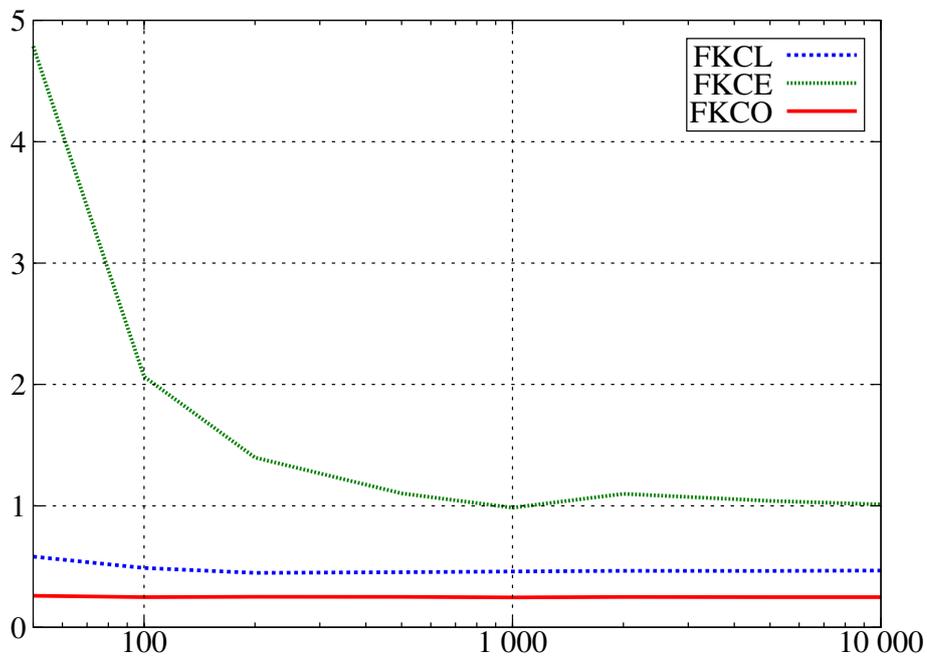


FIGURE 2.3: Évolution de l'EQM en fonction du nombre d'échantillons (section 3.1).

- Concernant la vraisemblance, les courbes du FKCE et du FKCL sont confondues, ce qui est conforme aux hypothèses des filtres équivalents.
- Concernant l'estimation des paramètres, le FKCE donne une meilleure estimation de Q que le FKCL. Par contre, le FKCE donne une très mauvaise estimation de F . On note aussi que l'estimation de F par le filtre libre reste éloignée des paramètres de simulation.
- Concernant les courbes d'EQM, nous remarquons que le FKCE donne une EQM plus élevée que le FKCL et que les performances des deux filtres non-supervisés restent assez éloignées de celles du filtre optimal, même pour un grand nombre d'échantillons.

Nous pouvons tirer 2 conclusions de cette expérience :

1. les filtres équivalents de type- α ne donnent pas les résultats escomptés sur des signaux de moins de 10 000 points et dégradent même le signal.
2. si nous voulons imposer une contrainte conforme à la matrice de covariance du bruit de modèle, ainsi que la décorrélation des bruits de mesure et de modèle, nous devons utiliser une autre méthode que les filtres équivalents.

3.2 Filtre équivalent de type β

Pour illustrer l'intérêt des filtres de Kalman couples équivalents de type β par rapport au filtre de Kalman couple original, nous avons simulé des signaux suivant le modèle du filtre de Kalman couple, dont les paramètres sont donnés dans la ligne 1 du tableau 2.2. Les signaux ont ensuite été restaurés par

0. le filtre de Kalman couple optimal (FKCO),
1. un filtre de Kalman couple libre (FKCL),
2. un filtre de Kalman couple équivalent de type β (FKCE) où $F^{x,t} = [0, -0.5]$.

Les paramètres d'initialisation des filtres non-supervisés sont donnés dans la ligne 2 du tableau 2.2. Les estimations moyennes par les deux filtres non-supervisés sont reportées dans les lignes 3 et 4. Un exemple de signal généré est représenté sur la figure 2.4a et ses restaurations par les différents filtres respectifs sur les figures 2.4b, 2.4c et 2.4d. Finalement, l'EQM des signaux lissés est représentée sur la figure 2.5 pour des signaux variant de 50 échantillons à 10 000 échantillons.

Les courbes nous amènent aux commentaires suivants :

- Concernant l'exemple de restauration, nous remarquons que la restauration du FKCL est très proche des observations et que les performances du FKCE sont proches de celles du filtre optimal.
- Concernant l'estimation moyenne des paramètres, le FKCE donne une meilleure estimation globale que le FKCL, qui est proche des paramètres de simulation. Néanmoins, la covariance de l'état initiale est très mal estimée par tous les filtres.
- Concernant les courbes d'EQM, l'EQM du FKCE est plus faible que celle du FKCL et tend vers celle du FKCO quand le nombre d'échantillons devient

TABLE 2.2: Paramètres de l'expérience de la section 3.2. Ligne 1 : paramètres de simulation ; Ligne 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 3 et 4 : estimation moyenne des paramètres par les différents filtres pour un signal de 1 000 échantillons (100 itérations EM).

	\hat{t}_0	Q_0	F	Q
θ	$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0.2 \\ 0.2 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & -0.5 \\ 1 & -0.5 \end{bmatrix}$	$\begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.5 \end{bmatrix}$
$\theta^{(0)}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & -0.5 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
$\theta_{FKCL}^{(100)}$	$\begin{bmatrix} 1.986 \\ -0.863 \end{bmatrix}$	$\begin{bmatrix} 0.010 & 0.002 \\ 0.002 & 0.030 \end{bmatrix}$	$\begin{bmatrix} -0.335 & -0.667 \\ 0.666 & -0.164 \end{bmatrix}$	$\begin{bmatrix} 0.579 & 0.070 \\ 0.070 & 0.456 \end{bmatrix}$
$\theta_{FKCE}^{(100)}$	$\begin{bmatrix} 1.096 \\ -0.863 \end{bmatrix}$	$\begin{bmatrix} 0.009 & 0.012 \\ 0.012 & 0.030 \end{bmatrix}$	$\begin{bmatrix} 0 & -0.5 \\ 0.998 & -0.499 \end{bmatrix}$	$\begin{bmatrix} 0.347 & 0.200 \\ 0.200 & 0.456 \end{bmatrix}$

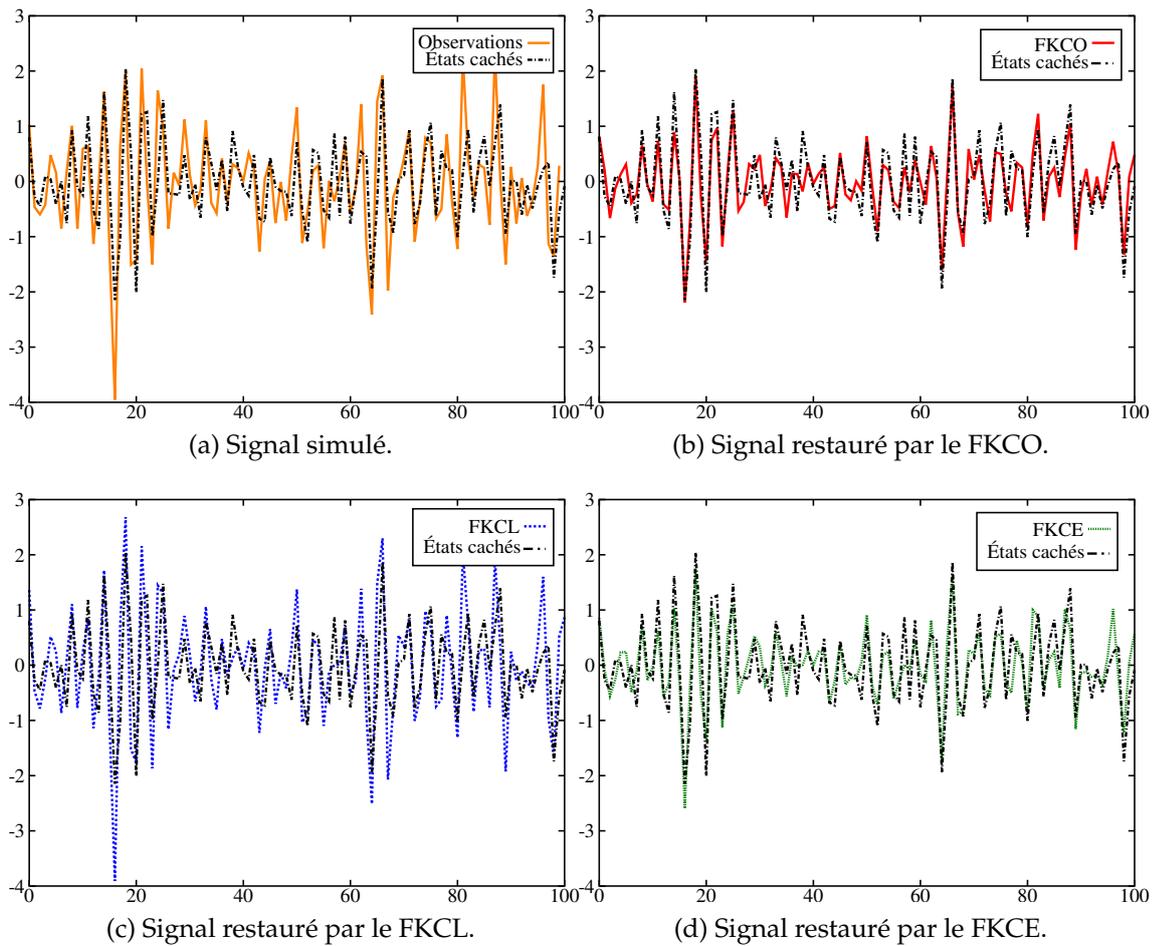


FIGURE 2.4: Exemple d'un signal de 100 échantillons simulé et de ses restaurations par les différents filtres (section 3.2).

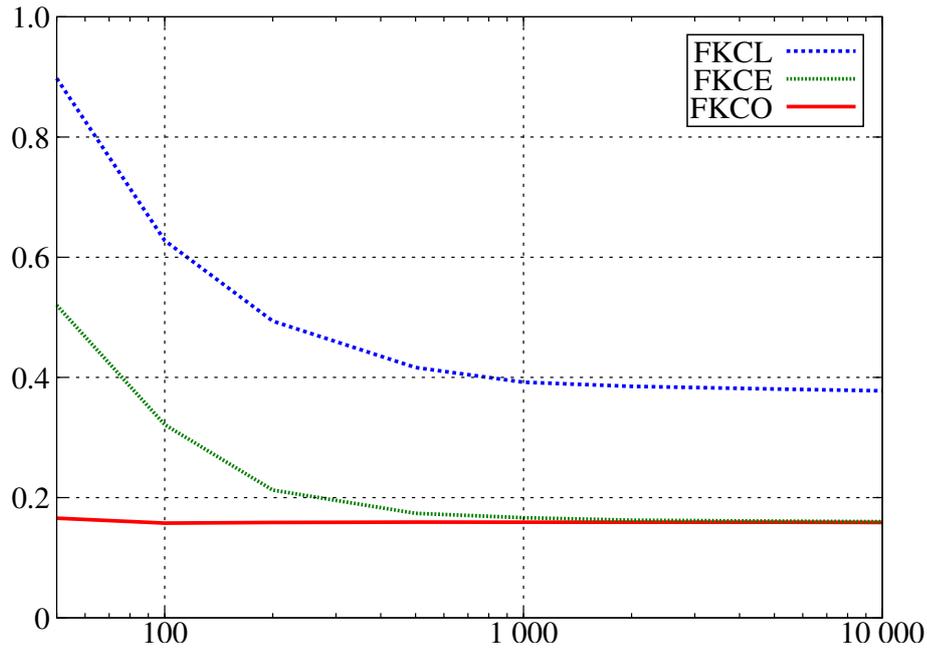


FIGURE 2.5: Évolution de l'EQM en fonction du nombre d'échantillons (section 3.2).

important.

Finalement, nous pouvons tirer de cette expérience deux conclusions :

1. le FKCE améliore significativement les estimations des paramètres et des états cachés par rapport au FKCL ;
2. comme attendu, les performances du FKCE se rapprochent des performances du filtre optimal quand le nombre d'échantillons devient important.

3.3 Filtre équivalent de type γ

Pour illustrer l'utilité des filtres de Kalman couple équivalents de type- γ par rapport au filtre de Kalman couple original, nous avons choisi de simuler des signaux suivant le modèle du filtre de Kalman couple, dont les paramètres sont répertoriés dans la ligne 1 du tableau 2.3. Nous avons ensuite restauré les signaux par

0. le filtre de Kalman couple optimal (FKCO),
1. un filtre de Kalman couple libre (FKCL),
2. un filtre de Kalman couple équivalent de type γ (FKCE) où $\mathbf{F}^{y,t} = [1, 0]$.

Les paramètres d'initialisation des filtres non-supervisés sont donnés dans la ligne 2 du tableau 2.3. Les estimations moyennes pour les signaux de 1000 échantillons par les deux filtres non-supervisés sont reportées dans les lignes 3 et 4. Un exemple de signal généré est représenté sur la figure 2.6a, ses restaurations par les différents filtres sur les figures respectives 2.6b, 2.6c et 2.6d. Finalement, l'EQM des signaux lissés est représentée sur la figure 2.7 pour des signaux variant de 50 échantillons à 10 000 échantillons.

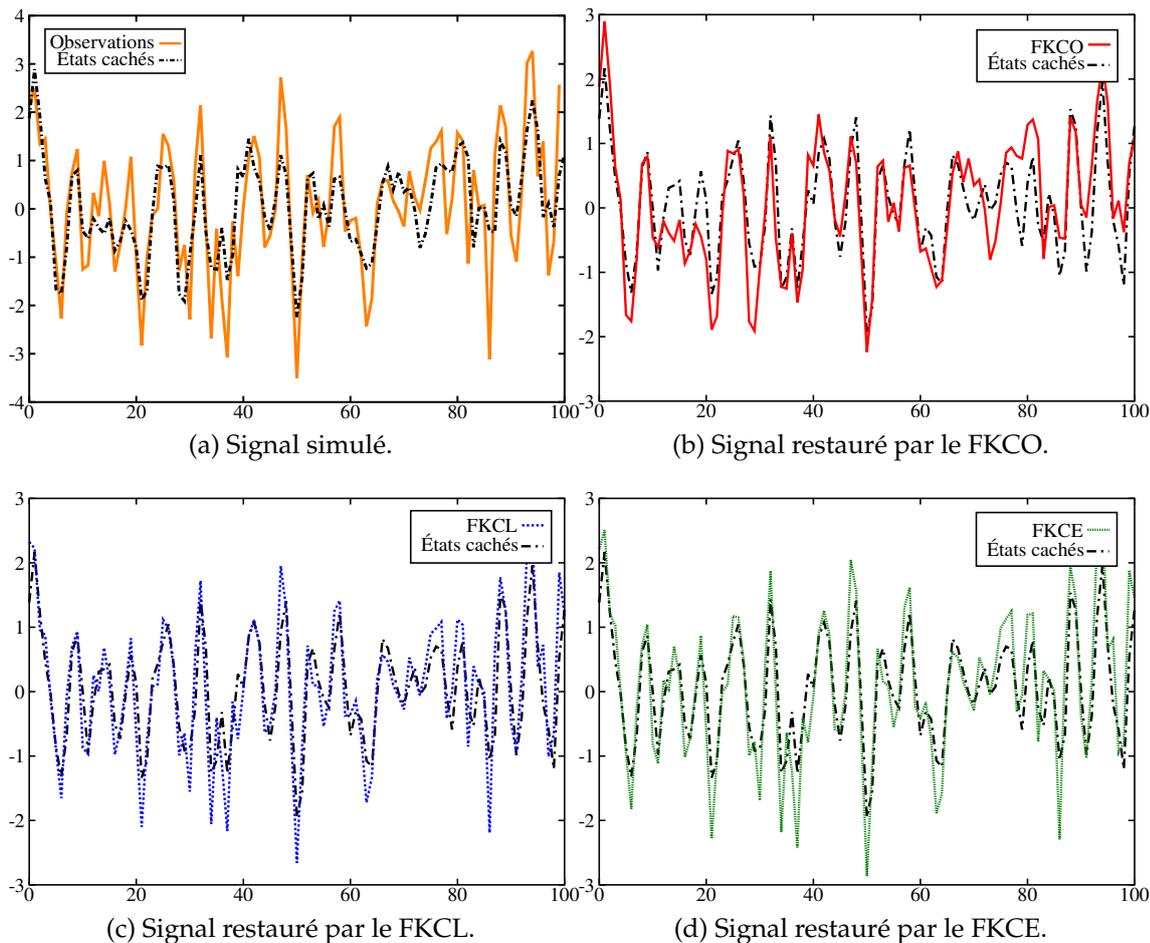


FIGURE 2.6: Exemple d'un signal de 100 échantillons simulé et de ses restaurations par les différents filtres (section 3.3).

Il convient d'analyser ces résultats séparément :

- Concernant la restauration, nous remarquons que les performances du FKCE et FKCL sont assez proches et restent éloignées de la restauration optimale.
- Concernant l'estimation moyenne des paramètres, le FKCE donne globalement une meilleure estimation que le FKCL, qui est proche des paramètres de simulation. Néanmoins, la covariance de l'état initial est très mal-estimée par tous les filtres non-supervisés.
- Concernant les courbes d'EQM, nous remarquons que les EQM du FKCE sont plus faibles que celles du FKCL. Quand la taille des signaux dépasse 500 points, les EQM du FKCE et du FKCO sont presque confondues.

Finalement, nous pouvons tirer de cette expérience deux conclusions :

1. L'utilisation d'un filtre équivalent améliore significativement les estimations des paramètres et des états cachés par rapport au filtre de Kalman couple original.
2. Comme attendu, les performances du filtre équivalent se rapprochent de celles du filtre optimal quand le nombre d'échantillons devient important.

TABLE 2.3: Paramètres de l'expérience de la section 3.3. Ligne 1 : paramètres de simulation ; Ligne 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 3 et 4 : estimation moyenne des paramètres par les différents filtres pour un signal de 1 000 échantillons (100 itérations EM).

	\hat{t}_0	Q_0	F	Q
θ	$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0.2 \\ 0.2 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.5 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 1 \end{bmatrix}$
$\theta^{(0)}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
$\theta_{FKCL}^{(100)}$	$\begin{bmatrix} 1.501 \\ -1.038 \end{bmatrix}$	$\begin{bmatrix} 0.009 & 0.004 \\ 0.004 & 0.059 \end{bmatrix}$	$\begin{bmatrix} 0.796 & -0.441 \\ 0.771 & 0.197 \end{bmatrix}$	$\begin{bmatrix} 0.492 & -0.085 \\ -0.085 & 0.908 \end{bmatrix}$
$\theta_{FKCE}^{(100)}$	$\begin{bmatrix} 0.971 \\ -1.038 \end{bmatrix}$	$\begin{bmatrix} 0.009 & 0.014 \\ 0.0144 & 0.059 \end{bmatrix}$	$\begin{bmatrix} 0.993 & -0.497 \\ \mathbf{1} & \mathbf{0} \end{bmatrix}$	$\begin{bmatrix} 0.308 & 0.117 \\ 0.117 & 0.908 \end{bmatrix}$

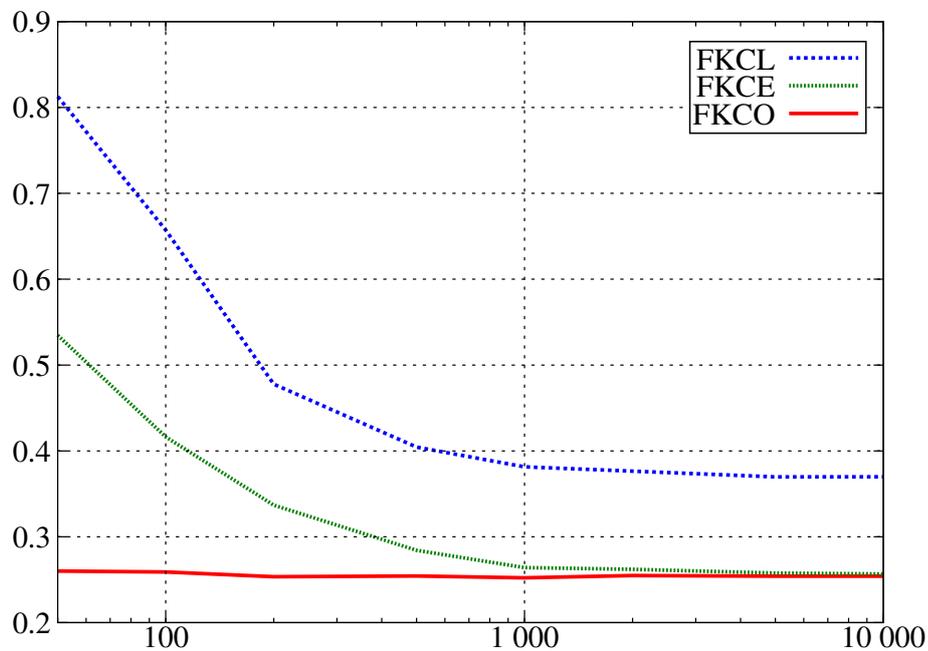


FIGURE 2.7: Évolution de l'EQM en fonction du nombre d'échantillons (section 3.3).

4 Conclusion

Dans ce chapitre, nous avons illustré le problème de l'identifiabilité du modèle du filtre de Kalman couple. Ce problème peut dégrader significativement les performances de l'algorithme EM original du filtre de Kalman couple puisqu'il est impossible de différencier deux jeux de paramètres équivalents selon le critère de la vraisemblance. Nous avons alors proposé de sélectionner un jeu de paramètres équivalents qui impose une forme particulière à F et à Q . Néanmoins, nous ne pouvons pas imposer n'importe quelle contrainte à ces matrices et trouver un filtre équivalent. Nous avons établis 3 cas où l'on peut toujours trouver un jeu de paramètres équivalent qui a une structure particulière. Les trois filtres équivalents ainsi trouvés ont été testés via des expériences numériques. Or, malheureusement, l'un de ces trois filtres qui impose une structure particulière à la matrice Q dégrade le signal comme le suggère la première expérience. On peut nuancer que les deux autres filtres équivalents qui imposent soit les matrices d'états $F^{x,x}$ et $F^{x,y}$ ou les matrices d'observations $F^{y,x}$ et $F^{y,y}$ donnent d'excellents résultats vis à vis du filtre de Kalman couple original, et que de plus, leurs performances tendent vers celles du filtre optimal quand le nombre d'échantillons devient assez élevé.

Dans le chapitre suivant, nous proposons d'incorporer des connaissances a priori sur les paramètres lors de l'étape de maximisation de l'algorithme EM, ce qui permet de traiter un plus grand nombre de configurations que les filtres équivalents.

Chapitre 3

Filtre de Kalman couple contraint

Ce chapitre poursuit l'étude de l'estimation des paramètres du modèle du filtre de Kalman couple par l'algorithme EM. Contrairement au chapitre précédent où la fonction de vraisemblance est préservée, nous étudions ici les contraintes possibles sur les matrices F et Q qui permettent d'inclure des connaissances partielles (provenant de la physique) que l'on a sur le modèle et qui modifie la fonction de vraisemblance.

À titre d'exemple, considérons le système composé de deux capteurs identiques, mais indépendants. Dans cette configuration, il est naturel de fixer la même matrice de covariance pour les bruits de mesure générés par les deux capteurs, soit

$$Q^{y,y} = \begin{bmatrix} Q^s & O \\ O & Q^s \end{bmatrix}, \quad (3.1)$$

où Q_1 est la covariance des bruits générés par les deux capteurs. Or, un tel modèle ne peut pas être en compte par les filtres équivalents du chapitre précédent car ces derniers préservent la matrice de covariance du bruit de mesure.

Dans ce chapitre, nous proposons d'étendre les travaux sur l'apprentissage partiel des paramètres du filtre de Kalman [20] au cas du filtre de Kalman couple. Ce problème ne doit pas être confondu avec le choix d'un jeu de paramètres équivalents puisqu'il modifie l'étape de maximisation de l'algorithme EM original. En effet, la fonction auxiliaire de vraisemblance (1.12) n'est plus maximisée par rapport à F et Q directement, mais en tenant compte des éventuelles contraintes.

Cela est inclu dans le problème de l'estimation partiel. Seulement peu de travaux sont dédiés à contraindre l'algorithme EM, *e.g.* [31,72] pour les modèles de mélange Gaussien, [59] pour les chaînes de Markov cachées, and [20] pour le filtre de Kalman. Dans le cadre du filtre de Kalman couple, le problème est particulièrement complexe, ce qui est aux multiples interactions entre les matrices F et Q . Un point clé est qu'il n'est pas possible d'imposer n'importe quel type de contraintes et de garder des formules de ré-estimation exactes des paramètres.

Les contraintes sur les matrices que nous considérons ne doivent pas être confondues avec les contraintes sur les états cachés (par les moyens d'égalités ou

inégalités algébriques), qui sont très étudiées dans la littérature [25, 67]. Il ne faut pas non-plus les confondre avec l'apprentissage supervisé partiel pour lequel un sous-ensemble des états cachés est connu et est réutilisé pour l'estimation des paramètres.

Ce chapitre est composé de trois sections. Nous présentons d'abord, l'ensemble des contraintes retenues qui imposent une structure particulière à F et Q , ainsi que les formules de ré-estimations des paramètres qui tiennent comptes de ces contraintes. Puis, dans la seconde section, nous décrivons une variante robuste des formules de ré-estimations, compatible avec l'étape E robuste, présentée dans le chapitre 1 section 2. Finalement, dans la dernière section, nous mettons en valeur le filtre contraint par une série d'expériences.

1 Expression des contraintes

Dans cette section, nous proposons un algorithme EM permettant d'introduire des contraintes sur les matrices F et Q tout en conservant des formules de ré-estimation exactes des paramètres du modèle du filtre de Kalman couple. Comme nous le verrons après, ces contraintes gardent l'étape d'espérance de l'algorithme EM original inchangée. Mais, l'étape de maximisation doit maintenant tenir compte des contraintes sur les paramètres. Toutes les contraintes que nous avons établies sont basées sur l'estimation séparée de blocs de lignes de la matrice de transition F . Or, pour conserver des expressions analytiques pour les formules de ré-estimation, nous devons imposer à Q d'être bloc-diagonale conjointement aux blocs de F . Par exemple, si nous souhaitons estimer séparément $F^{x,t}$ et $F^{y,t}$, alors, il faut imposer une décorrélation entre bruits de mesure et de modèle, soit $Q^{x,y} = O$. Dès lors, nous pourrions imposer des contraintes différentes et indépendantes sur les différents blocs de F et de Q . Pour la suite, nous ne tiendrons compte que de la partie de la fonction auxiliaire de vraisemblance qui dépend de F et Q soit $g_{\tilde{C}}(Q, F)$ (1.12). Notons dès à présent que les formules de ré-estimation de \hat{t}_0 et Q_0 restent identiques à celles du filtre de Kalman couple libre ; elles ne seront pas traitées dans ce chapitre. Toutes les démonstrations pour les formules de ré-estimations sont reportées dans la partie annexe du manuscrit section 2.

En conséquence, la section est donc organisée en 3 sous-sections : la première traite en détail de la décorrélation du bruit, le seconde des contraintes sur la matrice F et la dernière sur les contraintes possibles sur les blocs diagonaux de Q .

1.1 Décorrélation du bruit

Nous souhaitons pouvoir estimer chaque bloc de F indépendamment des autres blocs. Pour obtenir une solution analytique lors de la ré-estimation des paramètres, il faut que la matrice de covariance du bruit Q soit bloc-diagonale

conjointement aux blocs de F . Sous ces hypothèses, F et Q s'écrivent :

$$\begin{aligned} Q &= \sum_{i=1}^{n_{IN}} P_i^T Q_i P_i, \\ F &= \sum_{i=1}^{n_{IN}} P_i^T F_i, \end{aligned} \quad (3.2)$$

où les $\{P_i\}_{i=1:n_{IN}}$ sont les projecteurs des blocs diagonaux de Q ou des blocs de lignes de la matrice F . Nous rappelons que ces projecteurs sont mutuellement orthogonaux ($\forall i \neq j, P_i^T P_j = O_{n_t, n_t}$).

La fonction auxiliaire de vraisemblance (1.12) se décompose alors en n_{IN} fonctions indépendantes (Annexe : démonstration 3)

$$g_{\tilde{C}}(Q, F) = \sum_{i=1}^{n_{IN}} g_{\tilde{C}_i}(Q_i, F_i), \quad (3.3)$$

où

$$\tilde{C}_i = \begin{bmatrix} I_{n_t} & O_{n_t, n_i} \\ O_{n_t, n_t} & P_i \end{bmatrix} \tilde{C} \begin{bmatrix} I_{n_t} & O_{n_t, n_i} \\ O_{n_t, n_t} & P_i \end{bmatrix}^T, \quad (3.4)$$

qui peuvent être maximisées séparément.

1.2 Blocs de la matrice de transition

Nous cherchons à maximiser chaque sous-fonction auxiliaire de vraisemblance $g_{\tilde{C}_i}(Q_i, F_i)$ par rapport au bloc F_i . Toute sous-matrice F_i peut s'écrire

$$F_i = F_i^0 + F_i^b, \quad (3.5)$$

où F_i^0 est une matrice connue, et F_i^b , une matrice à estimer. Cela revient à maximiser $g_{\tilde{C}_i^b}(Q_i, F_i^b)$ par rapport à F_i^b (Annexe : démonstration 4), avec

$$\tilde{C}_i^b = \begin{bmatrix} I_{n_t} & O_{n_t, n_i} \\ -F_i^0 & I_{n_i} \end{bmatrix} \tilde{C}_i \begin{bmatrix} I_{n_t} & O_{n_t, n_i} \\ -F_i^0 & I_{n_i} \end{bmatrix}^T. \quad (3.6)$$

Il est possible de trouver des expressions analytiques pour l'estimation de \hat{F}_i^b pour quatre formes différentes de la sous-matrice F_i^b :

- (0) $F_i^b = O_{n_i, n_t}$ l'estimation est triviale.
- (1) $F_i^b = G_i$ où G_i est une matrice de taille $(n_i \times n_t)$ à estimer. Dans ce cas, la matrice \hat{F}_i^b qui maximise la sous-fonction auxiliaire de vraisemblance $g_{\tilde{C}_i^b}(Q_i, F_i^b)$ s'écrit (Annexe : démonstration 5 avec $M_i = I_{n_t}$)

$$\hat{F}_i^b = \tilde{C}_i^{b,[1,0]} \left[\tilde{C}_i^{b,[0,0]} \right]^{-1}, \quad (3.7)$$

où $\tilde{\mathbf{C}}_i^{b,[0,0]}$ et $\tilde{\mathbf{C}}_i^{b,[1,0]}$ correspondent aux sous-matrices de $\tilde{\mathbf{C}}_i^b$ de tailles respectives $(n_t \times n_t)$ et $(n_i \times n_t)$, et telles que

$$\tilde{\mathbf{C}}_i^b = \begin{bmatrix} \tilde{\mathbf{C}}_i^{b,[0,0]} & \tilde{\mathbf{C}}_i^{b,[0,1]} \\ \tilde{\mathbf{C}}_i^{b,[1,0]} & \tilde{\mathbf{C}}_i^{b,[1,1]} \end{bmatrix}.$$

Remarque: Cette formule est analogue à la formule de ré-estimation de \mathbf{F} (1.15) de l'algorithme EM original.

- (2) $\mathbf{F}_i^b = \mathbf{G}_i \mathbf{M}_i$ où \mathbf{G}_i est une matrice de taille $(n_i \times n_i^G)$ à estimer et \mathbf{M}_i , une matrice de rang plein et de taille $(n_i^G \times n_t)$, c.a.d $\text{rang}(\mathbf{M}_i) = n_i^G$. Dans cette configuration, la matrice $\hat{\mathbf{F}}_i^b$ qui maximise la sous-fonction auxiliaire de vraisemblance $g_{\tilde{\mathbf{C}}_i^b}(\mathbf{Q}_i, \mathbf{F}_i^b)$ est (Annexe : démonstration 5)

$$\hat{\mathbf{F}}_i^b = \tilde{\mathbf{C}}_i^{b,[1,0]} \mathbf{M}_i^T \left[\mathbf{M}_i \tilde{\mathbf{C}}_i^{b,[0,0]} \mathbf{M}_i^T \right]^{-1} \mathbf{M}_i. \quad (3.8)$$

Exemples:

- L'estimation de tous les paramètres du filtre de Kalman couple est obtenue en posant

$$\mathbf{G}_1 = \begin{bmatrix} \mathbf{F}^{x,x} & \mathbf{F}^{x,y} \\ \mathbf{F}^{y,x} & \mathbf{F}^{y,y} \end{bmatrix}, \mathbf{M}_1 = \mathbf{I}_{n_t} \text{ et } \mathbf{F}_1^0 = \mathbf{O}_{n_t, n_t}.$$

Dans ce cas, l'équation (3.2) s'écrit $\mathbf{F} = \mathbf{F}_1$ et $\tilde{\mathbf{C}}_i^b = \tilde{\mathbf{C}}_i$. Donc, l'équation (3.8) est la même que (1.15).

- Le filtre de Kalman original est un autre exemple de filtre de Kalman couple contraint, où

$$\mathbf{G}_1 = \begin{bmatrix} \mathbf{F}^{x,x} \\ \mathbf{F}^{y,x} \end{bmatrix}, \mathbf{M}_1 = [\mathbf{I}_{n_x}, \mathbf{O}_{n_x, n_t - n_x}] \text{ et } \mathbf{F}_1^0 = \mathbf{O}_{n_t, n_t}.$$

Dans ce cas, l'équation (3.8) est, aux notations prés, la même que celle présentée dans [65, chapitre 6, page 343].

- (3) $\mathbf{F}_i^b = \sum_{m=1}^{n_i^\lambda} \lambda_i^m \mathbf{U}_i^m$ où $\{\mathbf{U}_i^m\}_{m=1:n_i^\lambda}$ sont des matrices connues et indépendantes, et $\lambda_i = (\lambda_i^m)_{m=1:n_i^\lambda}$ est un vecteur de paramètres à estimer. De plus, la matrice \mathbf{Q}_i définie dans (3.2), doit être proportionnelle à une matrice symétrique et définie positive \mathbf{Q}_i^0 . Dans ce cas, le vecteur $\hat{\lambda}_i$ qui maximise la sous-fonction auxiliaire de vraisemblance $g_{\tilde{\mathbf{C}}_i^b}(\mathbf{Q}_i, \mathbf{F}_i^b)$ est (Annexe : démonstration 6)

$$\hat{\lambda}_i = \mathbf{A}_i^{-1} \mathbf{b}_i, \quad (3.9)$$

où

$$\mathbf{A}_i = \sum_{k=1}^{n_i^\lambda} \sum_{l=1}^{n_i^\lambda} \text{tr} \left[[\mathbf{Q}_i^0]^{-1} \mathbf{U}_i^k \tilde{\mathbf{C}}_i^{b,[0,0]} \mathbf{U}_i^l \right] \mathbf{E}_{k,l}, \quad (3.10)$$

avec $\mathbf{E}_{k,l}$, le (k, l) ^{ème} élément canonique de l'espace des matrices de taille $(n_i^\lambda \times n_i^\lambda)$, et où

$$\mathbf{b}_i = \sum_{k=1}^{n_i^\lambda} \text{tr} \left[[\mathbf{Q}_i^0]^{-1} \mathbf{U}_i^k \tilde{\mathbf{C}}_i^{b,[0,1]} \right] \mathbf{e}_k, \quad (3.11)$$

avec \mathbf{e}_k , le k ^{ème} élément canonique de l'espace des vecteurs de taille n_i^λ .

Exemple: Considérons le filtre de Kalman couple suivant qui permet d'estimer le gain entre observations et états cachés, quand la matrice de transition est partiellement connue :

– Pour \mathbf{F} :

$$\mathbf{F}^{x,x} = 0.5, \mathbf{F}^{x,y} = 0.5, \mathbf{F}^{y,y} = 0 \text{ et } \mathbf{F}^{y,x} = \lambda,$$

où λ doit être estimé. Dès lors,

$$\mathbf{F} = [\mathbf{F}_1^T, \mathbf{F}_2^T]^T$$

$$\text{avec } \mathbf{F}_1 = [\mathbf{F}^{x,x}, \mathbf{F}^{x,y}] = [0.5, 0.5] \text{ et } \mathbf{F}_2 = [\mathbf{F}^{y,x}, \mathbf{F}^{y,y}] = [\lambda, 0].$$

– Pour \mathbf{Q} :

$$\mathbf{Q}^{x,x} = \mathbf{Q}_1 = \sigma^2, \mathbf{Q}^{y,y} = \mathbf{Q}_2 = \gamma^2 \text{ et } \mathbf{Q}^{x,y} = 0,$$

où γ et σ doivent être estimés.

Dans cet exemple, les blocs de \mathbf{F} et \mathbf{Q} ont comme projecteurs

$$\mathbf{P}_1 = [1, 0] \text{ et } \mathbf{P}_2 = [0, 1].$$

Le bloc \mathbf{F}_1 est de type-(0), et le bloc \mathbf{F}_2 est de type-(3) avec

$$n_2^\lambda = 1, \mathbf{F}_2^0 = [0, 0], \text{ et } \mathbf{U}_2^1 = [1, 0].$$

Dès lors, l'estimation de \mathbf{F} s'écrit

$$\hat{\mathbf{F}} = \sum_{i=1}^{n_{IN}} \mathbf{P}_i^T \hat{\mathbf{F}}_i. \quad (3.12)$$

Nous supposons maintenant que \mathbf{F} est estimée, donc nous écrivons dans la suite $\hat{\mathbf{F}}$ à la place de \mathbf{F} .

1.3 Blocs de la matrice de covariance

Nous rappelons que la matrice de bruit \mathbf{Q} doit être bloc-diagonale pour permettre l'estimation par blocs indépendants de la matrice de transition \mathbf{F} . Nous supposons alors que chaque bloc \mathbf{Q}_i ($i = 1 : n_{IN}$) est unitairement équivalent à une matrice bloc-diagonale \mathbf{Q}_i^M c.a.d $\mathbf{Q}_i = \mathbf{M}_i \mathbf{Q}_i^M \mathbf{M}_i^T$, avec \mathbf{M}_i une matrice unitaire connue. Dès lors, il existe une relation d'équivalence unitaire entre

la matrice de bruit \mathbf{Q} et la matrice bloc-diagonale \mathbf{Q}^M telle que $\mathbf{Q} = \mathbf{M}\mathbf{Q}^M\mathbf{M}^T$, avec

$$\mathbf{M} = \sum_{i=1}^{n_{IN}} \mathbf{M}_i \mathbf{P}_i^T.$$

La fonction auxiliaire de vraisemblance peut s'exprimer en fonction de \mathbf{Q}^M et $\hat{\mathbf{F}}^M$ car $g_{\tilde{\mathbf{C}}}(\mathbf{Q}, \hat{\mathbf{F}}) = g_{\tilde{\mathbf{C}}^M}(\mathbf{Q}^M, \hat{\mathbf{F}}^M)$ avec

$$\begin{aligned} \hat{\mathbf{F}}^M &= \mathbf{M}^T \hat{\mathbf{F}}, \\ \tilde{\mathbf{C}}^M &= \begin{bmatrix} \mathbf{I}_{n_t} & \mathbf{O}_{n_t, n_t} \\ \mathbf{O}_{n_t, n_t} & \mathbf{M}^T \end{bmatrix} \tilde{\mathbf{C}} \begin{bmatrix} \mathbf{I}_{n_t} & \mathbf{O}_{n_t, n_t} \\ \mathbf{O}_{n_t, n_t} & \mathbf{M} \end{bmatrix}. \end{aligned} \quad (3.13)$$

De plus, \mathbf{Q}^M s'écrit

$$\mathbf{Q}^M = \sum_{i=1}^{n_{QB}} [\mathbf{P}_i^M]^T \mathbf{Q}_i^M \mathbf{P}_i^M, \quad (3.14)$$

où $\{\mathbf{P}_i^M\}_{i=1:n_{QB}}$ sont les projecteurs des blocs diagonaux de \mathbf{Q}^M de rang respectif n_i^M . Dès lors, la fonction auxiliaire se décompose en n_{QB} sous-fonctions

$$g_{\tilde{\mathbf{C}}}(\mathbf{Q}, \hat{\mathbf{F}}) = \sum_{i=1}^{n_{QB}} g_{\tilde{\mathbf{C}}_i^M}(\mathbf{Q}_i^M, \hat{\mathbf{F}}_i^M), \quad (3.15)$$

où

$$\begin{aligned} \hat{\mathbf{F}}_i^M &= \mathbf{P}_i^M \hat{\mathbf{F}}^M, \\ \tilde{\mathbf{C}}_i^M &= \begin{bmatrix} \mathbf{I}_{n_t} & \mathbf{O}_{n_t, n_t} \\ \mathbf{O}_{n_i^M, n_t} & \mathbf{P}_i^M \end{bmatrix} \tilde{\mathbf{C}}^M \begin{bmatrix} \mathbf{I}_{n_t} & \mathbf{O}_{n_t, n_t} \\ \mathbf{O}_{n_i^M, n_t} & \mathbf{P}_i^M \end{bmatrix}^T. \end{aligned} \quad (3.16)$$

Les blocs sont supposés être indépendants, donc chaque sous-fonction peut être maximisée individuellement. Des expressions analytiques ont été trouvées pour quatre formes de la matrice \mathbf{Q}_i^M :

- (O) $\mathbf{Q}_i^M = \mathbf{Q}_i^0$, aucune estimation n'est requise.
- (A) \mathbf{Q}_i^M doit être estimée. Dans ce cas, la matrice de covariance $\hat{\mathbf{Q}}_i^M$ qui maximise la fonction auxiliaire est donnée par (Annexe : démonstration 7)

$$\hat{\mathbf{Q}}_i^M = \frac{1}{N+1} \left[\begin{bmatrix} [\hat{\mathbf{F}}_i^M]^T \\ -\mathbf{I}_{n_i^M} \end{bmatrix}^T \tilde{\mathbf{C}}_i^M \begin{bmatrix} [\hat{\mathbf{F}}_i^M]^T \\ -\mathbf{I}_{n_i^M} \end{bmatrix} \right]. \quad (3.17)$$

- (B) $\mathbf{Q}_i^M = \lambda_i \mathbf{Q}_i^0$, où λ_i est un scalaire qui doit être estimé et \mathbf{Q}_i^0 , une matrice connue. Dans ce cas, le scalaire $\hat{\lambda}_i$ qui maximise la fonction auxiliaire de vraisemblance est donné par (Annexe : démonstration 8)

$$\hat{\lambda}_i = \frac{1}{[N+1]n_i^M} \text{tr} \left[[\mathbf{Q}_i^0]^{-1} \left[\begin{bmatrix} [\hat{\mathbf{F}}_i^M]^T \\ -\mathbf{I}_{n_i^M} \end{bmatrix}^T \tilde{\mathbf{C}}_i^M \begin{bmatrix} [\hat{\mathbf{F}}_i^M]^T \\ -\mathbf{I}_{n_i^M} \end{bmatrix} \right] \right]. \quad (3.18)$$

(C) $\mathbf{Q}_i^M = \sum_{j=1}^{n_i^B} [\mathbf{P}_i^j]^T \mathbf{M}_i^j \mathbf{R}_i^M [\mathbf{M}_i^j]^T \mathbf{P}_i^j$, où $\{\mathbf{M}_i^j\}_{j=1:n_i^B}$ sont des matrices connues inversibles, $\{\mathbf{P}_i^j\}_{j=1:n_i^B}$ des projecteurs orthogonaux connus, et \mathbf{R}_i^M une matrice symétrique et définie positive à estimer. La matrice $\hat{\mathbf{R}}_i^M$ qui maximise la fonction auxiliaire de vraisemblance est donnée par (Annexe : démonstration 9)

$$\hat{\mathbf{R}}_i^M = \frac{1}{[N+1]n_i^B} \sum_{j=1}^{n_i^B} [\mathbf{M}_i^j]^{-1} \mathbf{P}_i^j \begin{bmatrix} [\hat{\mathbf{F}}_i^M]^T \\ -\mathbf{I}_{n_i^M} \end{bmatrix}^T \tilde{\mathbf{C}}_i^M \begin{bmatrix} [\hat{\mathbf{F}}_i^M]^T \\ -\mathbf{I}_{n_i^M} \end{bmatrix} [\mathbf{P}_i^j]^T [\mathbf{M}_i^j]^{-T}. \quad (3.19)$$

Exemple: Considérons un système où deux capteurs identiques mesurent une grandeur physique. Il est légitime d'imposer la même matrice de covariance du bruit de mesure \mathbf{Q}^s pour les deux capteurs. Donc, dans ce cas, la matrice de covariance du bruit de mesure est de type-(C) :

$$\begin{aligned} \mathbf{Q}^{x,x} &= \mathbf{Q}_1^M \\ \mathbf{Q}^{y,y} &= \mathbf{Q}_2^M = \begin{bmatrix} \mathbf{Q}^s & \mathbf{O} \\ \mathbf{O} & \mathbf{Q}^s \end{bmatrix} \end{aligned}$$

avec

$$\mathbf{P}_1^1 = [\mathbf{I}, \mathbf{O}], \mathbf{P}_2^2 = [\mathbf{O}, \mathbf{I}], \mathbf{M}_2^1 = \mathbf{M}_2^2 = \mathbf{I} \text{ et } \mathbf{R}_2 = \mathbf{Q}^s.$$

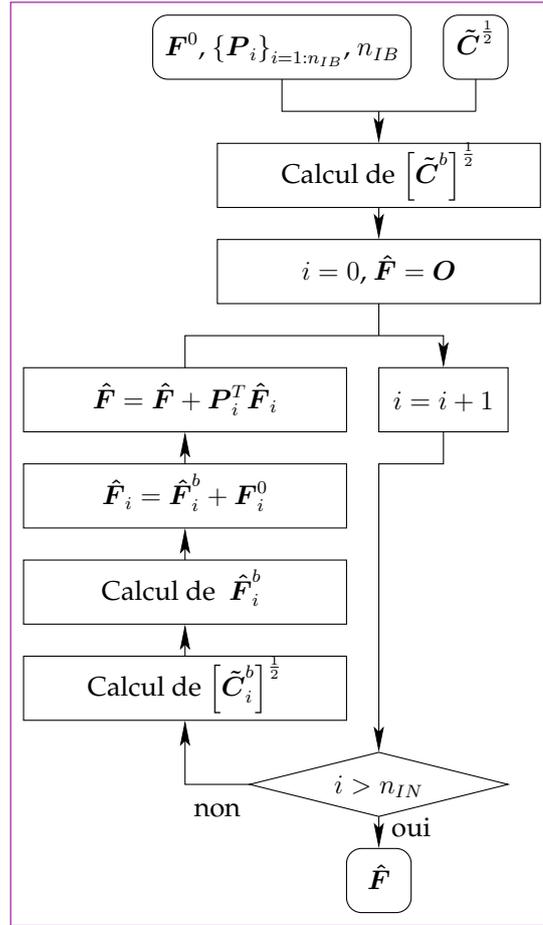
Nous venons de présenter un algorithme EM pour l'estimation non-supervisée des paramètres qui prend en compte certaines contraintes sur \mathbf{F} et sur \mathbf{Q} . En dépit du fait que ces contraintes semblent restrictives, cet algorithme reste très général comme nous serons amenés à l'illustrer dans la suite.

Remarque: En particulier, ces contraintes permettent de dégrader l'algorithme EM du filtre de Kalman couple pour le cas du filtre de Kalman, comme nous le détaillons dans la partie annexe section 3.

Cependant l'implémentation en l'état de cet algorithme conduit à de fortes instabilités numériques, comme celles constatées sur l'algorithme EM du filtre de Kalman couple présenté dans le chapitre 1. Nous présentons donc maintenant une version robuste de cet algorithme.

2 Ré-estimation robuste

A l'instar du chapitre 1, l'instabilité numérique vient du fait que les matrices de covariance ou corrélation censées être symétriques et positives peuvent perdre leurs propriétés après un certain nombre d'itérations de l'algorithme EM. Nous avons donc proposé de propager les matrices de covariances ou corrélation sous la forme $\mathbf{C} = \mathbf{T}^T \mathbf{T}$, où \mathbf{T} est une matrice triangulaire supérieure, pour qu'elles ne puissent pas perdre l'une ou l'autre de ces deux propriétés. Notre méthode reprend l'étape E présentée dans le chapitre 1.


 FIGURE 3.1: Estimation robuste et partielle de F .

2.1 Calcul de la matrice de transition

F peut se calculer à partir de $\tilde{C}^{\frac{1}{2}}$ (1.13). Notre méthode, schématisée dans la figure 3.1, se décompose en plusieurs étapes :

- (1) Une racine carrée de la matrice

$$\tilde{C}^b = \begin{bmatrix} \mathbf{I}_{n_t} & \mathbf{O}_{n_t, n_t} \\ -\mathbf{F}^0 & \mathbf{I}_{n_t} \end{bmatrix} \tilde{C} \begin{bmatrix} \mathbf{I}_{n_t} & \mathbf{O}_{n_t, n_t} \\ -\mathbf{F}^0 & \mathbf{I}_{n_t} \end{bmatrix}^T \quad (3.20)$$

est calculée à partir de la décomposition QR suivante :

$$\left[\tilde{C}^b \right]^{\frac{1}{2}} = \mathbf{Z}^* \tilde{C}^{\frac{1}{2}} \begin{bmatrix} \mathbf{I}_{n_t} & \mathbf{O}_{n_t, n_t} \\ -\mathbf{F}^0 & \mathbf{I}_{n_t} \end{bmatrix}^T. \quad (3.21)$$

- (2) Chaque bloc de F est estimé en utilisant l'algorithme suivant :

- (A) Une racine carrée de la matrice \tilde{C}_i^b (3.6) est calculée à partir de la décomposition QR suivante

$$\left[\tilde{C}_i^b \right]^{\frac{1}{2}} = \mathbf{Z}^* \left[\tilde{C}^b \right]^{\frac{1}{2}} \begin{bmatrix} \mathbf{I}_{n_t} & \mathbf{O}_{n_t, n_i} \\ \mathbf{O}_{n_t, n_t} & \mathbf{P}_i^T \end{bmatrix}. \quad (3.22)$$

- (B) \hat{F}_i^b est ensuite calculée selon le type de bloc de F_i :
- Pour un bloc de type (1),

$$\hat{F}_i^b = \left[\tilde{D}_i^{b,[0,1]} \right]^T \left[\tilde{D}_i^{b,[0,0]} \right]^{-1}, \quad (3.23)$$

où $\tilde{D}_i^{b,[0,0]}$ et $\tilde{D}_i^{b,[0,1]}$ correspondent aux sous-matrices de $\left[\tilde{C}_i^b \right]^{\frac{1}{2}}$ de tailles respectives $(n_t \times n_t)$ and $(n_t \times n_i)$, selon

$$\left[\tilde{C}_i^b \right]^{\frac{1}{2}} = \begin{bmatrix} \tilde{D}_i^{b,[0,0]} & \tilde{D}_i^{b,[0,1]} \\ \mathbf{O}_{n_i, n_t} & \tilde{D}_i^{b,[1,1]} \end{bmatrix}.$$

- Pour un bloc de type (2),
 - a. Une racine carrée de la matrice

$$\tilde{D}_i^M = \begin{bmatrix} M_i & \mathbf{O}_{n_i^G, n_t} \\ \mathbf{O}_{n_t, n_i} & \mathbf{I}_{n_t, n_t} \end{bmatrix} \tilde{C}_i^b \begin{bmatrix} M_i & \mathbf{O}_{n_i^G, n_t} \\ \mathbf{O}_{n_t, n_i} & \mathbf{I}_{n_t, n_t} \end{bmatrix}^T$$

est calculée grâce à la décomposition QR suivante

$$\left[\tilde{D}_i^M \right]^{\frac{1}{2}} = \mathbf{Z}^* \left[\tilde{C}_i^b \right]^{\frac{1}{2}} \begin{bmatrix} M_i & \mathbf{O}_{n_i^G, n_t} \\ \mathbf{O}_{n_t, n_i} & \mathbf{I}_{n_t, n_t} \end{bmatrix}^T.$$

- b. Ce qui permet d'écrire

$$\hat{F}_i^b = \left[\tilde{E}_i^{M,[0,1]} \right]^T \left[\tilde{E}_i^{M,[0,0]} \right]^{-1} M_i, \quad (3.24)$$

où $\tilde{E}_i^{M,[0,0]}$ et $\tilde{E}_i^{M,[0,1]}$ correspondent aux sous-matrices de $\left[\tilde{D}_i^M \right]^{\frac{1}{2}}$ de tailles respectives $(n_i^G \times n_i^G)$ and $(n_i^G \times n_t)$, selon

$$\left[\tilde{D}_i^M \right]^{\frac{1}{2}} = \begin{bmatrix} \tilde{E}_i^{M,[0,0]} & \tilde{E}_i^{M,[0,1]} \\ \mathbf{O} & \tilde{E}_i^{M,[1,1]} \end{bmatrix}.$$

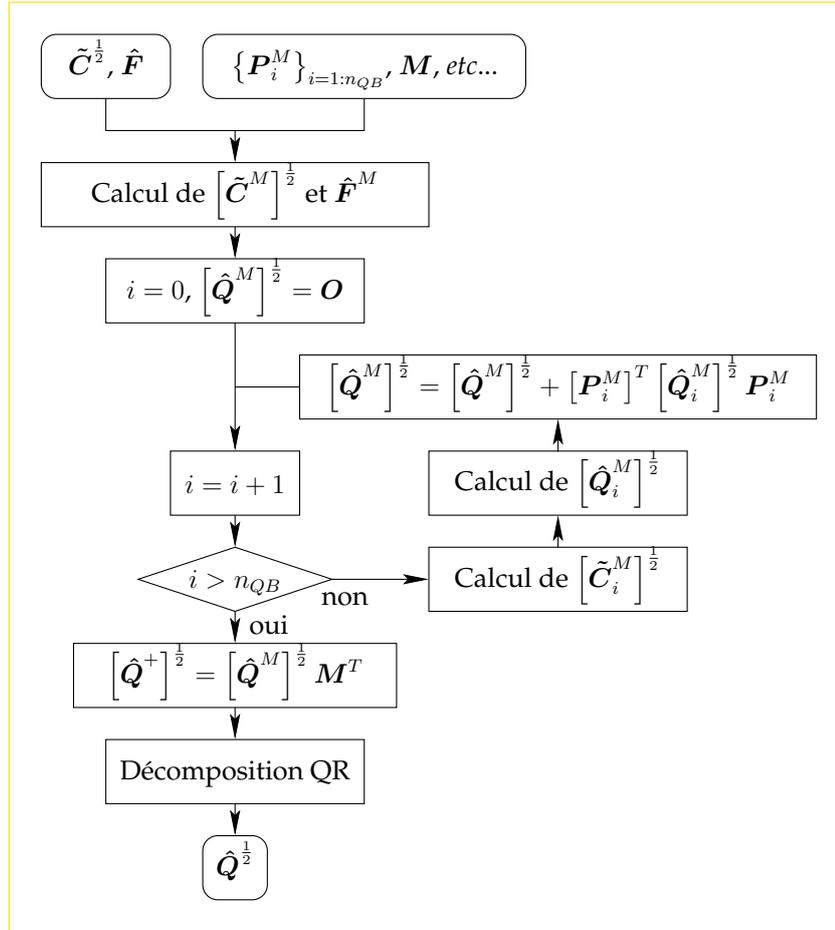
- Pour un bloc de type (3), l'équation (3.9) est ré-utilisée pour l'estimation.

(C) Nous obtenons \hat{F}_i selon $\hat{F}_i = \hat{F}_i^b + \hat{F}_i^0$.

(3) Finalement, \hat{F} est reconstruite à partir des \hat{F}_i , en utilisant l'équation (3.12). À ce stade, nous disposons d'un algorithme robuste pour estimer F selon toutes les contraintes exprimées dans la section précédente.

2.2 Calcul de la matrice de covariance du bruit

Afin de garantir la positivité et la symétrie de Q dans notre algorithme, nous proposons d'estimer $Q^{\frac{1}{2}}$ à la place de celle-ci. Le calcul de $Q^{\frac{1}{2}}$, cf. figure 3.2, est réalisé à partir de $\tilde{C}^{\frac{1}{2}}$ et se décompose en trois étapes :


 FIGURE 3.2: Estimation robuste et partielle de Q .

- (1) Une racine carrée de \tilde{C}^M (3.13) est donnée par la décomposition QR :

$$\left[\tilde{C}^M\right]^{\frac{1}{2}} = \mathbf{Z}^* \tilde{C}^{\frac{1}{2}} \begin{bmatrix} \mathbf{I}_{n_t} & \mathbf{O}_{n_t, n_t} \\ \mathbf{O}_{n_t, n_t} & \mathbf{M} \end{bmatrix}. \quad (3.25)$$

- (2) Chaque bloc Q_i^M est estimé en utilisant l'algorithme suivant :

- (a) Une racine carrée de \tilde{C}_i^M (3.16) est donnée par la décomposition QR

$$\left[\tilde{C}_i^M\right]^{\frac{1}{2}} = \mathbf{Z}^* \left[\tilde{C}^M\right]^{\frac{1}{2}} \begin{bmatrix} \mathbf{I}_{n_t} & \mathbf{O}_{n_t, n_i^M} \\ \mathbf{O}_{n_t, n_t} & \mathbf{P}_i^M \end{bmatrix}^T. \quad (3.26)$$

- (b) $\left[\hat{Q}_i^M\right]^{\frac{1}{2}}$ est ensuite calculée selon le type de bloc

- Pour un bloc de type (O), $\left[\hat{Q}_i^M\right]^{\frac{1}{2}} = [Q_i^0]^{\frac{1}{2}}$.
- Pour un bloc de type (A), $\left[\hat{Q}_i^M\right]^{\frac{1}{2}}$ est calculée grâce à la décomposition QR suivante

$$\left[\hat{Q}_i^M\right]^{\frac{1}{2}} = \frac{1}{\sqrt{N+1}} \mathbf{Z}^* \left[\tilde{C}_i^M\right]^{\frac{1}{2}} \left[\hat{F}_i^M, -\mathbf{I}_{n_i}\right]^T. \quad (3.27)$$

- Pour un bloc de type (B), $[\hat{Q}_i^M]^{\frac{1}{2}}$ est calculée en remplaçant λ_i et Q_i^0 respectivement par $\lambda_i^{\frac{1}{2}}$ et $[Q_i^0]^{\frac{1}{2}}$ dans l'équation (3.18).
- Pour un bloc de type (C), $[\hat{Q}_i^M]^{\frac{1}{2}}$ est calculée grâce à la méthode suivante
 - Une racine carrée de la matrice R_i^M (3.19) est calculée grâce à l'algorithme basé sur les décompositions QR suivantes :

$$[R_i^M]^{\frac{1}{2}} = \frac{1}{[N+1]n_i^B} [W_i^{n_i^B}]^{\frac{1}{2}},$$

où

$$W_i^k = \sum_{j=1}^k [M_i^j]^{-T} P_i^j \begin{bmatrix} [\hat{F}_i^M]^T \\ -I_{n_i^M} \end{bmatrix}^T \tilde{C}_i^M \begin{bmatrix} [\hat{F}_i^M]^T \\ -I_{n_i^M} \end{bmatrix} [P_i^j]^T [M_i^j]^{-1}. \quad (3.28)$$

Le calcul de $[W_i^k]^{\frac{1}{2}}$ se conduit de la manière suivante :

- (a) Pour $k = 0$, $[W_i^0]^{\frac{1}{2}} = O_{n_i^B, n_i^B}$.
- (b) Pour $k = 1$ à n_i^B , $[W_i^k]^{\frac{1}{2}}$ est calculé à partir de la décomposition QR suivante :

$$\begin{bmatrix} [W_i^k]^{\frac{1}{2}} \\ O_{n_i^B, n_i^B} \end{bmatrix} = Z^* \begin{bmatrix} [W_i^{k-1}]^{\frac{1}{2}} \\ \tilde{C}_i^M \begin{bmatrix} [\hat{F}_i^M]^T \\ -I_{n_i^M} \end{bmatrix} [P_i^j]^T [M_i^j]^{-1} \end{bmatrix}. \quad (3.29)$$

- Puis, $[\hat{Q}_i^M]^{\frac{1}{2}}$ est obtenu par la décomposition QR

$$[\hat{Q}_i^M]^{\frac{1}{2}} = Z^* \sum_{j=1}^{n_i^B} [P_i^j]^T [R_i^M]^{\frac{1}{2}} [M_i^j]^T P_i^j. \quad (3.30)$$

- (3) Dès lors, nous obtenons une racine carrée de \hat{Q}^M grâce à l'expression :

$$[\hat{Q}^M]^{\frac{1}{2}} = \sum_{i=1}^{n_{QB}} [P_i^M]^T [\hat{Q}_i^M]^{\frac{1}{2}} P_i^M. \quad (3.31)$$

- (4) Finalement, une racine carrée de \hat{Q} est obtenue par la décomposition QR

$$\hat{Q}^{\frac{1}{2}} = Z^* \underbrace{[\hat{Q}^M]^{\frac{1}{2}}}_{[\hat{Q}_+^M]^{\frac{1}{2}}} M^T. \quad (3.32)$$

3 Évaluations

Cette section propose de comparer le filtre de Kalman couple contraint avec le filtre de Kalman couple libre et le filtre de Kalman couple équivalent. Nous proposons 4 séries d'expériences qui comparent les performances des filtres équivalents et contraints dans les conditions décrites dans la section 2 du chapitre 2 et des filtres de plus en plus contraints. Tous les résultats sont la moyenne de 1000 expériences pour un nombre d'échantillons variant de 50 à 10 000. Nous avons retenu comme critère de qualité de la restauration, l'erreur quadratique moyenne (EQM) entre les états simulés et les états restaurés.

3.1 Comparaison avec le filtre équivalent de type α

L'expérience 3.1 du chapitre 2 montre un comportement décevant des filtres équivalents de type α dans la mesure où la matrice F est très mal estimée. Or, les contraintes que nous avons proposées dans ce chapitre permettent de prendre en compte le même type d'hypothèses et une comparaison des deux filtres est intéressante.

Nous avons donc repris le même schéma expérimental que dans la section 3.1 du chapitre 2. Nous avons simulé des signaux suivant le modèle du filtre de Kalman couple, avec une décorrélation entre bruits de mesure et de modèle, dont les paramètres sont donnés dans la ligne 1 du tableau 3.1. Les signaux ont ensuite été restaurés par

0. le filtre de Kalman couple optimal (FKCO),
1. un filtre de Kalman couple libre (FKCL),
2. un filtre de Kalman couple équivalent de type α (FKCE) où $Q^{x,y} = 0$ et $Q^{x,x} = 0.5$.
3. un filtre de Kalman couple contraint (FKC contraint) où
 - nous avons imposé $Q^{x,y} = 0$ et donc estimé $F^{x,t}$ et $F^{y,t}$ séparément,
 - nous avons imposé une contrainte de type-O sur $Q^{x,x} = 0.5$
 - et $Q^{y,y}$ est laissé libre.

Les paramètres d'initialisation des filtres non-supervisés sont donnés dans la ligne 2 du tableau 3.1. Les estimations moyennes par les trois filtres non-supervisés sont reportées dans les lignes 3 à 5. Les courbes de vraisemblance sont représentées sur la figure 3.3. Un exemple de signal généré est reporté sur la figure 3.4a, ses restaurations par les différents filtres sur les figures 3.4b, 3.4c, 3.4d et 3.4e. Les EQM pour le FKCL, le FKCE, le FKC contraint et le FKCO sont reportées sur la figure 3.5.

Il convient d'analyser séparément ces résultats :

- Concernant l'exemple de restauration d'un signal de 100 échantillons, le FKCE ne restaure pas le signal, contrairement aux FKCL et FKC contraint dont les performances sont proches. On note aussi que leurs performances restent assez éloignées de celles du filtre optimal.

TABLE 3.1: Paramètres de l'expérience de la section 3.1. Ligne 1 : paramètres de simulation ; Ligne 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 3 à 5 : estimation moyenne des paramètres par les différents filtres pour un signal d'une taille de 1 000 échantillons.

	\hat{t}_0	Q_0	F	Q
θ	$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0.1 \\ 0.1 & 1.5 \end{bmatrix}$	$\begin{bmatrix} 0.5 & -0.2 \\ 1 & 0.5 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0 \\ 0 & 0.7 \end{bmatrix}$
$\theta^{(0)}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix}$
$\theta_{FKCL}^{(100)}$	$\begin{bmatrix} 0.754 \\ -0.788 \end{bmatrix}$	$\begin{bmatrix} 0.006 & -0.006 \\ -0.006 & 0.057 \end{bmatrix}$	$\begin{bmatrix} 0.763 & -0.253 \\ 1.060 & 0.242 \end{bmatrix}$	$\begin{bmatrix} 0.485 & 0.137 \\ 0.137 & 0.666 \end{bmatrix}$
$\theta_{FKCE}^{(100)}$	$\begin{bmatrix} 0.727 \\ -0.788 \end{bmatrix}$	$\begin{bmatrix} 0.012 & -0.010 \\ -0.010 & 0.057 \end{bmatrix}$	$\begin{bmatrix} 0.537 & -0.127 \\ 0.586 & 0.468 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0 \\ 0 & 0.666 \end{bmatrix}$
$\theta_{FKC\text{ constraint}}^{(100)}$	$\begin{bmatrix} 0.906 \\ -0.703 \end{bmatrix}$	$\begin{bmatrix} 0.011 & -0.006 \\ -0.006 & 0.053 \end{bmatrix}$	$\begin{bmatrix} 0.564 & -0.205 \\ 1.011 & 0.452 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0 \\ 0 & 0.595 \end{bmatrix}$

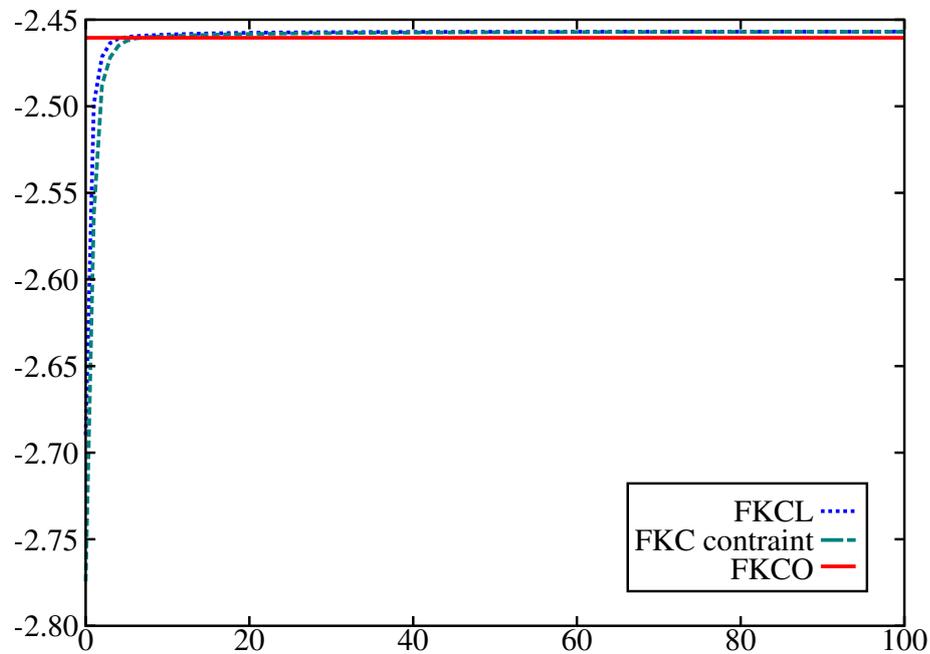
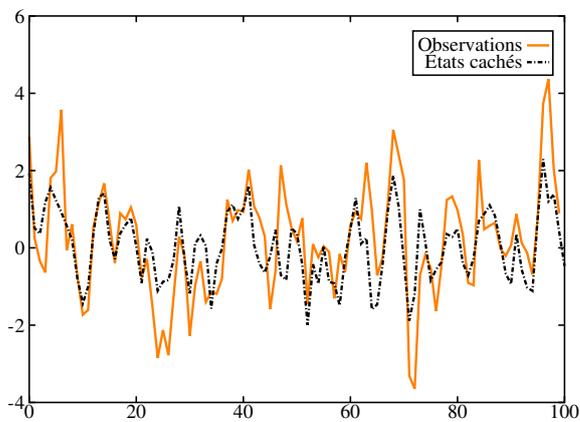
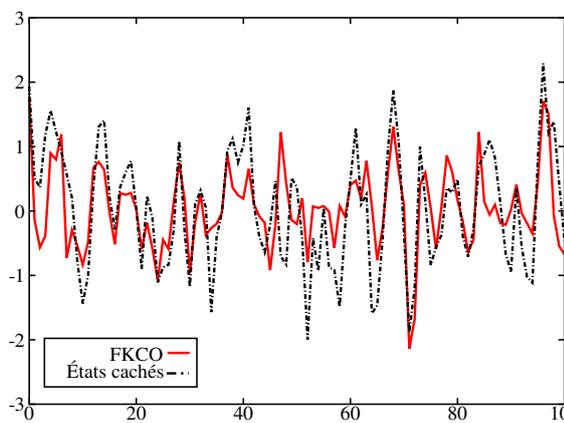


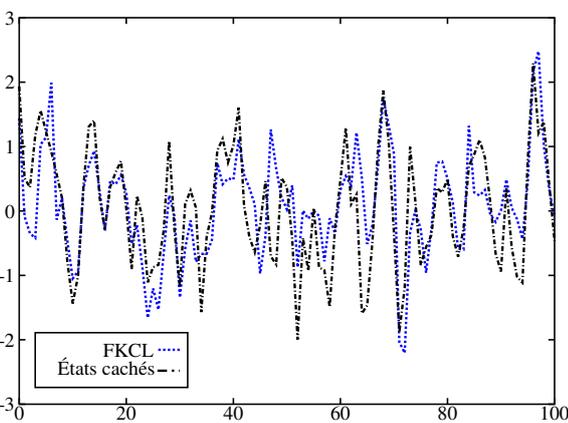
FIGURE 3.3: Vraisemblances moyennes des différents filtres pour un signal de 1 000 échantillons (section 3.1).



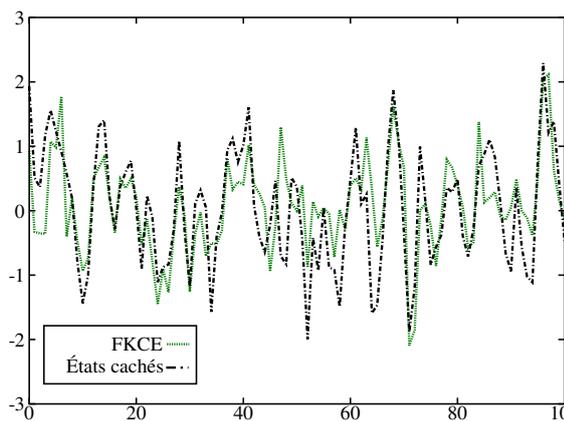
(a) Signal simulé.



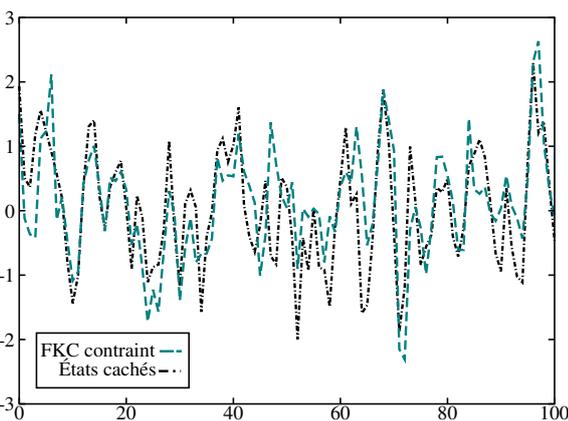
(b) Signal restauré par le FKCO.



(c) Signal restauré par le FKCL.



(d) Signal restauré par le FKCE.



(e) Signal restauré par le FKC contraint.

FIGURE 3.4: Exemple d'un signal de 100 échantillons simulé et de ses restaurations par les différents filtres (section 3.1).

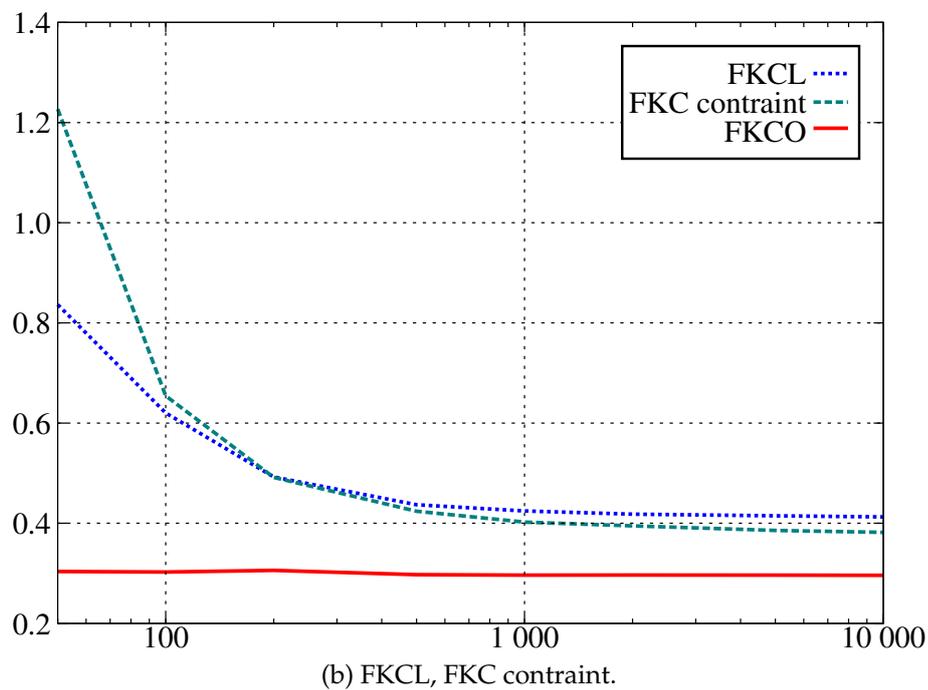
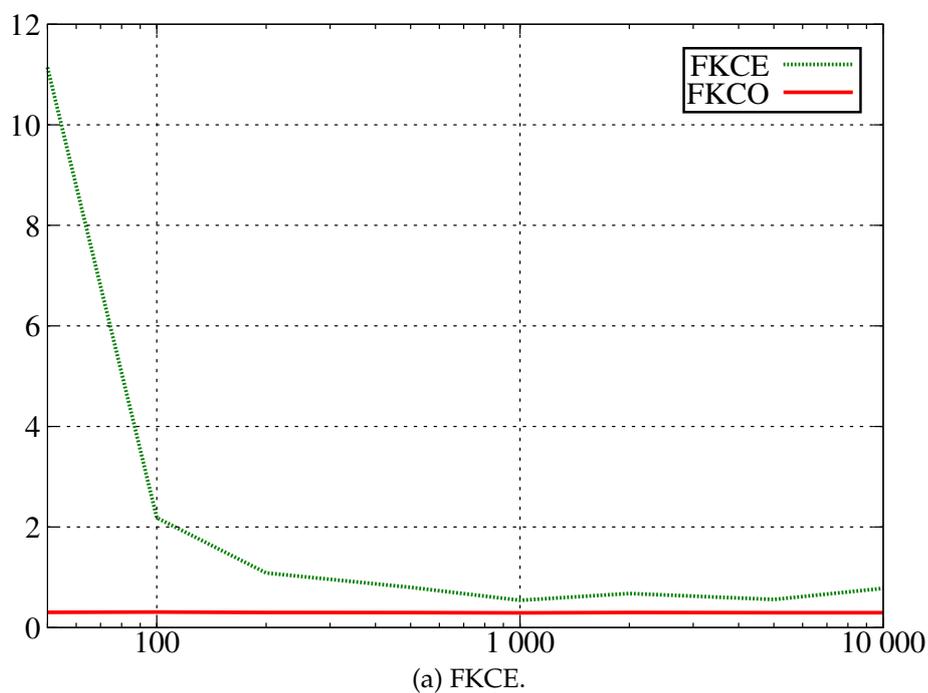


FIGURE 3.5: Evolution de l'EQM en fonction du nombre d'échantillons (section 3.1).

- Concernant la courbe de vraisemblance, on remarque que la vraisemblance dans le modèle contraint croît moins vite que celle du modèle libre. Ce résultat était attendu puisque l’algorithme contraint l’étape de maximisation de l’algorithme EM, ce qui n’est pas le cas des algorithmes libre ou équivalent.
- Concernant l’estimation des paramètres, le FKC contraint donne une meilleure estimation que le FKCL pour l’estimation des paramètres. Par contre, on peut nuancer ce bon résultat par le fait que l’estimation reste assez éloignée des paramètres de simulation. On peut également ajouter qu’on retrouve la mauvaise estimation du filtre équivalent.
- Concernant les courbes d’EQM, on remarque que le FKC contraint donne de meilleurs résultats que le FKCL quand le nombre d’échantillons dépasse 200 points. Cependant, ses performances restent éloignées de celles du filtre optimal.

Dans cette expérience, nous retenons une conclusion : les filtres contraints permettent d’estimer les paramètres et les états cachés quand $Q^{x,x}$ est connue et $Q^{x,y} = O$ lorsque le nombre d’échantillons est suffisant.

3.2 Comparaison avec le filtre équivalent de type β

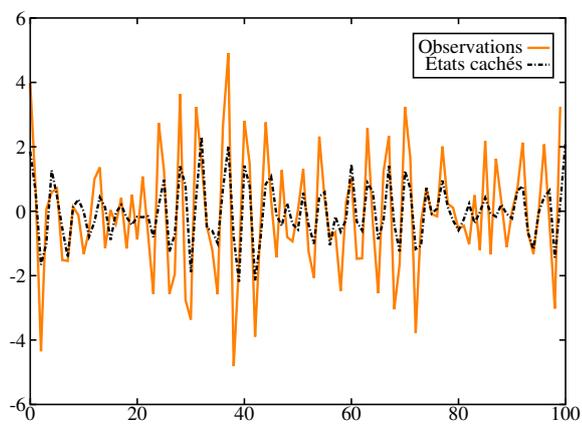
L’expérience de la section 3.2 du chapitre 2 montre un bon comportement des filtres équivalents de type β à la fois pour l’estimation des paramètres et des états cachés. Or, les filtres contraints permettent aussi d’imposer le même type de contraintes sur la matrice de transition F (avec la contrainte supplémentaire que les bruits de mesure et de modèle soient décorrélés). Dès lors, il est intéressant de comparer ces deux algorithmes d’apprentissage car l’un ne fait pas d’approximation et l’autre réduit l’espace de recherche des paramètres.

Nous avons donc simulé des signaux respectant les hypothèses du filtre équivalent de type- β dont les paramètres sont donnés dans la ligne 1 du tableau 3.2. Les signaux ont ensuite été restaurés par

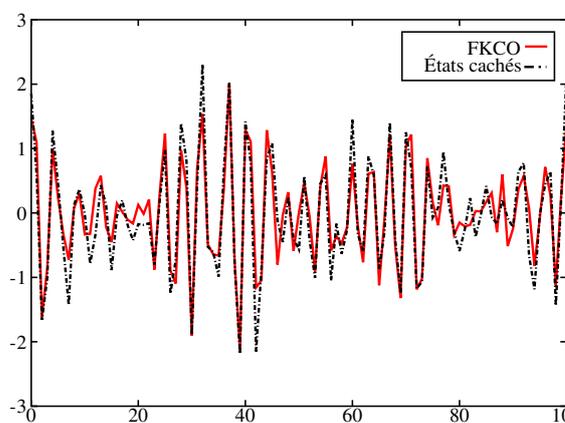
0. le filtre de Kalman couple optimal (FKCO),
1. un filtre de Kalman couple libre (FKCL),
2. un filtre de Kalman couple équivalent de type β (FKCE), où $F^{x,t} = [0.2, -0.4]$,
3. un filtre de Kalman couple contraint (FKC contraint) où
 - nous avons imposé $Q^{x,y} = 0$ et donc estimé $F^{x,t}$ et $F^{y,t}$ séparément,
 - nous avons imposé une contrainte de type-O sur $F^{x,t} = [0.2, -0.4]$
 - et $Q^{x,x}$, $Q^{y,y}$ et $F^{y,t}$ sont laissés libres.

Les paramètres d’initialisation des filtres non-supervisés sont donnés dans la ligne 2 du tableau 3.2. Les estimations moyennes par les trois filtres non-supervisés sont reportées dans les lignes 3 à 5. Un exemple de signal généré est représenté sur la figure 3.6a et ses restaurations respectives sur les figures 3.6b, 3.6c, 3.6d et 3.6e. L’EQM est représentée sur la figure 3.7.

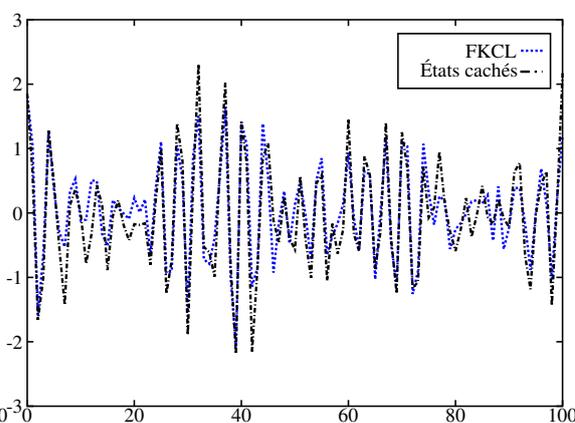
Il convient d’analyser les résultats séparément :



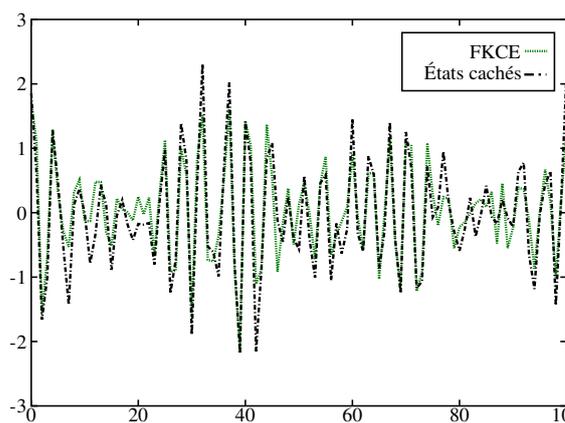
(a) Signal simulé.



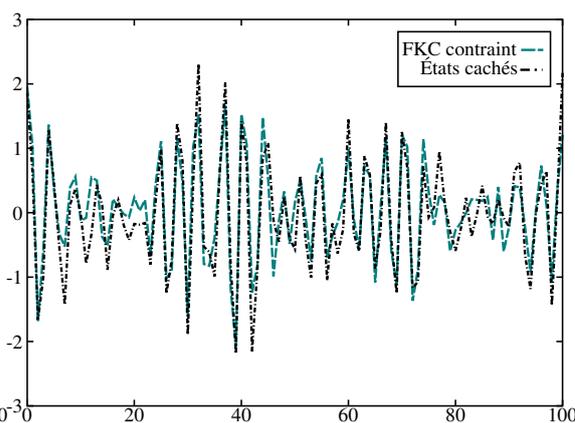
(b) Signal restauré par le FKCO.



(c) Signal restauré par le FKCL.



(d) Signal restauré par le FKCE.



(e) Signal restauré par le FKC contraint.

FIGURE 3.6: Exemple d'un signal de 100 échantillons simulé et de ses restaurations par les différents filtres (section 3.2).

TABLE 3.2: Paramètres de l'expérience de la section 3.2. Ligne 1 : paramètres de simulation ; Ligne 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 3 à 5 : estimation moyenne des paramètres par les différents filtres pour un signal d'une taille de 1 000 échantillons (100 itérations EM).

	\hat{t}_0	Q_0	F	Q
θ	$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0.2 \\ 0.2 & 1.0 \end{bmatrix}$	$\begin{bmatrix} 0.2 & -0.4 \\ 2 & -0.5 \end{bmatrix}$	$\begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 1.0 \end{bmatrix}$
$\theta^{(0)}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.2 & -0.4 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix}$
$\theta_{FKCL}^{(100)}$	$\begin{bmatrix} 1.222 \\ -1.042 \end{bmatrix}$	$\begin{bmatrix} 0.006 & 0.007 \\ 0.007 & 0.034 \end{bmatrix}$	$\begin{bmatrix} 0.052 & -0.406 \\ 1.768 & -0.352 \end{bmatrix}$	$\begin{bmatrix} 0.107 & 0.041 \\ 0.041 & 1.416 \end{bmatrix}$
$\theta_{FKCE}^{(100)}$	$\begin{bmatrix} 1.003 \\ -1.042 \end{bmatrix}$	$\begin{bmatrix} 0.006 & 0.010 \\ 0.010 & 0.034 \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.2} & \mathbf{-0.4} \\ 1.999 & -0.500 \end{bmatrix}$	$\begin{bmatrix} 0.097 & 0.141 \\ 0.141 & 1.416 \end{bmatrix}$
$\theta_{FKC\text{ contraint}}^{(100)}$	$\begin{bmatrix} 1.125 \\ -0.834 \end{bmatrix}$	$\begin{bmatrix} 0.006 & 0.009 \\ 0.009 & 0.035 \end{bmatrix}$	$\begin{bmatrix} \mathbf{0.2} & \mathbf{-0.4} \\ 1.946 & -0.445 \end{bmatrix}$	$\begin{bmatrix} 0.210 & \mathbf{0} \\ \mathbf{0} & 0.938 \end{bmatrix}$

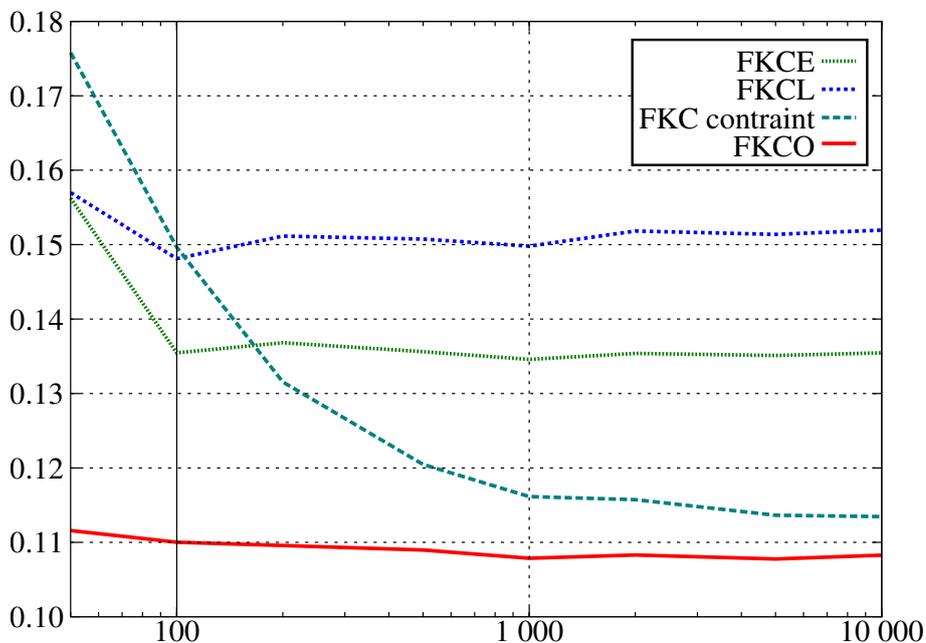


FIGURE 3.7: Évolution des EQM pour le lissage en fonction du nombre d'échantillons (section 3.2).

- Concernant l'exemple de restauration, tous les filtres donnent des résultats assez similaires. On remarque aussi que le filtre optimal procure une légère amélioration de l'estimation des états cachés.
- Concernant l'estimation des paramètres, les filtres équivalents et contraints donnent tous les deux une meilleure estimation de la matrice F que le filtre libre, et qui est proche des paramètres de simulation.
- Concernant les courbes d'EQM, on remarque que sur des petits signaux, le FKCE donne des meilleurs résultats que le FKC contraint. Par contre, sur de grands signaux, la tendance s'inverse. Il faut aussi noter que les résultats de ces deux filtres restent supérieures à celles du FKCL. On note finalement que le FKCE et le FKCL sont très loins de la convergence vers l'EQM optimale dans cette expérience.

Nous pouvons tirer de cette expérience 2 conclusions principales :

1. Les filtres équivalents ou contraints améliorent l'estimation des états cachés et des paramètres du modèle.
2. Il semblerait que les filtres équivalents soient légèrement plus sensibles aux maxima locaux de la fonction de vraisemblance que les filtres contraints.

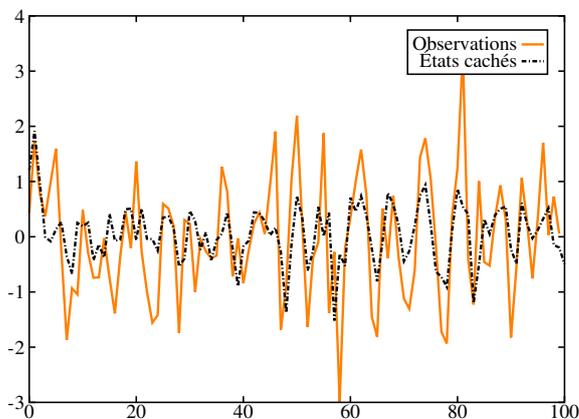
3.3 Comparaison avec le filtre équivalent de type γ

L'expérience de la section 3.3 du chapitre 2 montre un bon comportement des filtres équivalents de type γ dans l'estimation des paramètres et des états cachés. Or, les filtres contraints permettent d'imposer le même type de contraintes sur la matrice de transition F (avec la contrainte supplémentaire que les bruits de mesure et de modèle soient décorrélés). Dès lors, il est intéressant de comparer ces deux algorithmes d'apprentissage car l'un ne fait pas d'approximation et l'autre réduit l'espace de recherche des paramètres du filtre.

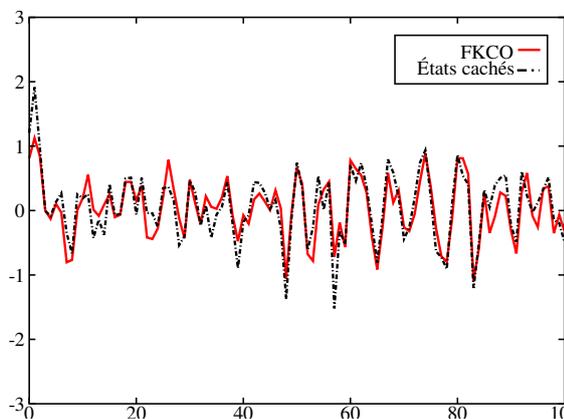
Nous avons donc simulé des signaux respectant les hypothèses du filtre équivalent de type- γ dont les paramètres sont donnés dans la ligne 1 du tableau 3.3. Les signaux ont ensuite été restaurés par

0. le filtre de Kalman couple optimal (FKCO),
1. un filtre de Kalman couple libre (FKCL),
2. un filtre de Kalman couple équivalent de type γ (FKCE), où $F^{y,t} = [1, 0]$,
3. un filtre de Kalman couple contraint (FKC contraint) où
 - nous avons imposé $Q^{x,y} = 0$ et donc estimé $F^{x,t}$ et $F^{y,t}$ séparément,
 - nous avons imposé une contrainte de type-O sur $F^{y,t} = [1, 0]$,
 - et $Q^{x,x}$, $Q^{y,y}$ et $F^{x,t}$ sont laissés libres.

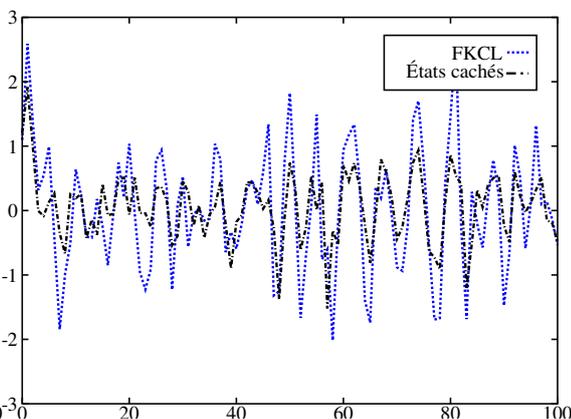
Les paramètres d'initialisation des filtres non-supervisés sont donnés dans la ligne 2 du tableau 3.3. Les estimations moyennes par les trois filtres non-supervisés sont reportées dans les lignes 3 à 5. Un exemple de signal généré est représenté sur la figure 3.8a et ses restaurations respectives sur les figures 3.8b, 3.8c, 3.8d et 3.8e. L'EQM des états lissés est représentées sur la figure 3.9.



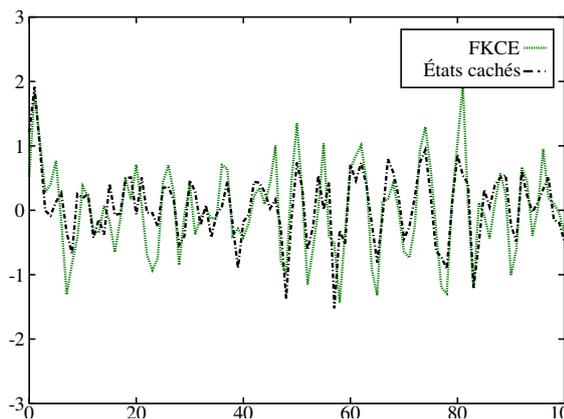
(a) Signal simulé.



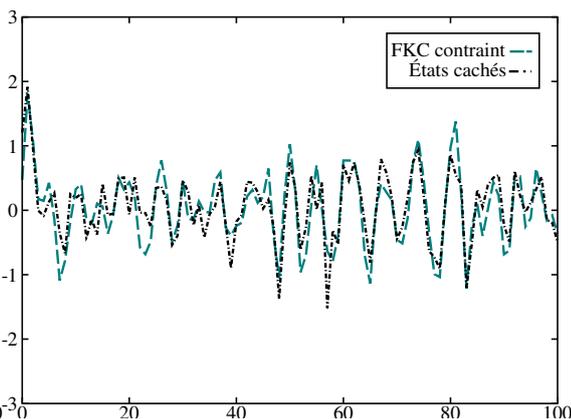
(b) Signal restauré par le FKCO.



(c) Signal restauré par le FKCL.



(d) Signal restauré par le FKCE.



(e) Signal restauré par le FKC contraint.

FIGURE 3.8: Exemple d'un signal de 100 échantillons simulé et de ses restaurations par les différents filtres (section 3.3).

TABLE 3.3: Paramètres de l'expérience de la section 3.3. Ligne 1 : paramètres de simulation ; Ligne 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 3 à 5 : estimation moyenne des paramètres par les différents filtres pour un signal d'une taille de 1 000 échantillons (100 itérations EM).

	\hat{t}_0	Q_0	F	Q
θ	$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0.1 \\ 0.1 & 1.5 \end{bmatrix}$	$\begin{bmatrix} 0.8 & -0.4 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.1 & 0.2 \\ 0.2 & 1 \end{bmatrix}$
$\theta^{(0)}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
$\theta_{FKCL}^{(100)}$	$\begin{bmatrix} 1.648 \\ -1.171 \end{bmatrix}$	$\begin{bmatrix} 0.010 & 0.006 \\ 0.006 & 0.067 \end{bmatrix}$	$\begin{bmatrix} 0.662 & -0.463 \\ 0.673 & 0.133 \end{bmatrix}$	$\begin{bmatrix} 0.445 & -0.146 \\ -0.146 & 0.748 \end{bmatrix}$
$\theta_{FKCL}^{(100)}$	$\begin{bmatrix} 0.975 \\ -1.171 \end{bmatrix}$	$\begin{bmatrix} 0.007 & 0.013 \\ 0.013 & 0.067 \end{bmatrix}$	$\begin{bmatrix} 0.796 & -0.399 \\ \mathbf{1} & \mathbf{0} \end{bmatrix}$	$\begin{bmatrix} 0.193 & 0.004 \\ 0.004 & 0.748 \end{bmatrix}$
$\theta_{FKC\text{ contraint}}^{(100)}$	$\begin{bmatrix} 1.012 \\ -1.383 \end{bmatrix}$	$\begin{bmatrix} 0.007 & 0.011 \\ 0.011 & 0.068 \end{bmatrix}$	$\begin{bmatrix} 0.767 & -0.399 \\ \mathbf{1} & \mathbf{0} \end{bmatrix}$	$\begin{bmatrix} 0.218 & \mathbf{0} \\ \mathbf{0} & 0.728 \end{bmatrix}$

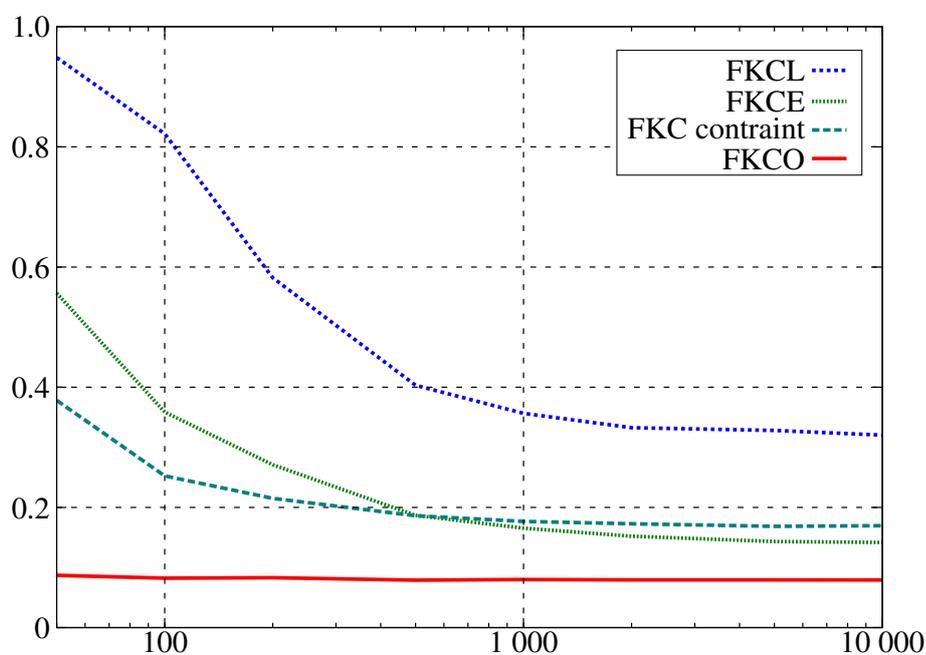


FIGURE 3.9: Évolution de l'EQM en fonction du nombre d'échantillons (section 3.3).

Ces expériences conduisent aux commentaires suivants :

- Concernant l'exemple de restauration, nous remarquons que le filtre contraint donne une meilleure estimation que le filtre équivalent, qui lui-même donne une meilleure que celle du filtre libre. On peut nuancer ces résultats par le fait que les performances du filtre contraint restent assez éloignées de celles du filtre optimal.
- Concernant l'estimation des paramètres, les filtres équivalents et contraints donnent tous les deux une meilleure estimation de la matrice F que le filtre libre, et qui est proche des paramètres de simulation.
- Concernant les courbes d'EQM, on remarque que pour un faible nombre d'échantillons, le filtre contraint donne de meilleures performances que le filtre équivalent. Par contre, pour un grand nombre d'échantillons, le FKCE devient plus performant que le FKC contraint. Ces deux filtres donnent significativement de meilleures performances que le FKCL et qui se rapprochent de celles du filtre optimal quand le nombre d'échantillons devient assez grand.

Cette expérience conduit à des conclusions similaires à celle de l'expérience précédente :

1. Les filtres équivalents ou contraints améliorent l'estimation des états cachés et des paramètres du modèle par rapport au filtre libre.
2. Pour un grand nombre d'échantillons, les performances du filtre équivalent se rapprochent de celles du filtre optimal. Par contre, les performances du filtre contraint ne convergent pas vers celle du filtre optimal.

On peut conclure que quand les hypothèses sont similaires, il est préférable d'utiliser un filtre contraint plutôt qu'un filtre équivalent. Cependant, le résultat est plus mitigé quand on doit faire des approximations pour utiliser le filtre contraint à la place du filtre équivalent. Finalement, on doit faire attention que les hypothèses du filtre contraint ne soit pas trop éloignées du modèle sinon on risque de dégrader significativement l'estimation des paramètres et la restauration du signal.

3.4 Comparaison entre filtres contraints

Dans cette dernière expérience, nous avons choisi d'illustrer l'apport des connaissances a-priori sur l'estimation des paramètres du filtre de Kalman couple. Pour cela, nous avons utilisé des filtres de plus en plus contraints pour estimer ces paramètres et restaurer les états cachés. Nous avons donc simulé des signaux respectant les hypothèses du filtre de Kalman dont les paramètres sont donnés dans la ligne 1 du tableau 3.4. Les signaux ont ensuite été restaurés par

0. le filtre de Kalman couple optimal (FKCO),
1. un filtre de Kalman couple libre (FKCL),

TABLE 3.4: Paramètres de l'expérience de la section 3.4. Ligne 1 : paramètres de simulation ; Ligne 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 3 à 6 : estimation moyenne des paramètres par les différents filtres pour des signaux d'une taille de 100 échantillons ; Lignes 7 à 10 : estimation moyenne des paramètres pour des signaux d'une taille de 1 000 échantillons (100 itérations EM).

	\hat{t}_0	Q_0	F	Q
θ	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.7 \end{bmatrix}$	$\begin{bmatrix} 0.2 & -0.4 \\ 0.8 & 0.4 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0 \\ 0 & 1.5 \end{bmatrix}$
$\theta^{(0)}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$
100 échantillons				
$\theta_{FKCL}^{(100)}$	$\begin{bmatrix} -0.156 \\ 0.130 \end{bmatrix}$	$\begin{bmatrix} 0.025 & -0.037 \\ -0.037 & 0.140 \end{bmatrix}$	$\begin{bmatrix} 0.357 & -0.338 \\ 0.909 & 0.241 \end{bmatrix}$	$\begin{bmatrix} 0.955 & -0.005 \\ -0.005 & 0.849 \end{bmatrix}$
$\theta_{FKC \text{ constraint } 1}^{(100)}$	$\begin{bmatrix} -0.155 \\ -0.030 \end{bmatrix}$	$\begin{bmatrix} 0.039 & -0.054 \\ -0.054 & 0.141 \end{bmatrix}$	$\begin{bmatrix} 0.187 & -0.373 \\ 0.820 & 0.361 \end{bmatrix}$	$\begin{bmatrix} 1.076 & 0 \\ 0 & 1.047 \end{bmatrix}$
$\theta_{FKC \text{ constraint } 2}^{(100)}$	$\begin{bmatrix} -0.167 \\ 0.044 \end{bmatrix}$	$\begin{bmatrix} 0.038 & -0.043 \\ -0.043 & 0.119 \end{bmatrix}$	$\begin{bmatrix} 0.194 & -0.389 \\ 0.810 & 0.389 \end{bmatrix}$	$\begin{bmatrix} 0.493 & 0 \\ 0 & 1.478 \end{bmatrix}$
$\theta_{FKC \text{ constraint } 3}^{(100)}$	$\begin{bmatrix} -0.169 \\ 0.051 \end{bmatrix}$	$\begin{bmatrix} 0.037 & -0.042 \\ -0.042 & 0.119 \end{bmatrix}$	$\begin{bmatrix} 0.195 & -0.391 \\ 0.805 & 0.391 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0 \\ 0 & 1.5 \end{bmatrix}$
1 000 échantillons				
$\theta_{FKCL}^{(100)}$	$\begin{bmatrix} -0.070 \\ 0.281 \end{bmatrix}$	$\begin{bmatrix} 0.025 & -0.042 \\ -0.042 & 0.110 \end{bmatrix}$	$\begin{bmatrix} 0.335 & -0.372 \\ 0.842 & 0.264 \end{bmatrix}$	$\begin{bmatrix} 0.780 & 0.146 \\ 0.146 & 1.255 \end{bmatrix}$
$\theta_{FKC \text{ constraint } 1}^{(100)}$	$\begin{bmatrix} -0.110 \\ 0.278 \end{bmatrix}$	$\begin{bmatrix} 0.036 & -0.042 \\ -0.042 & 0.108 \end{bmatrix}$	$\begin{bmatrix} 0.199 & -0.398 \\ 0.809 & 0.382 \end{bmatrix}$	$\begin{bmatrix} 0.774 & 0 \\ 0 & 1.301 \end{bmatrix}$
$\theta_{FKC \text{ constraint } 2}^{(100)}$	$\begin{bmatrix} -0.110 \\ 0.266 \end{bmatrix}$	$\begin{bmatrix} 0.038 & -0.041 \\ -0.041 & 0.105 \end{bmatrix}$	$\begin{bmatrix} 0.200 & -0.400 \\ 0.800 & 0.400 \end{bmatrix}$	$\begin{bmatrix} 0.498 & 0 \\ 0 & 1.495 \end{bmatrix}$
$\theta_{FKC \text{ constraint } 3}^{(100)}$	$\begin{bmatrix} -0.106 \\ 0.262 \end{bmatrix}$	$\begin{bmatrix} 0.038 & -0.040 \\ -0.040 & 0.104 \end{bmatrix}$	$\begin{bmatrix} 0.200 & -0.400 \\ 0.800 & 0.400 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0 \\ 0 & 1.5 \end{bmatrix}$

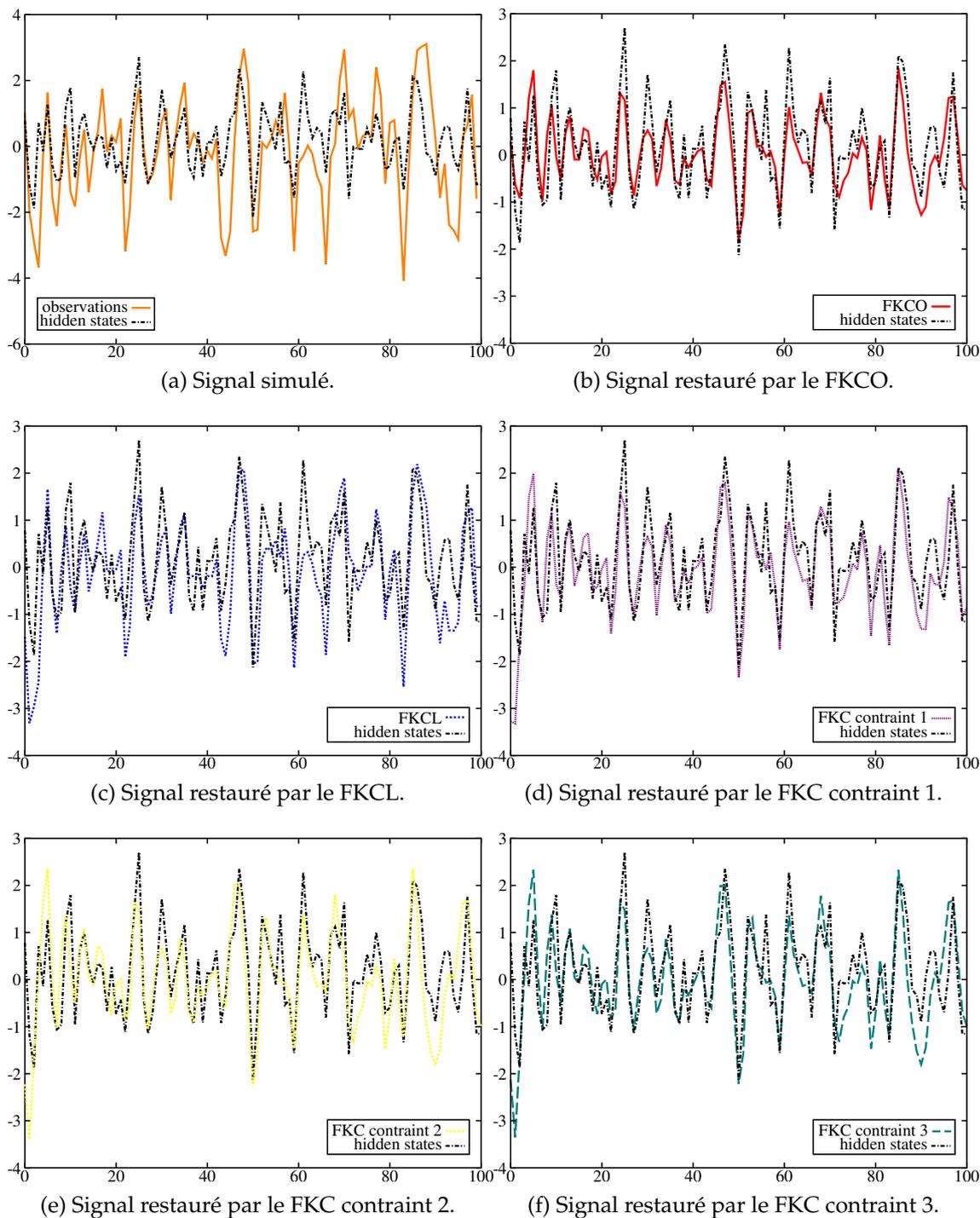


FIGURE 3.10: Exemple d'un signal de 100 échantillons simulé et de ses restaurations par les différents filtres (section 3.4).

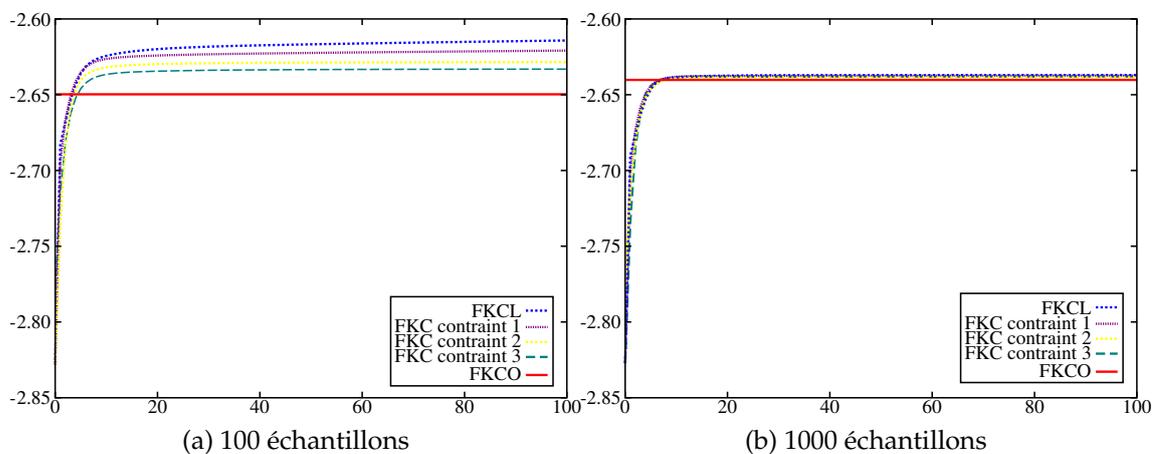


FIGURE 3.11: Vraisemblances moyennes des différents filtres (section 3.4).

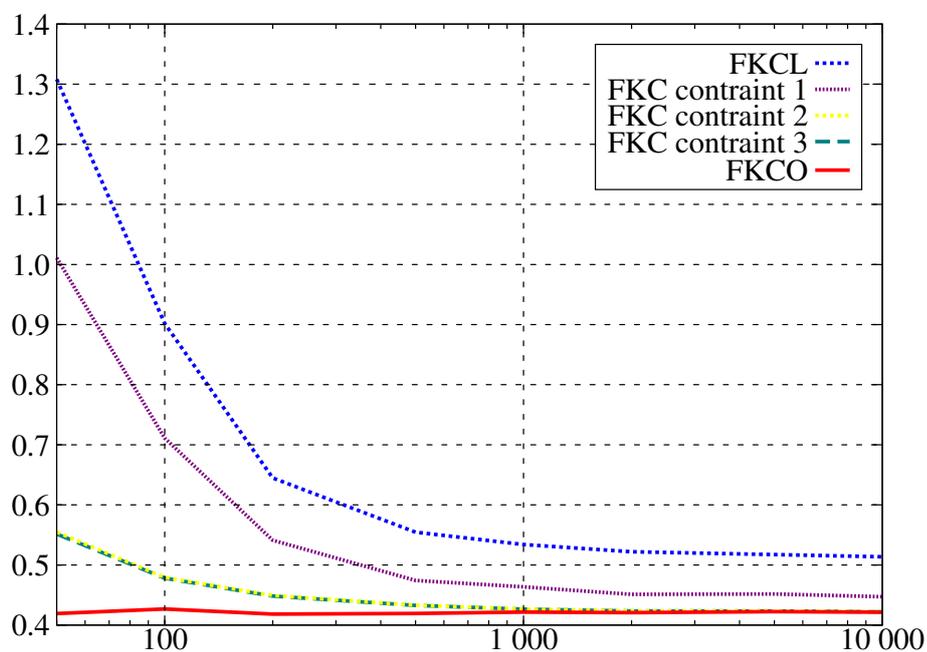


FIGURE 3.12: Évolution de l'EQM en fonction du nombre d'échantillons (section 3.4).

2. un filtre de Kalman couple contraint (FKC contraint 1) où
- nous avons imposé $Q^{x,y} = 0$ et donc estimé $F^{x,t}$ et $F^{y,t}$ séparément,
 - nous avons imposé une contrainte de type-2 sur

$$F^{x,t} = G_x [1, -2],$$

où G_x est un scalaire à estimer,

- une contrainte de type-2 sur

$$F^{y,t} = [1, 0] + G_y [-1, 2],$$

où G_y est un scalaire à estimer,

- et $Q^{x,x}$, $Q^{y,y}$ sont laissés libres.

3. un filtre de Kalman couple contraint (FKC contraint 2) où
- nous avons imposé une contrainte de type-3 sur

$$F = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \lambda \begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix},$$

où λ est un scalaire à estimer,

- et imposé conjointement une contrainte de type B sur la matrice

$$Q = \gamma \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix},$$

où γ est un scalaire à estimer.

4. un filtre de Kalman couple contraint (FKC contraint 3) où
- nous avons imposé une contrainte de type-3

$$F = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \lambda \begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix},$$

où λ est un scalaire à estimer,

- et la matrice Q est connue.

Les paramètres d'initialisation des filtres non-supervisés sont donnés dans la ligne 2 du tableau 3.4. Les estimations moyennes par les quatre filtres non-supervisés sont reportées dans les lignes 3 à 6 pour des signaux de 100 points et 7 à 10 pour des signaux de 1000 points. Un exemple de signal généré est représenté sur la figure 3.10a et ses restaurations respectives sur les figures 3.10b, 3.10c, 3.10d, 3.10e et 3.10f. Les courbes de vraisemblances moyennes pour les différents filtres sont reportées sur les figures 3.11a et 3.11b pour des tailles de signaux respectives de 100 et 1000 échantillons. L'EQM des signaux lissés est représentée sur la figure 3.12.

Il convient d'analyser ces résultats séparément :

- Concernant la restauration, plus le filtre est contraint, plus la restauration s'approche des vraies valeurs des états cachés. On note aussi que les filtres contraints 2 et 3 ont des performances très proches.

- Concernant les courbes de vraisemblances, plus le filtre est contraint, plus il a tendance à converger rapidement. Cela se remarque particulièrement sur la courbe de vraisemblance moyenne pour 100 échantillons. De plus, on remarque que le filtre libre n’a pas fini de converger pour 100 échantillons contrairement aux deux autres filtres. On peut nuancer que plus le modèle est contraint, plus la vraisemblance finale est faible.
- Concernant les estimations moyennes des paramètres, nous remarquons que les filtres contraints sont proches des paramètres de simulation. En particulier, les filtres contraints 2 et 3 ont presque convergé vers les vraies valeurs, même pour le cas difficile de 100 échantillons.
- Concernant les courbes d’EQM, on remarque que plus le filtre est contraint, plus l’EQM est faible. De plus, les filtres contraints 2 et 3 ont des performances presque similaires. Tous les filtres semblent converger vers l’EQM de la solution optimale quand le nombre d’échantillons devient important.

Nous pouvons tirer 2 conclusions de cette expérience :

- L’ajout de contraintes sur les paramètres permet d’une part, d’améliorer leur estimation et d’autre part, d’améliorer la restauration des états cachés, et cela même dans le cas difficile des petits signaux.
- Par contre, on peut nuancer que plus les paramètres sont contraints, moins bonne est la vraisemblance finale. Donc, on peut s’attendre à ce que l’algorithme contraint puisse rester piéger par des maxima locaux de la vraisemblance différents de ceux des algorithmes original et équivalent.

4 Conclusion

Dans ce chapitre, nous avons établi des contraintes sur les paramètres du filtre de Kalman couple qui permettent des formules de ré-estimation exacte. Contrairement à celles présentées dans le chapitre précédent, ces contraintes modifient le chemin parcouru par la vraisemblance. Nous avons également comparé les filtres équivalents aux filtres contraints. Mis à part pour le filtre équivalent de type α , les filtres équivalents et contraints ont montré des performances similaires quand les hypothèses des modèles équivalents étaient respectées. Nous avons également comparé plusieurs filtres contraints pour mesurer l’influence des connaissances a-priori sur les paramètres et sur l’estimation de l’algorithme EM. Nous avons illustré que les contraintes améliorent l’estimation et la restauration. Mais, on peut nuancer que plus le système est contraint, moins la vraisemblance croît rapidement. Donc, le fait de contraindre le système peut amener l’algorithme EM à rester piéger sur un maximum local de la vraisemblance.

Dans le chapitre suivant, nous proposons de comparer nos algorithmes d’apprentissage aux deux algorithmes EM du filtre de Kalman sur des données qui ne sont plus forcément adaptées au filtre de Kalman couple. De plus, il est envisageable de contraindre l’étape de maximisation de l’algorithme EM et d’utiliser un jeu de paramètres équivalents, comme nous le faisons pour le filtre de Kalman.

Chapitre 4

Discussion

Ce chapitre propose une discussion sur l'apport du filtre de Kalman couple vis-à-vis du filtre de Kalman. Même si le premier modèle est strictement plus général que le second, le nombre de paramètres inconnus est plus important. Cela rend la tâche des algorithmes EM plus difficile puisque l'estimation est réalisée dans des espaces de plus grandes dimensions. Il est alors intéressant de comprendre si la présence de contraintes sur le modèle couple –réduisant la dimension de l'espace de recherche–, peut apporter des gains en termes de restauration de signaux par rapport à un filtre de Kalman à nombre de paramètres identique.

Aussi, ce chapitre reporte t'il les résultats d'expériences comparatives dans trois situations distinctes, permettant d'apprécier le comportement des modèles et des algorithmes d'estimation. La première section se concentre sur des données simulées, issues des modèles que nous souhaitons évaluer. La seconde section évalue les algorithmes sur des données simulées, mais ne suivant pas les modèles. La dernière section reporte les résultats comparatifs du filtre de Kalman et du filtre de Kalman couple sur des données réelles issues de l'application de biométrie dynamique d'iris présentée dans le chapitre 5.

1 Données théoriques

Dans ces premières séries d'expériences, nous avons choisi d'illustrer les performances du filtre de Kalman couple sur des données théoriques par rapport au filtre de Kalman. Nous avons réalisé deux séries d'expériences, l'une comparant différents filtres de Kalman couple avec deux filtres de Kalman sur des données adaptées au filtre de Kalman, et l'autre comparant encore ces filtres sur des données adaptées au filtre de Kalman couple. Tous les résultats sont la moyenne de 1 000 expériences et nous avons mesuré la qualité de la restauration par l'erreur quadratique moyenne (EQM). La taille des signaux varie de 50 à 10 000 échantillons.

TABLE 4.1: Paramètres de l'expérience de la section 1.1. Ligne 1 : paramètres de simulation ; Ligne 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 3 à 6 : estimation moyenne des paramètres par les différents filtres pour un signal d'une taille de 1 000 échantillons (100 itérations EM).

	\hat{t}_0	Q_0	F	Q
θ	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.8 & 0 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0 \\ 0 & 0 \end{bmatrix}$
$\theta^{(0)}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
$\theta_{FKE}^{(100)}$	$\begin{bmatrix} 0.843 \\ * \end{bmatrix}$	$\begin{bmatrix} 0.008 & * \\ * & * \end{bmatrix}$	$\begin{bmatrix} 0.797 & \mathbf{0} \\ \mathbf{1} & \mathbf{0} \end{bmatrix}$	$\begin{bmatrix} 0.508 & \mathbf{0} \\ \mathbf{0} & 1.507 \end{bmatrix}$
$\theta_{FK\text{ constraint}}^{(100)}$	$\begin{bmatrix} 0.881 \\ * \end{bmatrix}$	$\begin{bmatrix} 0.009 & * \\ * & * \end{bmatrix}$	$\begin{bmatrix} 0.791 & \mathbf{0} \\ \mathbf{1} & \mathbf{0} \end{bmatrix}$	$\begin{bmatrix} 0.521 & \mathbf{0} \\ \mathbf{0} & 1.497 \end{bmatrix}$
$\theta_{FKCE}^{(100)}$	$\begin{bmatrix} 0.863 \\ 0.090 \end{bmatrix}$	$\begin{bmatrix} 0.007 & 0.000 \\ 0.000 & 0.931 \end{bmatrix}$	$\begin{bmatrix} 0.740 & 0.031 \\ \mathbf{1} & \mathbf{0} \end{bmatrix}$	$\begin{bmatrix} 0.678 & -0.217 \\ -0.217 & 1.187 \end{bmatrix}$
$\theta_{FKC\text{ constraint}}^{(100)}$	$\begin{bmatrix} 0.904 \\ 0.200 \end{bmatrix}$	$\begin{bmatrix} 0.009 & -0.013 \\ -0.013 & 0.881 \end{bmatrix}$	$\begin{bmatrix} 0.735 & 0.031 \\ \mathbf{1} & \mathbf{0} \end{bmatrix}$	$\begin{bmatrix} 0.525 & \mathbf{0} \\ \mathbf{0} & 1.437 \end{bmatrix}$

1.1 Données suivant le modèle du filtre de Kalman

Il est intéressant de vérifier que le filtre de Kalman couple non-supervisé obtient des performances de restauration similaire à celles du filtre de Kalman non-supervisé sur des signaux adaptés au filtre de Kalman. Pour cela, nous avons simulé des signaux avec les paramètres donnés dans la ligne 1 du tableau 4.1. Ces données ont ensuite été restaurées par

0. le filtre de Kalman optimal (FKO),
1. un filtre de Kalman équivalent (FKE cf. Annexe 3.1) où $F^{y,x} = 1$,
2. un filtre de Kalman contraint (FK contraint cf. Annexe 3.2) où $F^{y,x} = 1$,
3. un filtre de Kalman couple équivalent (FKCE) où $F^{y,t} = [1, 0]$,
4. un filtre de Kalman couple contraint (FKC contraint) où
 - nous avons imposé $Q^{x,y} = 0$ et donc une estimation séparée de $F^{x,t}$ et $F^{y,t}$,
 - nous avons imposé une contrainte de type 0 sur $F^{y,t} = [1, 0]$ et
 - nous avons laissé $F^{x,y}$, $Q^{x,x}$ et $Q^{y,y}$, libres.

Les paramètres d'initialisation des filtres non-supervisés sont donnés dans la ligne 2 du tableau 4.1. Les estimations moyennes par les quatre filtres non-supervisés sont reportées dans les lignes 3 à 6. Un exemple de signal généré est représenté sur la figure 4.2a et ses restaurations respectives sur les figures 4.2b, 4.2c, 4.2d, 4.2e et 4.2f. L'EQM est représentée sur la figure 4.1.

Il est intéressant d'analyser ces résultats séparément :

- Concernant la restauration, les performances des filtres de Kalman non-supervisés sont très proches de celles du filtre de Kalman optimal. Par contre,

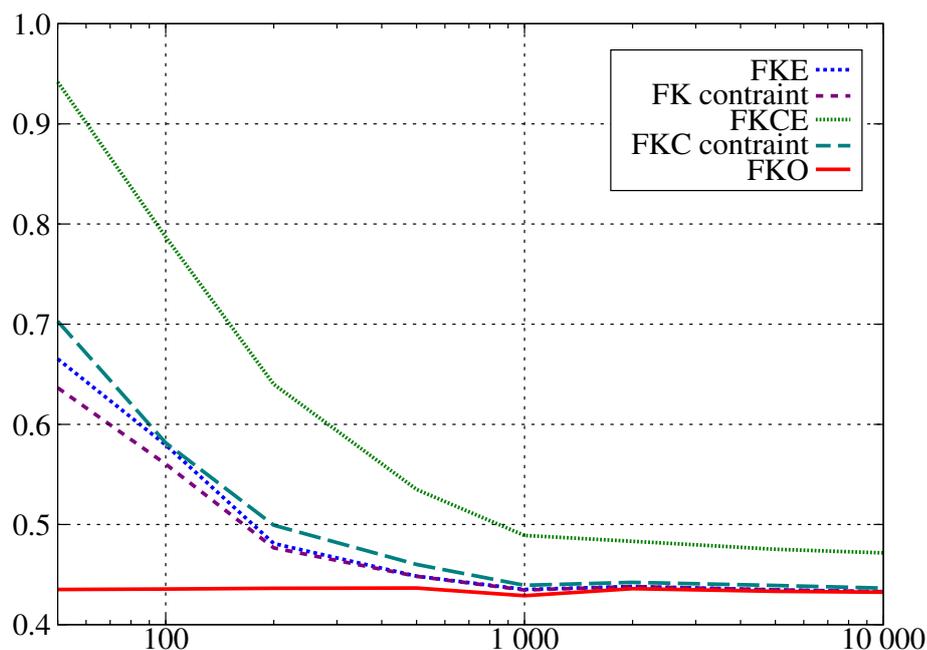


FIGURE 4.1: Évolution de l'EQM en fonction du nombre d'échantillons (section 1.1).

les filtres de Kalman couples intègrent une partie des observations dans le signal restauré, ce qui dégrade légèrement leur performance.

- Concernant les estimations moyennes des paramètres, nous remarquons que plus le modèle est contraint, meilleure est l'estimation. De plus, les paramètres estimés sont très proches des paramètres de simulation pour tous les filtres.
- Concernant les courbes d'EQM, nous remarquons que plus le modèle est contraint, meilleures sont les performances. Nous pouvons aussi ajouter que le filtre de Kalman couple contraint donne des résultats similaires à ceux des filtres de Kalman non-supervisés. Quand le nombre d'échantillons dépasse 1000, les performances des filtres de Kalman et du filtre de Kalman couple contraint se confondent avec celles du filtre optimal.

Ces résultats conduisent à deux conclusions principales :

- Comme prévu, les filtres de Kalman non-supervisés donnent de meilleurs résultats que ceux de Kalman couple. Cependant, les différences de performances ne sont pas significatives.
- Le grand nombre de paramètres du filtre de Kalman couple ne l'empêche pas d'être optimal quand le nombre d'échantillons tend vers l'infini.

1.2 Données suivant le modèle du filtre de Kalman couple

Il est maintenant intéressant de vérifier que le filtre de Kalman ne peut pas restaurer correctement des signaux suivant le modèle du filtre de Kalman couple. Nous avons simulé des signaux dont les paramètres sont donnés dans la ligne 1

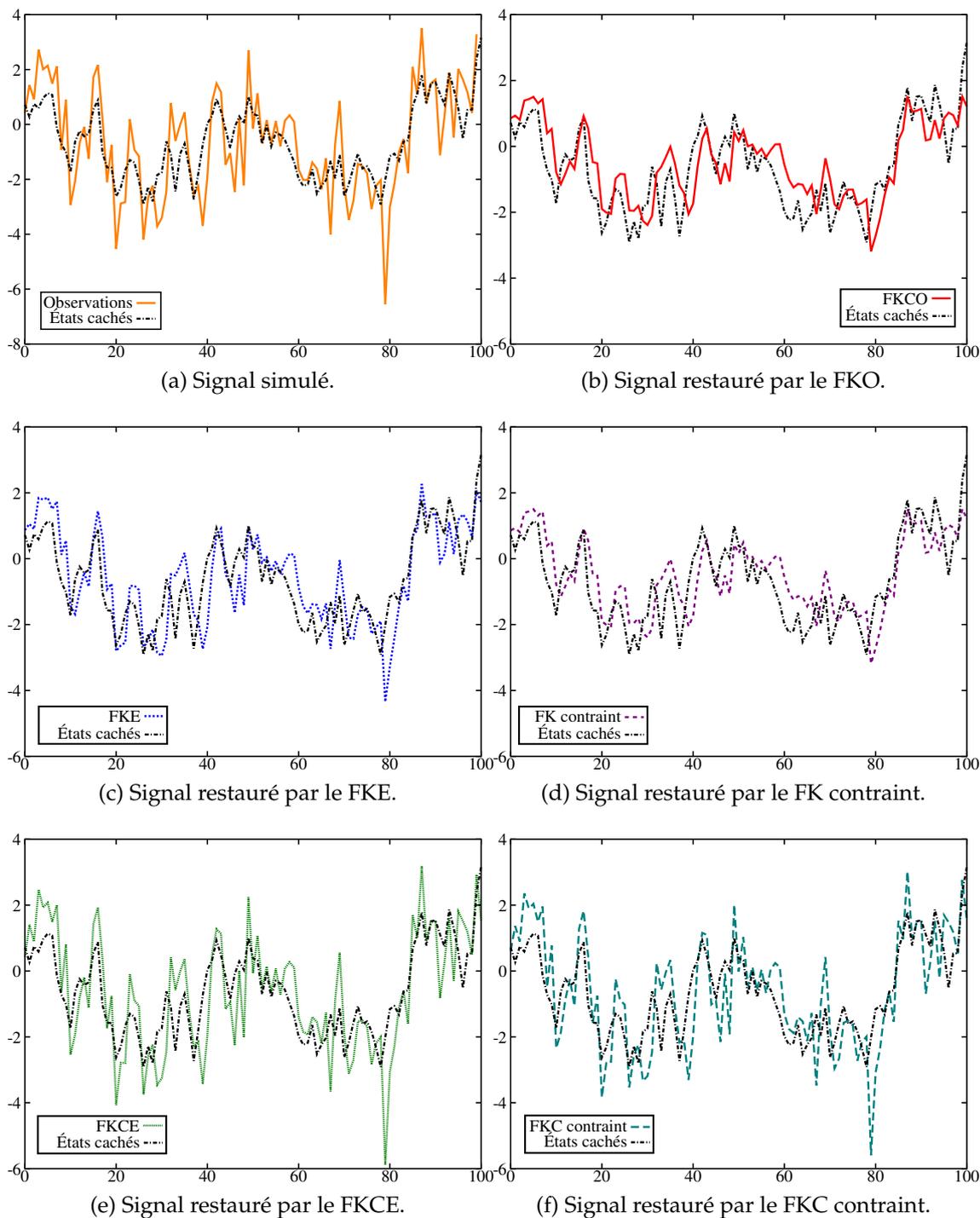


FIGURE 4.2: Exemple d'un signal de 100 échantillons simulé et de ses restaurations par les différents filtres (section 1.1).

TABLE 4.2: Paramètres de l'expérience de la section 1.2. Ligne 1 : paramètres de simulation ; Ligne 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 3 à 6 : estimation moyenne des paramètres par les différents filtres pour un signal d'une taille de 1 000 échantillons (100 itérations EM).

	\hat{t}_0	Q_0	F	Q
θ	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.8 & -0.4 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0 \\ 0 & 1.5 \end{bmatrix}$
$\theta^{(0)}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
$\theta_{FKE}^{(100)}$	$\begin{bmatrix} 1.133 \\ * \end{bmatrix}$	$\begin{bmatrix} 0.006 & * \\ * & * \end{bmatrix}$	$\begin{bmatrix} 0.284 & \mathbf{0} \\ \mathbf{1} & \mathbf{0} \end{bmatrix}$	$\begin{bmatrix} 2.265 & \mathbf{0} \\ \mathbf{0} & 0.377 \end{bmatrix}$
$\theta_{FK\text{ constraint}}^{(100)}$	$\begin{bmatrix} 1.113 \\ * \end{bmatrix}$	$\begin{bmatrix} 0.007 & * \\ * & * \end{bmatrix}$	$\begin{bmatrix} 0.287 & \mathbf{0} \\ \mathbf{1} & \mathbf{0} \end{bmatrix}$	$\begin{bmatrix} 2.207 & \mathbf{0} \\ \mathbf{0} & 0.436 \end{bmatrix}$
$\theta_{FKCE}^{(100)}$	$\begin{bmatrix} 1.046 \\ -0.136 \end{bmatrix}$	$\begin{bmatrix} 0.012 & 0.024 \\ 0.024 & 0.133 \end{bmatrix}$	$\begin{bmatrix} 0.792 & -0.401 \\ \mathbf{1} & \mathbf{0} \end{bmatrix}$	$\begin{bmatrix} 0.557 & -0.104 \\ -0.104 & 1.362 \end{bmatrix}$
$\theta_{FKC\text{ constraint}}^{(100)}$	$\begin{bmatrix} 1.069 \\ -0.176 \end{bmatrix}$	$\begin{bmatrix} 0.013 & 0.023 \\ 0.023 & 0.134 \end{bmatrix}$	$\begin{bmatrix} 0.779 & -0.401 \\ \mathbf{1} & \mathbf{0} \end{bmatrix}$	$\begin{bmatrix} 0.538 & \mathbf{0} \\ \mathbf{0} & 1.473 \end{bmatrix}$

du tableau 4.2. Les données ont ensuite été restaurées par

0. le filtre de Kalman couple optimal (FKCO),
1. un filtre de Kalman équivalent (FKE) où $F^{y,x} = 1$,
2. un filtre de Kalman contraint (FK contraint) où $F^{y,x} = 1$,
3. un filtre de Kalman couple équivalent (FKCE) où $F^{y,t} = [1, 0]$,
4. un filtre de Kalman couple contraint (FKC contraint) où
 - nous avons imposé $Q^{x,y} = 0$ et donc une estimation séparée de $F^{x,t}$ et $F^{y,t}$,
 - nous avons imposé une contrainte de type 0 sur $F^{y,t} = [1, 0]$ et
 - nous avons laissé $F^{x,y}$, $Q^{x,x}$ et $Q^{y,y}$ libres.

Les paramètres d'initialisation des filtres non-supervisés sont donnés dans la ligne 2 du tableau 4.2. Les estimations moyennes par les quatre filtres non-supervisés sont reportées dans les lignes 3 à 6. Un exemple de signal généré est représenté sur la figure 4.4a et ses restaurations respectives sur les figures 4.4b, 4.4c, 4.4d, 4.4e et 4.4f. L'EQM est représentée sur la figure 4.3.

Ces résultats sont analysés séparément :

- Concernant la restauration, nous remarquons que le FK contraint donne des résultats très proches de ceux du filtre optimal, que le FKCE est relativement proche et que les deux filtres de Kalman recopient seulement les observations (aucun filtrage).
- Concernant les estimations moyennes des paramètres, nous remarquons que le FKC contraint donne une estimation très proche des paramètres de simulation, que le FKCE en donne relativement une bonne estimation et

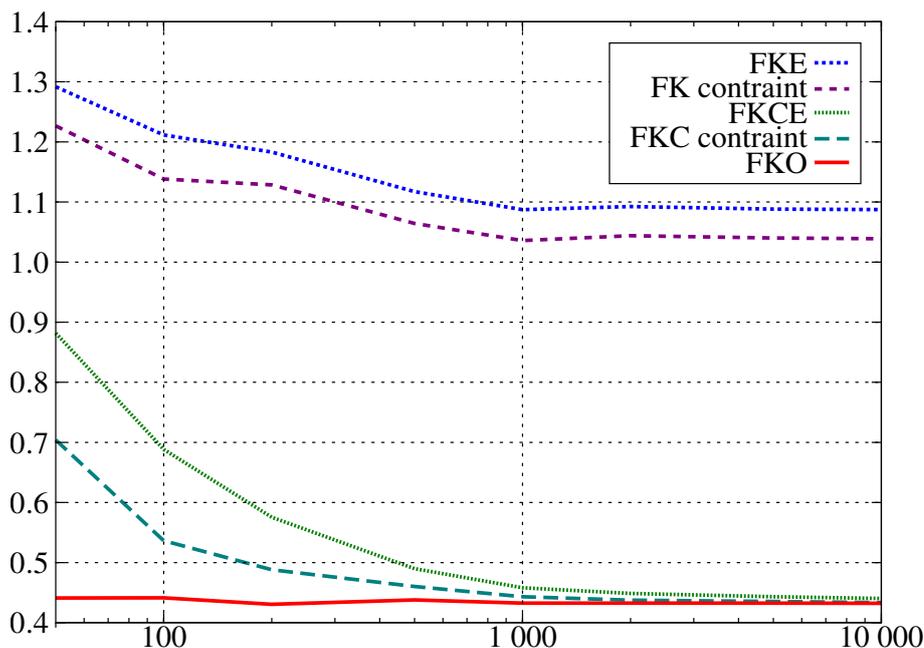


FIGURE 4.3: Évolution de l'EQM en fonction du nombre d'échantillons (section 1.2).

que les filtres de Kalman restent très éloignés des paramètres optimaux.

- Concernant les courbes d'EQM, nous remarquons que l'EQM du FKC contraint est très proche de celle du filtre optimal, et que les performances du FKCE tendent vers celles du FKCO quand le nombre d'échantillons devient important. Par contre, les EQM des filtres de Kalman sont plus élevées que celles des autres filtres et surtout ne convergent pas vers celle du filtre optimal.

De ces résultats, nous pouvons conclure que

1. Comme attendu, le filtre de Kalman non-supervisé n'est pas adapté à la restauration de données suivant le modèle du filtre de Kalman couple.
2. Les FKCE et FKC contraint montrent encore une fois des performances intéressantes sur des données qui leur sont adaptées et s'avèrent être asymptotiquement efficaces quand le nombre d'échantillons tend vers l'infini.

2 Données simulées

Dans cette seconde série d'expériences, nous travaillons sur des données simulées ne suivant plus le modèle du filtre de Kalman couple. Nous souhaitons montrer l'intérêt du filtre de Kalman couple par rapport au filtre de Kalman classique dans ce contexte. Comme pour les expériences précédentes, nous avons mesuré la qualité de la restauration par l'EQM. Tous les résultats sont la moyenne de 100 expériences pour des tailles de signaux différentes.

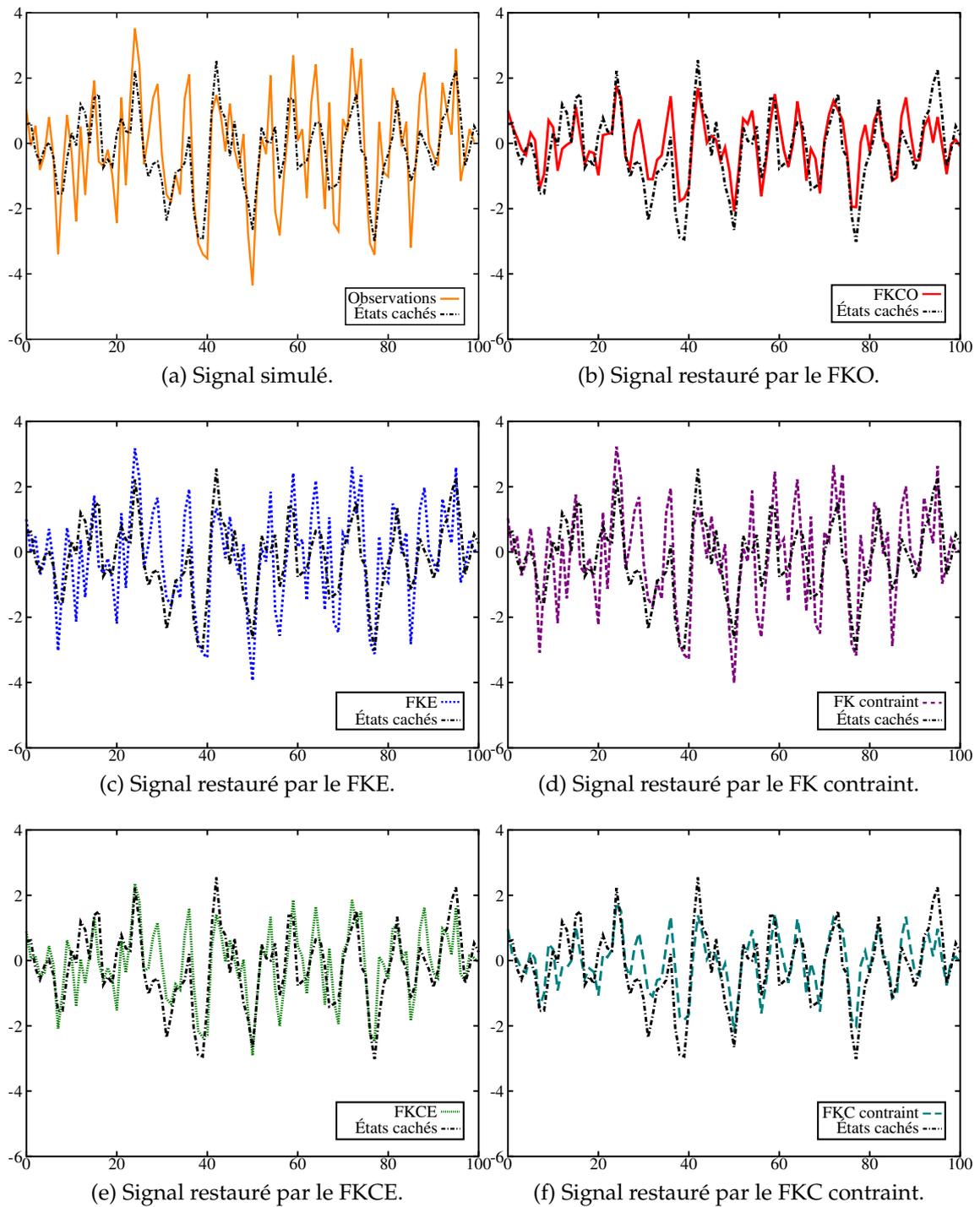


FIGURE 4.4: Exemple d'un signal de 100 échantillons simulé et de ses restaurations par les différents filtres (section 1.2).

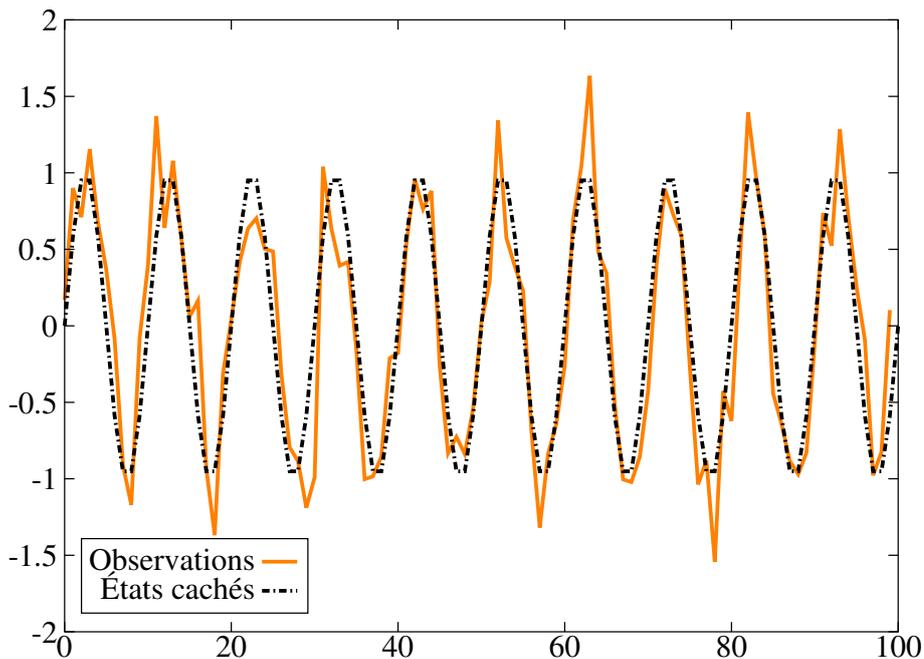


FIGURE 4.5: Exemple de signal généré (section 2.1).

2.1 Signal sinusoïdal

Dans cette première expérience, nous avons considéré le problème de la restauration d'un signal sinusoïdal $x(t) = \cos(10\pi t)$, échantillonné entre les dates $t_1 = 0$ et $t_2 = 1 - T_E$ où $T_E = 1/N$ est la période d'échantillonnage. Les signaux échantillonnés x sont dégradés par des bruits gaussiens centrés de covariance $\mathbf{Q}^{y:y} = 0.1$. Ils sont ensuite restaurés par

1. un filtre de Kalman équivalent (FKE) où $\mathbf{F}^{y,x} = 1$,
2. un filtre de Kalman couple équivalent (FKCE) où $\mathbf{F}^{y,t} = [1, 0]$,
3. un filtre de Kalman augmenté ($n_x = 2$) équivalent (FKAE) où $\mathbf{F}^{y,x} = [1, 0]$,
4. un filtre de Kalman augmenté ($n_x = 2$) contraint (FKA contraint) où
 - nous avons imposé $\mathbf{F}^{y,x} = [1, 0]$ et
 - nous avons laissé $\mathbf{F}^{x,x}$ libre.
5. un filtre de Kalman couple augmenté ($n_x = 2$) équivalent (FKCAE) où $\mathbf{F}^{y,t} = [1, 0, 0]$,
6. un filtre de Kalman couple augmenté ($n_x = 2$) contraint (FKCA contraint) où
 - nous avons imposé une estimation séparée de $\mathbf{F}^{x,t}$ et $\mathbf{F}^{y,t}$ et une décorrélation des bruits de mesure et de modèle,
 - nous avons imposé $\mathbf{F}^{y,t} = [1, 0, 0]$ et
 - nous avons laissé $\mathbf{F}^{x,t}$ libre.

Les paramètres d'initialisation des filtres non-supervisés sont donnés dans les lignes 1 et 2 du tableau 4.3. Les estimations moyennes des filtres non-supervisés sont reportées dans les lignes 3 à 8 pour 100 échantillons et 9 à 14 pour 1000 échantillons. Un exemple de signal généré est représenté sur la figure 4.5. Les EQM in-

TABLE 4.3: Paramètres de l'expérience de la section 2.1. Lignes 1 et 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 4 à 9 : estimation moyenne des paramètres par les différents filtres pour un signal de 100 échantillons ; Lignes 10 à 15 : estimation moyenne des paramètres par les différents filtres pour un signal de 1 000 échantillons (100 ou 600 itérations EM).

	\hat{t}_0	Q_0	F	Q
Initialisation				
$\theta^{(0)}$	$O_{2,1}$	I_2	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$	I_2
$\theta_{AF}^{(0)}$	$O_{3,1}$	I_3	$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$	I_3
100 échantillons				
$\theta_{FKE}^{(100)}$	$\begin{bmatrix} 0.030 \\ * \end{bmatrix}$	$\begin{bmatrix} 0.001 & * \\ * & * \end{bmatrix}$	$\begin{bmatrix} 0.683 & 0 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.321 & 0 \\ 0 & 0.006 \end{bmatrix}$
$\theta_{FKCE}^{(100)}$	$\begin{bmatrix} 0.009 \\ -0.725 \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.002 \\ 0.002 & 0.003 \end{bmatrix}$	$\begin{bmatrix} 1.485 & -0.835 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.026 & 0.053 \\ 0.053 & 0.134 \end{bmatrix}$
$\theta_{FKAE}^{(600)}$	$\begin{bmatrix} 0.008 \\ -4.263 \\ * \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.000 & * \\ 0.000 & 0.001 & * \\ * & * & * \end{bmatrix}$	$\begin{bmatrix} 1.106 & -0.140 & 0 \\ 3.143 & 0.512 & 0 \\ 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.000 & 0 \\ 0.000 & 0.002 & 0 \\ 0 & 0 & 0.096 \end{bmatrix}$
$\theta_{FKAc.}^{(600)}$	$\begin{bmatrix} 0.008 \\ -2.782 \\ * \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.000 & * \\ 0.000 & 0.001 & * \\ * & * & * \end{bmatrix}$	$\begin{bmatrix} 0.944 & -0.215 & 0 \\ 1.698 & 0.673 & 0 \\ 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.000 & 0 \\ 0.000 & 0.001 & 0 \\ 0 & 0 & 0.096 \end{bmatrix}$
$\theta_{FKCAE}^{(600)}$	$\begin{bmatrix} 0.016 \\ -0.145 \\ -1.228 \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.000 & -0.000 \\ 0.000 & 0.100 & -0.029 \\ -0.000 & -0.029 & 0.009 \end{bmatrix}$	$\begin{bmatrix} 1.435 & -0.152 & -0.424 \\ 2.253 & 0.132 & 0.598 \\ 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.003 & 0.034 & -0.009 \\ 0.034 & 1.280 & -0.325 \\ -0.009 & -0.325 & 0.084 \end{bmatrix}$
$\theta_{FKCAc.}^{(600)}$	$\begin{bmatrix} 0.030 \\ -2.052 \\ -0.765 \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.000 & -0.000 \\ 0.000 & 0.005 & -0.002 \\ -0.000 & -0.002 & 0.002 \end{bmatrix}$	$\begin{bmatrix} 0.975 & -0.199 & -0.020 \\ 1.558 & 0.555 & 0.431 \\ 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.006 & 0.040 & 0 \\ 0.040 & 0.313 & 0 \\ 0 & 0 & 0.082 \end{bmatrix}$
1 000 échantillons				
$\theta_{FKE}^{(100)}$	$\begin{bmatrix} 0.099 \\ * \end{bmatrix}$	$\begin{bmatrix} 0.001 & * \\ * & * \end{bmatrix}$	$\begin{bmatrix} 0.981 & 0 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.019 & 0 \\ 0 & 0.086 \end{bmatrix}$
$\theta_{FKCE}^{(100)}$	$\begin{bmatrix} 0.016 \\ 0.825 \end{bmatrix}$	$\begin{bmatrix} 0.001 & -0.003 \\ -0.003 & 0.052 \end{bmatrix}$	$\begin{bmatrix} 0.852 & 0.125 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.022 & 0.007 \\ 0.007 & 0.097 \end{bmatrix}$
$\theta_{FKAE}^{(600)}$	$\begin{bmatrix} -0.019 \\ 0.812 \\ * \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.001 & * \\ 0.001 & 0.008 & * \\ * & * & * \end{bmatrix}$	$\begin{bmatrix} 0.599 & 0.147 & 0 \\ 1.830 & 0.192 & 0 \\ 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.063 & -0.159 & 0 \\ -0.159 & 1.306 & 0 \\ 0 & 0 & 0.032 \end{bmatrix}$
$\theta_{FKAc.}^{(600)}$	$\begin{bmatrix} -0.010 \\ 1.258 \\ * \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.001 & * \\ 0.001 & 0.008 & * \\ * & * & * \end{bmatrix}$	$\begin{bmatrix} 0.755 & 0.157 & 0 \\ 1.364 & 0.033 & 0 \\ 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.058 & -0.240 & 0 \\ -0.240 & 1.927 & 0 \\ 0 & 0 & 0.035 \end{bmatrix}$
$\theta_{FKCAE}^{(600)}$	$\begin{bmatrix} 0.094 \\ -13.684 \\ -2.175 \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.000 & -0.000 \\ 0.000 & 0.042 & 0.015 \\ -0.000 & 0.015 & 0.018 \end{bmatrix}$	$\begin{bmatrix} 0.995 & -0.014 & 0.024 \\ 1.753 & 0.846 & -0.123 \\ 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.004 & 0.008 & -0.001 \\ 0.008 & 0.242 & -0.008 \\ -0.001 & -0.008 & 0.096 \end{bmatrix}$
$\theta_{FKCAc.}^{(600)}$	$\begin{bmatrix} -0.033 \\ 9.202 \\ -1.299 \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.000 & -0.000 \\ 0.000 & 0.076 & -0.029 \\ -0.000 & -0.029 & 0.015 \end{bmatrix}$	$\begin{bmatrix} 0.937 & 0.017 & 0.042 \\ -1.704 & 0.853 & 0.940 \\ 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.005 & -0.003 & 0 \\ -0.003 & 0.250 & 0 \\ 0 & 0 & 0.097 \end{bmatrix}$

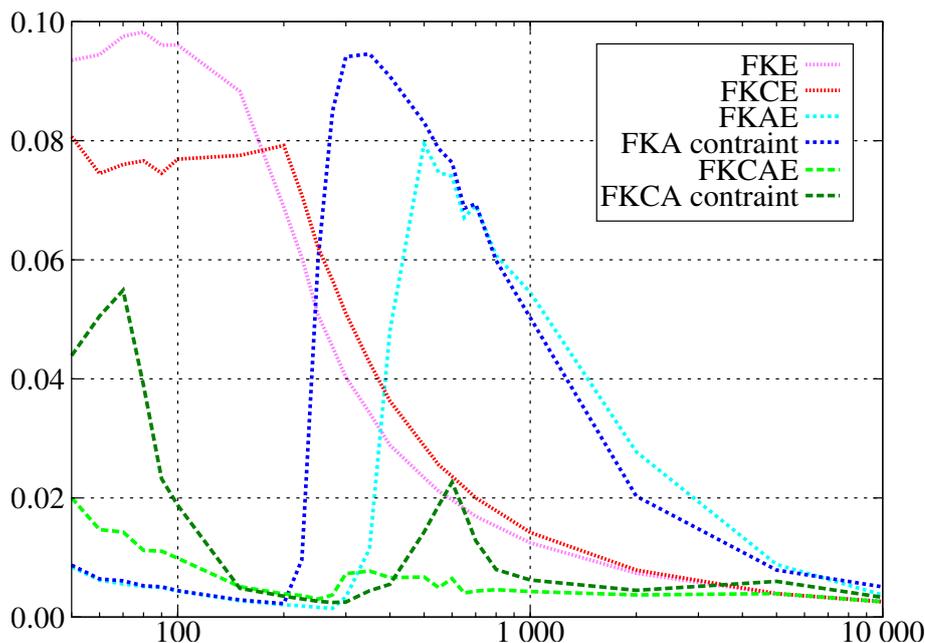


FIGURE 4.6: Évolution de l'EQM en fonction du nombre d'échantillons (section 2.1).

stantannées de restauration sont respectivement représentées sur les figures 4.7a pour les signaux de 100 échantillons, 4.7b pour les signaux de 400 échantillons et 4.7c pour les signaux de 800 échantillons. L'EQM est reportée sur la figure 4.6.

Il convient d'analyser ces résultats séparément :

- Concernant l'estimation des paramètres, on remarque que les valeurs pour 100 ou 1000 échantillons sont très différentes. Pour 100 points, tous les modèles, excepté le FKE, ont une covariance de bruit de modèle faible, ce qui montre leur bonne adéquation aux données. Pour 1000 points, on remarque une covariance de bruit de modèle plus élevée pour les filtres de Kalman augmenté vis-à-vis des autres filtres.
- Concernant les courbes d'EQM, on remarque que pour un faible nombre d'échantillons, les modèles augmentés donnent de meilleurs résultats. On peut noter aussi que les modèles de Kalman augmentés subissent deux décrochages plus importants que ceux que subissent les filtres de Kalman couplés augmentés. De plus, il est intéressant de remarquer que les décrochages des modèles contraints sont plus importants que ceux des modèles équivalents !

Nous pouvons tirer 3 conclusions de cette expérience :

1. Les filtres équivalents ou contraints sont très performants sur des signaux sinusoïdaux.
2. Par contre, le fait de contraindre peut introduire de nouveaux maxima locaux sur la fonction de vraisemblance, et ainsi piéger l'algorithme EM comme le montre cette expérience.
3. Finalement, sur des signaux sinusoïdaux, les filtres équivalents augmen-

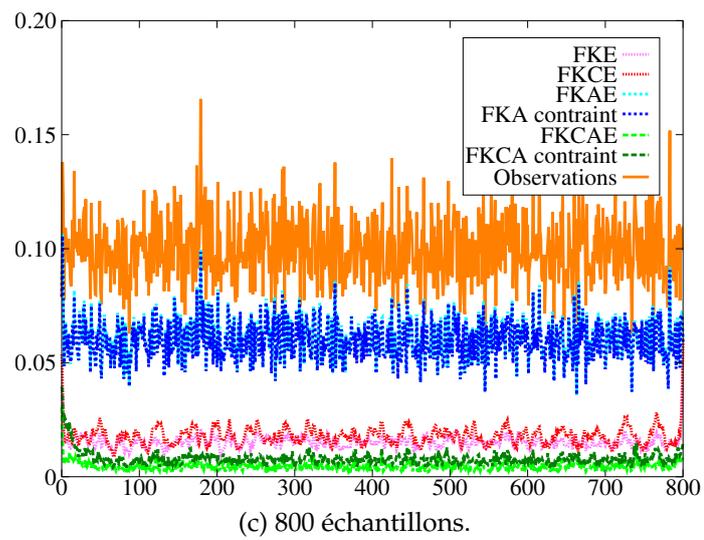
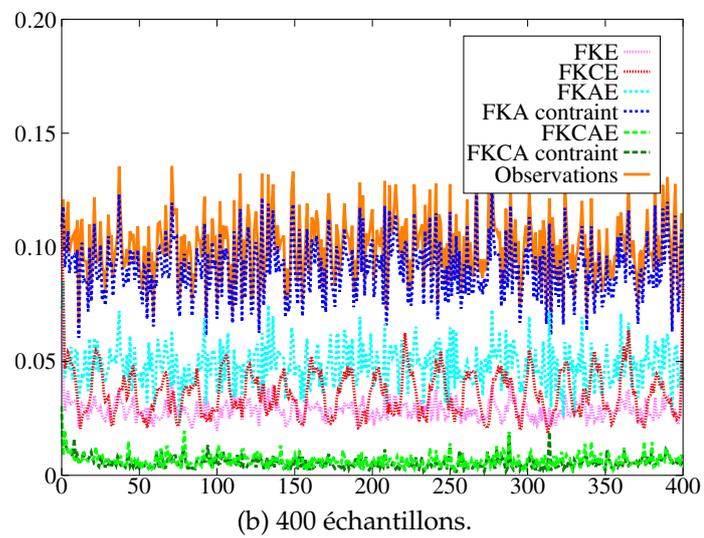
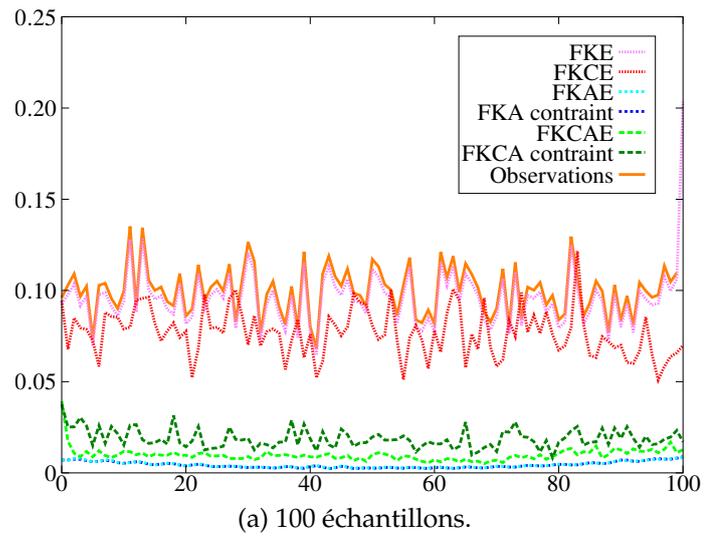


FIGURE 4.7: EQM instantanée pour les différents filtres (section 2.1).

TABLE 4.4: Paramètres de l'expérience de la section 2.2. Lignes 1 à 3 : valeurs d'initialisation de l'algorithme EM ; Lignes 4 à 6 : estimation moyenne des paramètres par les différents filtres pour un signal d'une taille de 1 000 échantillons (100 itérations EM).

	\hat{t}_0	Q_0	F	Q
$\theta_{FK}^{(0)}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
$\theta_{FKC\ constraint1}^{(0)}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0.5 & 0.5 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
$\theta_{FKC\ constraint2}^{(0)}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.999 & 0 \\ 0.5 & 0.5 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
$\theta_{FK}^{(100)}$	/	/	$\begin{bmatrix} 0.998 & \mathbf{0} \\ \mathbf{1} & \mathbf{0} \end{bmatrix}$	$\begin{bmatrix} 0.001 & \mathbf{0} \\ \mathbf{0} & 0.098 \end{bmatrix}$
$\theta_{FKC\ constraint1}^{(100)}$	/	/	$\begin{bmatrix} 0.997 & \mathbf{0} \\ \mathbf{0.5} & \mathbf{0.5} \end{bmatrix}$	$\begin{bmatrix} 0.003 & \mathbf{0} \\ \mathbf{0} & 0.138 \end{bmatrix}$
$\theta_{FKC\ constraint2}^{(100)}$	/	/	$\begin{bmatrix} \mathbf{0.999} & \mathbf{0} \\ \mathbf{0.5} & \mathbf{0.5} \end{bmatrix}$	$\begin{bmatrix} 0.002 & \mathbf{0} \\ \mathbf{0} & 0.127 \end{bmatrix}$

tés semblent plus stables que les filtres contraints pour un grand nombre d'échantillons.

2.2 Signal asservi

Dans cette seconde expérience, nous avons considéré le problème de l'estimation d'un signal dont l'observation est solution de l'équation différentielle

$$\frac{\partial \mathbf{y}}{\partial t}(t) + \mathbf{D}^{y,y} \mathbf{y}(t) = \mathbf{x}(t) + \boldsymbol{\omega}^y(t). \quad (4.1)$$

Ce modèle est analogue à un capteur qui ne réagit pas instannément à la quantité observée (capteur de force). Si nous imposons $\mathbf{D}^{y,y} = \mathbf{O}$, alors nous retrouvons un système de mesure parfait. Par contre, plus le terme $\mathbf{D}^{y,y}$ est important, moins les états cachés ont d'influence sur les observations.

Nous avons choisi l'estimation d'un signal pris entre les dates $t = 0$ et $t = 1$ avec différents échantillonnages. Nous notons N , le nombre d'échantillons. Dans le cadre de ces hypothèses, l'équation (4.1) est compatible avec un filtre de Kalman couple dont la matrice de transition est de la forme

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}^{x,x} & \mathbf{O} \\ \mathbf{I} - \boldsymbol{\nu} & \boldsymbol{\nu} \end{bmatrix} \quad (4.2)$$

avec

$$\boldsymbol{\nu} = N [\mathbf{I} + N \mathbf{D}^{y,y}]^{-1} \mathbf{D}^{y,y}.$$

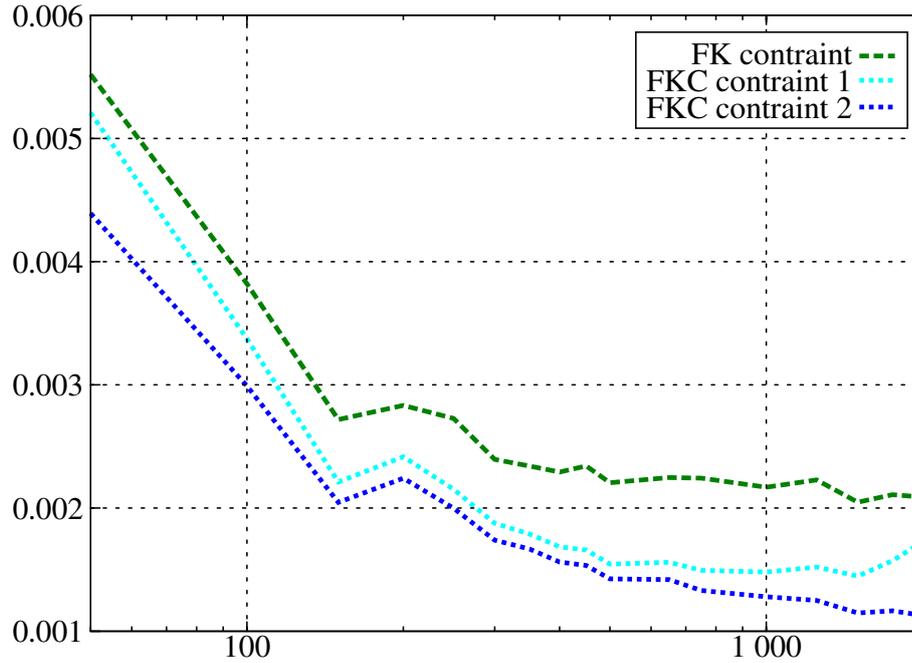


FIGURE 4.8: Évolution de l'EQM en fonction du nombre d'échantillons (section 2.2).

Il faut noter que le terme ν dépend du nombre d'observations N . Afin de rester dans un cadre simple, nous avons pris le signal exponentiel à restaurer $x(t) = \exp(-[D^{x,x}]^{-1}t)$. Avec ce signal, l'observation "parfaite" $y(t)$ s'écrit

$$y(t) = A \exp(-[D^{x,x}]^{-1}t) + (1 - A) \exp(-[D^{y,y}]^{-1}t)$$

avec

$$A = \frac{1}{1 + [D^{x,x}]^{-1} \times D^{y,y}}.$$

Remarques:

- Le signal $x(t)$ est compatible avec les filtres de Kalman ou de Kalman couple dont l'équation d'état est de la forme

$$\delta = \mathbf{F}^{x,x} = [1 + N/D^{x,x}]^{-1} \times N/D^{x,x}. \quad (4.3)$$

- Quand le terme $D^{y,y}$ est nul, nous retrouvons le modèle du filtre de Kalman avec le jeu de paramètres

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

$$\mathbf{Q} = \begin{bmatrix} Q^{x,x} & 0 \\ 0 & Q^{y,y} \end{bmatrix}$$

- Quand le nombre d'échantillons devient trop important (>2000), alors les états cachés ont une influence tellement faible sur les observations que l'algorithme EM du filtre de Kalman couple reste piégé sur une solution sous-optimale.

Nous avons ensuite dégradé l'observation "parfaite" par l'ajout d'un bruit additif gaussien centré de variance 0.1. Nous avons ensuite restauré le signal par

1. un filtre de Kalman contraint (FK contraint) où $F^{y,x} = 1$,
2. un filtre de Kalman couple contraint (FKC contraint 1) où $Q^{x,y} = 0$, $F^{y,t} = [I - \nu \quad \nu]$ et $F^{x,y} = 0$.
3. un filtre de Kalman couple contraint (FKC contraint 2) où $Q^{x,y} = 0$, $F^{y,t} = [I - \nu \quad \nu]$ et $F^{x,t} = [\delta, \quad O]$

Les paramètres d'initialisation des filtres sont reportés dans les lignes 1 à 3 du tableau 4.4. Les estimations moyennes sont données dans les lignes 4 à 6 pour les signaux de 1 000 points. Un exemple de signal généré est représenté sur la figure 4.9a et ses restaurations sur la figure 4.9b. L'EQM en fonction du nombre d'échantillons est reportée sur la figure 4.8.

Il convient d'analyser séparément ces courbes

- Concernant l'exemple de restauration, on remarque que les performances des filtres non-supervisés sont assez proches.
- Concernant l'estimation des paramètres, on remarque que, pour les deux filtres de Kalman couple, le bruit de modèle estimé est faible. On peut aussi noter que tous les filtres donnent presque la même estimation de $F^{x,x}$.
- Concernant les courbes d'EQM, on remarque que le filtre de Kalman couple contraint 2 donne les meilleurs résultats. Les résultats du FKC contraint 1 sont assez proches du FKC contraint 2, ce qui n'est pas le cas du FK contraint.

Cette expérience montre une conclusion intéressante : le filtre de Kalman couple permet de réaliser l'estimation dans des modèles où l'observation n'est pas l'état caché dégradé par un bruit additif. Ici, le filtre de Kalman couple a permis de restaurer un signal qui est une réponse retardée des états cachés.

3 Suivi de la pupille dans une séquence vidéo

Pour illustrer l'intérêt des filtres contraints, nous avons choisi le problème du suivi de la pupille dans une vidéo. Des travaux précédents [40] proposent d'utiliser un filtre de Kalman avec un état augmenté pour estimer à la fois, position et vitesse de la pupille. Nous souhaitons donc affiner le suivi de la pupille grâce à un algorithme d'apprentissage des paramètres. Les positions $x^{pupille}$, $y^{pupille}$ de la pupille sont obtenues par l'algorithme de segmentation présenté dans le chapitre suivant. Nous mesurons les performances de l'estimation par l'EQM sur la prédiction qui est principalement utilisée dans les algorithmes de suivi. Nous définissons les observations

$$\mathbf{y}_n = \begin{bmatrix} x_n^{pupille} \\ y_n^{pupille} \end{bmatrix}$$

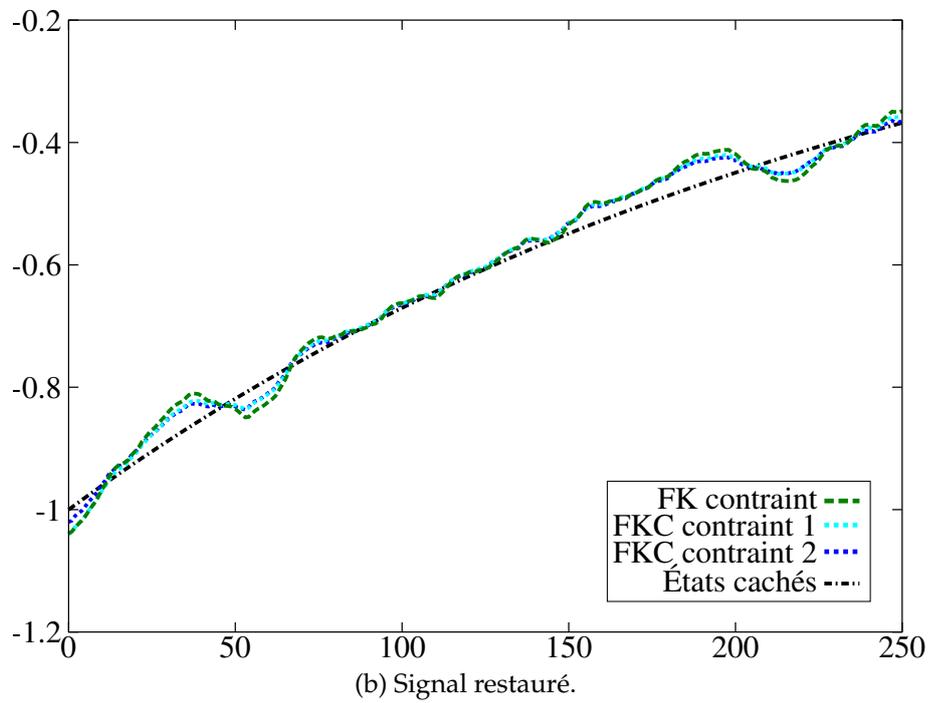
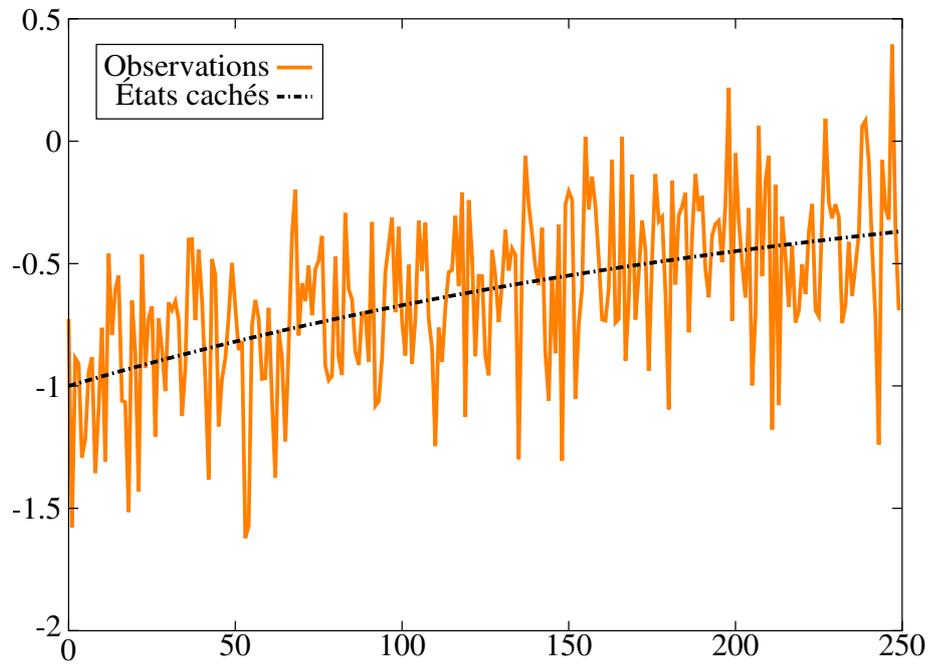


FIGURE 4.9: Signal simulé et restaurations (section 2.2).

TABLE 4.5: Valeurs d'initialisation des paramètres de l'expérience de la section 3

	\hat{t}_0	Q_0	F	Q
$\theta^{(0)}$	$O_{6,1}$	$50I_6$	$\begin{bmatrix} I_2 & I_2 & O_{2,2} \\ O_{2,2} & O_{2,2} & O_{2,2} \\ I_2 & O_{2,2} & O_{2,2} \end{bmatrix}$	I_6

et les états augmentés

$$\mathbf{x}_n = \begin{bmatrix} \bar{x}_n^{pupille} \\ \bar{y}_n^{pupille} \\ r_n^1 \\ r_n^2 \end{bmatrix} \left. \begin{array}{l} \} z_n \\ \} r_n \end{array} \right\},$$

où z_n et r_n correspondent respectivement à l'état caché (position de la pupille) et aux processus latents (dérivée de la position).

Nous avons testé les algorithmes d'apprentissage sur 15 vidéos issues de 3 individus différents et de la base de données MBGC. Les paramètres ont été appris par

1. un filtre de Kalman modélisant un mouvement brownien (FKMB), *c.a.d* que
 - nous avons imposé $F^{x,x} = \begin{bmatrix} I_2 & T_E I_2 \\ O_{2,2} & O_{2,2} \end{bmatrix}$ où T_E représente la fréquence d'échantillonnage,
 - nous avons imposé $F^{y,x} = [I_2, O_{2,2}]$,
 - $Q^{x,x}$ et $Q^{y,y}$ restent libres.
2. un filtre de Kalman augmenté contraint (FKA contraint 1) où
 - nous avons estimé $F^{x,t}$ et $F^{y,t}$ séparément et imposé $Q^{x,y} = O_{4,2}$,
 - nous avons imposé $F^{y,t} = [I_2, O_{2,2}, O_{2,2}]$,
 - $F^{x,t}$, $Q^{x,x}$ et $Q^{y,y}$ restent libres.
3. un filtre de Kalman augmenté contraint (FKA contraint 2) où
 - nous avons estimé $F^{x,t}$ et $F^{y,t}$ séparément et imposé $Q^{x,y} = O_{4,2}$,
 - nous avons imposé $F^{y,t} = [I_2, O_{2,2}, O_{2,2}]$,
 - $F^{x,t}$, $Q^{x,x}$ et $Q^{y,y}$ restent libres.

Remarque: Le filtre de Kalman couple ne s'est pas avéré performant pour restaurer ce type de données, il n'a donc pas été utilisé lors de cette expérience.

Les paramètres d'initialisation des filtres non-supervisés sont donnés dans la ligne 1 du tableau 4.5. Les estimations et les erreurs de prédiction après 1000 itérations EM sont reportées dans les tableaux 4.6 pour le premier filtre, 4.7 pour le second filtre et 4.8 pour le dernier filtre.

Il convient d'analyser ces résultats séparément :

- Concernant l'estimation des paramètres par le premier filtre, on remarque une grande variabilité de la matrice de covariance. De plus, les covariances

TABLE 4.6: Paramètres estimés par le premier filtre non-supervisé pour les différentes vidéos et EQM de prédiction (section 3).

Vidéo	$Q^{z,z}$	$Q^{r,r}$	$Q^{y,y}$	EQM
02463d1980	$\begin{bmatrix} 2.18 & 0.27 \\ 0.27 & 1.69 \end{bmatrix}$	$\begin{bmatrix} 4.31 & 0.24 \\ 0.24 & 2.55 \end{bmatrix}$	$\begin{bmatrix} 0.01 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	9392.20
02463d1981	$\begin{bmatrix} 13.72 & -4.29 \\ -4.29 & 1.99 \end{bmatrix}$	$\begin{bmatrix} 4.09 & 6.92 \\ 6.92 & 14.53 \end{bmatrix}$	$\begin{bmatrix} 25.77 & -11.90 \\ -11.90 & 5.54 \end{bmatrix}$	1861.69
02463d1982	$\begin{bmatrix} 126.76 & 4.87 \\ 4.87 & 1.81 \end{bmatrix}$	$\begin{bmatrix} 0.80 & -0.14 \\ -0.14 & 4.20 \end{bmatrix}$	$\begin{bmatrix} 395.57 & 13.02 \\ 13.02 & 0.43 \end{bmatrix}$	8811.61
02463d1983	$\begin{bmatrix} 3.75 & -0.45 \\ -0.45 & 2.95 \end{bmatrix}$	$\begin{bmatrix} 11.98 & -1.11 \\ -1.11 & 8.99 \end{bmatrix}$	$\begin{bmatrix} 0.01 & -0.00 \\ -0.00 & 0.01 \end{bmatrix}$	1805.03
02463d1984	$\begin{bmatrix} 384.17 & 25.54 \\ 25.54 & 71.52 \end{bmatrix}$	$\begin{bmatrix} 2.74 & -0.23 \\ -0.23 & 2.66 \end{bmatrix}$	$\begin{bmatrix} 953.22 & 96.77 \\ 96.77 & 57.98 \end{bmatrix}$	19808.95
04327d1357	$\begin{bmatrix} 9.52 & -20.94 \\ -20.94 & 54.43 \end{bmatrix}$	$\begin{bmatrix} 10.78 & 1.52 \\ 1.52 & 0.74 \end{bmatrix}$	$\begin{bmatrix} 17.83 & -58.55 \\ -58.55 & 192.36 \end{bmatrix}$	4629.67
04327d1358	$\begin{bmatrix} 43.42 & 6.12 \\ 6.12 & 9.17 \end{bmatrix}$	$\begin{bmatrix} 0.65 & 0.03 \\ 0.03 & 0.41 \end{bmatrix}$	$\begin{bmatrix} 86.13 & 10.74 \\ 10.74 & 1.35 \end{bmatrix}$	5169.98
04327d1359	$\begin{bmatrix} 55.13 & 22.58 \\ 22.58 & 24.18 \end{bmatrix}$	$\begin{bmatrix} 7.96 & -2.52 \\ -2.52 & 1.36 \end{bmatrix}$	$\begin{bmatrix} 64.56 & 105.20 \\ 105.20 & 329.39 \end{bmatrix}$	3711.57
04327d1360	$\begin{bmatrix} 1.17 & 0.15 \\ 0.15 & 1.94 \end{bmatrix}$	$\begin{bmatrix} 2.44 & 1.27 \\ 1.27 & 7.90 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.01 \end{bmatrix}$	2403.09
04327d1361	$\begin{bmatrix} 5.51 & -4.54 \\ -4.54 & 9.77 \end{bmatrix}$	$\begin{bmatrix} 13.89 & 3.18 \\ 3.18 & 4.50 \end{bmatrix}$	$\begin{bmatrix} 0.05 & -0.01 \\ -0.01 & 0.02 \end{bmatrix}$	4984.76
04394d1613	$\begin{bmatrix} 5.61 & 0.20 \\ 0.20 & 2.65 \end{bmatrix}$	$\begin{bmatrix} 23.19 & 1.06 \\ 1.06 & 4.45 \end{bmatrix}$	$\begin{bmatrix} 0.02 & 0.00 \\ 0.00 & 0.01 \end{bmatrix}$	1869.89
04394v1420	$\begin{bmatrix} 2.25 & -0.16 \\ -0.16 & 7.35 \end{bmatrix}$	$\begin{bmatrix} 14.97 & 1.19 \\ 1.19 & 3.42 \end{bmatrix}$	$\begin{bmatrix} 0.01 & 0.00 \\ 0.00 & 0.01 \end{bmatrix}$	2676.12
04394v1421	$\begin{bmatrix} 3.52 & 0.07 \\ 0.07 & 1.02 \end{bmatrix}$	$\begin{bmatrix} 12.72 & -0.22 \\ -0.22 & 1.90 \end{bmatrix}$	$\begin{bmatrix} 0.05 & -0.00 \\ -0.00 & 0.00 \end{bmatrix}$	1434.23
04394v1422	$\begin{bmatrix} 0.86 & 0.01 \\ 0.01 & 0.54 \end{bmatrix}$	$\begin{bmatrix} 0.87 & -0.01 \\ -0.01 & 0.59 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	1757.13
04394v1423	$\begin{bmatrix} 196.06 & -1.62 \\ -1.62 & 0.56 \end{bmatrix}$	$\begin{bmatrix} 1.16 & 0.15 \\ 0.15 & 1.24 \end{bmatrix}$	$\begin{bmatrix} 530.00 & -10.83 \\ -10.83 & 0.22 \end{bmatrix}$	3628.49

TABLE 4.7: Paramètres estimés par le second filtre non-supervisé pour les différentes vidéos et EQM de prédiction (section 3).

Vidéo	$F^{r,z}$	$F^{r,r}$	$Q^{z,z}$	$Q^{r,r}$	$Q^{y,y}$	EQM
02463d1980	$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.83 & -0.04 \\ 0.00 & 0.84 \end{bmatrix}$	$\begin{bmatrix} 0.01 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 2.03 & 0.29 \\ 0.29 & 1.16 \end{bmatrix}$	$\begin{bmatrix} 0.05 & 0.02 \\ 0.02 & 0.01 \end{bmatrix}$	9367.48
02463d1981	$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.34 & -0.01 \\ 0.60 & 0.71 \end{bmatrix}$	$\begin{bmatrix} 16.54 & -1.18 \\ -1.18 & 0.09 \end{bmatrix}$	$\begin{bmatrix} 5.39 & 2.47 \\ 2.47 & 1.17 \end{bmatrix}$	$\begin{bmatrix} 20.51 & -1.89 \\ -1.89 & 0.18 \end{bmatrix}$	1822.06
02463d1982	$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.82 & -0.06 \\ 0.02 & 0.77 \end{bmatrix}$	$\begin{bmatrix} 0.03 & 0.00 \\ 0.00 & 0.05 \end{bmatrix}$	$\begin{bmatrix} 2.37 & 0.37 \\ 0.37 & 2.15 \end{bmatrix}$	$\begin{bmatrix} 0.75 & -0.04 \\ -0.04 & 0.13 \end{bmatrix}$	8446.61
02463d1983	$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.80 & -0.05 \\ 0.01 & 0.83 \end{bmatrix}$	$\begin{bmatrix} 0.04 & -0.00 \\ -0.00 & 0.03 \end{bmatrix}$	$\begin{bmatrix} 5.44 & -0.05 \\ -0.05 & 3.62 \end{bmatrix}$	$\begin{bmatrix} 0.01 & -0.00 \\ -0.00 & 0.01 \end{bmatrix}$	1733.60
02463d1984	$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.92 & 0.18 \\ -0.17 & 0.80 \end{bmatrix}$	$\begin{bmatrix} 324.60 & 16.66 \\ 16.66 & 1.04 \end{bmatrix}$	$\begin{bmatrix} 3.01 & -5.82 \\ -5.82 & 12.55 \end{bmatrix}$	$\begin{bmatrix} 966.31 & 84.05 \\ 84.05 & 11.98 \end{bmatrix}$	19480.28
04327d1357	$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.88 & -0.10 \\ -0.06 & 0.88 \end{bmatrix}$	$\begin{bmatrix} 0.01 & 0.00 \\ 0.00 & 0.01 \end{bmatrix}$	$\begin{bmatrix} 4.42 & -1.58 \\ -1.58 & 5.27 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	4338.61
04327d1358	$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.72 & -0.14 \\ -0.02 & 0.68 \end{bmatrix}$	$\begin{bmatrix} 0.08 & 0.05 \\ 0.05 & 0.05 \end{bmatrix}$	$\begin{bmatrix} 8.93 & 0.52 \\ 0.52 & 3.47 \end{bmatrix}$	$\begin{bmatrix} 12.76 & 9.35 \\ 9.35 & 6.85 \end{bmatrix}$	5110.55
04327d1359	$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.55 & 0.59 \\ -0.29 & 0.51 \end{bmatrix}$	$\begin{bmatrix} 9.09 & 13.72 \\ 13.72 & 21.75 \end{bmatrix}$	$\begin{bmatrix} 18.62 & 6.91 \\ 6.91 & 2.71 \end{bmatrix}$	$\begin{bmatrix} 66.24 & 114.34 \\ 114.34 & 307.31 \end{bmatrix}$	3667.81
04327d1360	$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.58 & 0.05 \\ -0.08 & 0.71 \end{bmatrix}$	$\begin{bmatrix} 0.03 & 0.00 \\ 0.00 & 0.21 \end{bmatrix}$	$\begin{bmatrix} 2.27 & 0.81 \\ 0.81 & 4.15 \end{bmatrix}$	$\begin{bmatrix} 0.02 & -0.05 \\ -0.05 & 0.69 \end{bmatrix}$	2385.61
04327d1361	$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.76 & -0.06 \\ 0.00 & 0.69 \end{bmatrix}$	$\begin{bmatrix} 0.24 & 0.06 \\ 0.06 & 0.12 \end{bmatrix}$	$\begin{bmatrix} 5.76 & 0.67 \\ 0.67 & 6.60 \end{bmatrix}$	$\begin{bmatrix} 2.58 & -0.55 \\ -0.55 & 0.35 \end{bmatrix}$	4897.54
04394d1613	$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.93 & -0.33 \\ 0.08 & 0.90 \end{bmatrix}$	$\begin{bmatrix} 0.11 & 0.18 \\ 0.18 & 0.52 \end{bmatrix}$	$\begin{bmatrix} 3.87 & -0.24 \\ -0.24 & 0.82 \end{bmatrix}$	$\begin{bmatrix} 0.57 & 0.18 \\ 0.18 & 0.34 \end{bmatrix}$	1853.89
04394v1420	$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.90 & -0.00 \\ -0.01 & 0.96 \end{bmatrix}$	$\begin{bmatrix} 0.01 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 3.37 & 0.47 \\ 0.47 & 2.18 \end{bmatrix}$	$\begin{bmatrix} 0.00 & -0.00 \\ -0.00 & 0.00 \end{bmatrix}$	2637.17
04394v1421	$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.88 & -0.02 \\ -0.02 & 0.78 \end{bmatrix}$	$\begin{bmatrix} 0.03 & -0.01 \\ -0.01 & 0.02 \end{bmatrix}$	$\begin{bmatrix} 3.47 & 0.35 \\ 0.35 & 0.92 \end{bmatrix}$	$\begin{bmatrix} 0.62 & -0.31 \\ -0.31 & 0.21 \end{bmatrix}$	1393.13
04394v1422	$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.49 & -0.14 \\ -0.06 & 0.74 \end{bmatrix}$	$\begin{bmatrix} 0.02 & 0.00 \\ 0.00 & 0.02 \end{bmatrix}$	$\begin{bmatrix} 1.14 & 0.15 \\ 0.15 & 0.47 \end{bmatrix}$	$\begin{bmatrix} 0.03 & -0.01 \\ -0.01 & 0.03 \end{bmatrix}$	1754.03
04394v1423	$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.71 & -0.26 \\ -0.03 & 0.74 \end{bmatrix}$	$\begin{bmatrix} 42.65 & 1.83 \\ 1.83 & 0.09 \end{bmatrix}$	$\begin{bmatrix} 7.68 & 1.87 \\ 1.87 & 0.46 \end{bmatrix}$	$\begin{bmatrix} 523.50 & -32.20 \\ -32.20 & 1.99 \end{bmatrix}$	3574.08

TABLE 4.8: Paramètres estimés par le troisième filtre non-supervisé pour les différentes vidéos et EQM de prédiction (section 3).

Vidéo	$F^{r,r}$	$Q^{z,z}$	$Q^{r,r}$	$Q^{y,y}$	EQM
02463d1980	$\begin{bmatrix} 0.85 & 0.00 \\ 0.00 & 0.85 \end{bmatrix}$	$\begin{bmatrix} 0.05 & 0.02 \\ 0.02 & 0.01 \end{bmatrix}$	$\begin{bmatrix} 1.40 & 0.00 \\ 0.00 & 1.40 \end{bmatrix}$	$\begin{bmatrix} 0.11 & 0.04 \\ 0.04 & 0.02 \end{bmatrix}$	9368.22
02463d1981	$\begin{bmatrix} 0.86 & 0.00 \\ 0.00 & 0.86 \end{bmatrix}$	$\begin{bmatrix} 0.11 & -0.03 \\ -0.03 & 0.01 \end{bmatrix}$	$\begin{bmatrix} 3.41 & 0.00 \\ 0.00 & 3.41 \end{bmatrix}$	$\begin{bmatrix} 31.38 & -8.21 \\ -8.21 & 2.15 \end{bmatrix}$	1767.68
02463d1982	$\begin{bmatrix} 0.80 & 0.00 \\ 0.00 & 0.80 \end{bmatrix}$	$\begin{bmatrix} 0.04 & 0.02 \\ 0.02 & 0.05 \end{bmatrix}$	$\begin{bmatrix} 2.27 & 0.00 \\ 0.00 & 2.27 \end{bmatrix}$	$\begin{bmatrix} 0.77 & 0.01 \\ 0.01 & 0.12 \end{bmatrix}$	8446.58
02463d1983	$\begin{bmatrix} 0.82 & 0.00 \\ 0.00 & 0.82 \end{bmatrix}$	$\begin{bmatrix} 0.32 & -0.01 \\ -0.01 & 0.01 \end{bmatrix}$	$\begin{bmatrix} 4.28 & 0.00 \\ 0.00 & 4.28 \end{bmatrix}$	$\begin{bmatrix} 0.03 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	1730.59
02463d1984	$\begin{bmatrix} 0.79 & 0.00 \\ 0.00 & 0.79 \end{bmatrix}$	$\begin{bmatrix} 275.68 & 14.62 \\ 14.62 & 0.94 \end{bmatrix}$	$\begin{bmatrix} 15.57 & 0.00 \\ 0.00 & 15.57 \end{bmatrix}$	$\begin{bmatrix} 961.98 & 81.02 \\ 81.02 & 10.80 \end{bmatrix}$	19469.10
04327d1357	$\begin{bmatrix} 0.89 & 0.00 \\ 0.00 & 0.89 \end{bmatrix}$	$\begin{bmatrix} 0.01 & -0.00 \\ -0.00 & 0.01 \end{bmatrix}$	$\begin{bmatrix} 4.95 & 0.00 \\ 0.00 & 4.95 \end{bmatrix}$	$\begin{bmatrix} 0.00 & -0.00 \\ -0.00 & 0.00 \end{bmatrix}$	4345.80
04327d1358	$\begin{bmatrix} 0.72 & 0.00 \\ 0.00 & 0.72 \end{bmatrix}$	$\begin{bmatrix} 0.25 & 0.12 \\ 0.12 & 0.07 \end{bmatrix}$	$\begin{bmatrix} 5.45 & 0.00 \\ 0.00 & 5.45 \end{bmatrix}$	$\begin{bmatrix} 17.37 & 11.17 \\ 11.17 & 7.19 \end{bmatrix}$	5087.66
04327d1359	$\begin{bmatrix} 0.65 & 0.00 \\ 0.00 & 0.65 \end{bmatrix}$	$\begin{bmatrix} 50.16 & 17.75 \\ 17.75 & 7.08 \end{bmatrix}$	$\begin{bmatrix} 2.46 & 0.00 \\ 0.00 & 2.46 \end{bmatrix}$	$\begin{bmatrix} 60.30 & 105.55 \\ 105.55 & 347.66 \end{bmatrix}$	3675.18
04327d1360	$\begin{bmatrix} 0.70 & 0.00 \\ 0.00 & 0.70 \end{bmatrix}$	$\begin{bmatrix} 0.10 & 0.49 \\ 0.49 & 2.95 \end{bmatrix}$	$\begin{bmatrix} 2.33 & 0.00 \\ 0.00 & 2.33 \end{bmatrix}$	$\begin{bmatrix} 0.03 & -0.06 \\ -0.06 & 0.18 \end{bmatrix}$	2385.85
04327d1361	$\begin{bmatrix} 0.72 & 0.00 \\ 0.00 & 0.72 \end{bmatrix}$	$\begin{bmatrix} 0.28 & 0.13 \\ 0.13 & 0.17 \end{bmatrix}$	$\begin{bmatrix} 6.31 & 0.00 \\ 0.00 & 6.31 \end{bmatrix}$	$\begin{bmatrix} 2.45 & -0.45 \\ -0.45 & 0.39 \end{bmatrix}$	4901.69
04394d1613	$\begin{bmatrix} 0.89 & 0.00 \\ 0.00 & 0.89 \end{bmatrix}$	$\begin{bmatrix} 0.10 & 0.02 \\ 0.02 & 0.03 \end{bmatrix}$	$\begin{bmatrix} 3.56 & 0.00 \\ 0.00 & 3.56 \end{bmatrix}$	$\begin{bmatrix} 0.58 & 0.18 \\ 0.18 & 0.17 \end{bmatrix}$	1760.62
04394v1420	$\begin{bmatrix} 0.93 & 0.00 \\ 0.00 & 0.93 \end{bmatrix}$	$\begin{bmatrix} 0.09 & 0.02 \\ 0.02 & 0.01 \end{bmatrix}$	$\begin{bmatrix} 2.70 & 0.00 \\ 0.00 & 2.70 \end{bmatrix}$	$\begin{bmatrix} 0.01 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	2632.55
04394v1421	$\begin{bmatrix} 0.86 & 0.00 \\ 0.00 & 0.86 \end{bmatrix}$	$\begin{bmatrix} 0.09 & -0.01 \\ -0.01 & 0.01 \end{bmatrix}$	$\begin{bmatrix} 2.06 & 0.00 \\ 0.00 & 2.06 \end{bmatrix}$	$\begin{bmatrix} 0.86 & -0.28 \\ -0.28 & 0.11 \end{bmatrix}$	1392.62
04394v1422	$\begin{bmatrix} 0.62 & 0.00 \\ 0.00 & 0.62 \end{bmatrix}$	$\begin{bmatrix} 0.36 & 0.06 \\ 0.06 & 0.02 \end{bmatrix}$	$\begin{bmatrix} 0.67 & 0.00 \\ 0.00 & 0.67 \end{bmatrix}$	$\begin{bmatrix} 0.03 & 0.00 \\ 0.00 & 0.00 \end{bmatrix}$	1753.85
04394v1423	$\begin{bmatrix} 0.71 & 0.00 \\ 0.00 & 0.71 \end{bmatrix}$	$\begin{bmatrix} 119.51 & 3.34 \\ 3.34 & 0.12 \end{bmatrix}$	$\begin{bmatrix} 0.50 & 0.00 \\ 0.00 & 0.50 \end{bmatrix}$	$\begin{bmatrix} 522.19 & -23.54 \\ -23.54 & 1.06 \end{bmatrix}$	3558.04

estimées sont très importantes ce qui montre une mauvaise adéquation entre le modèle et les données.

- Concernant l'estimation des paramètres par le second filtre, on remarque que les matrices de covariances sont encore très variable. Par contre, les matrices de transition F sont assez proches entre deux vidéos du même individu.
- Concernant l'estimation des paramètres par le dernier filtre, on remarque que la matrice F varie pour les vidéos d'un même individu.
- Concernant les EQM de prédiction, on remarque que le deuxième et troisième filtre donnent à peu près les mêmes résultats qui, de toute façon, restent supérieurs à ceux du premier filtre.

Remarque: L'EQM est fortement déterminée par les images où la segmentation de l'iris a échoué, comme nous le verrons dans le chapitre suivant. En effet, comme l'erreur initiale est très forte, la perte d'une position de l'oeil au cours d'une vidéo augmentera considérablement l'erreur moyenne.

Nous pouvons conclure de cette expérience que la matrice de transition F semble caractéristique de l'individu. Cependant, puisque les paramètres sont liés à l'individu, alors il est impossible de les ré-utiliser pour poursuivre l'oeil dans une vidéo inconnue.

4 Conclusion

L'intérêt principal du filtre de Kalman couple réside dans la possibilité pour l'utilisateur d'affiner les interactions entre observations successives et états cachés. En contre-partie, cela augmente le nombre de paramètres que nous avons à estimer si aucune connaissance a priori est disponible. En pratique, grâce à la physique sous-jacente du modèle, les paramètres sont souvent partiellement connus ou dépendent des autres paramètres.

Dans ce contexte, nous avons étudié le comportement du filtre de Kalman couple sur des données qui ne sont pas adaptées. Dans le cadre de restauration de signaux dégradés par un bruit additif, ce filtre est moins performant que le filtre de Kalman. Par contre, dans une problématique d'asservissement, le filtre de Kalman couple permet d'une part, de débruiter les données, et d'autre part, d'estimer l'équation d'état des états cachés.

Finalement, nous avons appliqué les algorithmes EM contraints pour suivre le mouvement de la pupille dans une vidéo par un filtre de Kalman. Les performances ainsi obtenues ont dépassé celles du modèle classique qui revient à supposer un mouvement brownien pour la pupille. Les résultats ont même montré que les paramètres appris sont caractéristiques de l'individu.

On peut aussi noter que les résultats obtenus n'auraient pas été possible sans l'utilisation d'algorithmes robustes numériquement.

Chapitre 5

Biométrie dynamique de l'iris

Ce chapitre intègre les résultats de poursuite obtenus dans la dernière expérience du chapitre précédent dans une application de biométrie dynamique de l'iris. Il s'agit ici de fusionner les meilleures images d'iris extraites d'une vidéo pour améliorer l'identification d'une personne par rapport à un système plus classique, basé sur des images photographiques.

Le chapitre est organisé en quatre sections. Les principes sur lesquels reposent la biométrie par l'iris sont exposés dans la section 1. Les algorithmes d'enrôlement¹ de l'iris et de fusion d'informations, que nous avons proposés, sont présentés dans la section 2. Toutes les techniques mises en œuvre pour atteindre des traitements temps-réels sont décrites dans la section 3. Les résultats des expériences sur la base de données MBGC (*Multiple Biometric Grand Challenge*) [53] sont finalement reportés et analysés dans la section 4.

1 Principe

La biométrie l'iris est généralement acceptée comme l'une des méthodes les plus fiables pour la reconnaissance d'êtres humains, grâce à la grande variabilité des textures de l'iris et à sa stabilité au cours de la vie [32]. L'iris est un anneau situé au centre de l'œil, cf. Figure 5.1, entre la pupille (zone sombre), la sclère et les paupières. Sa texture contient de nombreuses informations : les crêtes, les sillons, les cryptes, les nævus et les tâches, cf. Figure 5.2.

L'iris possède des qualités indéniables pour la biométrie : chaque individu possède deux iris uniques différents (un par œil), il est difficilement falsifiable ou altérable et finalement, il possède de très nombreux points de comparaisons. Il est donc considéré comme très efficace pour les applications haute-sécurité.

Néanmoins, l'acquisition de l'image d'un iris reste très contraignant pour l'utilisateur. En effet, celui-ci doit se trouver à une distance proche du capteur (environ 30cm) pour permettre, d'une part, une résolution suffisante et, d'autre part, une bonne focalisation de l'image. L'iris peut aussi être occulté par les paupières,

1. L'enrôlement est l'étape qui transforme un œil en un code caractéristique de l'individu

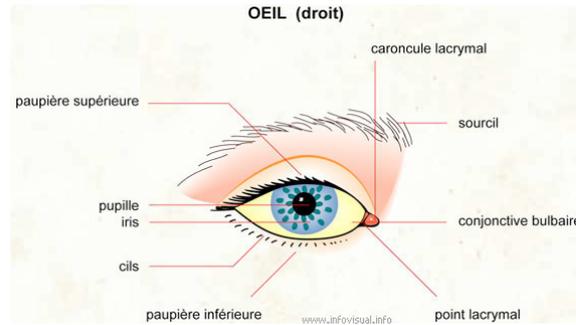


FIGURE 5.1: Schéma de l'œil (source : www.infovisual.com).

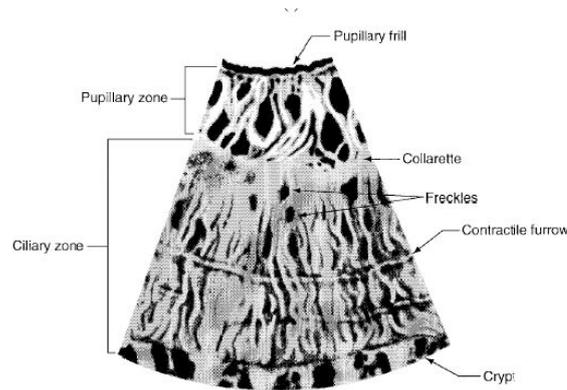


FIGURE 5.2: Texture de l'iris [9].

les cils et les reflets, ce qui peut entraver le processus de reconnaissance.

La biométrie de l'iris transforme cet iris en un code caractéristique de l'individu, appelé iris-code [13].

1.1 Identification et critères d'évaluation

Les systèmes biométriques reposent sur le calcul d'une distance entre deux codes issus de deux individus différents. La distance ainsi calculée est ensuite comparée à seuil de décision λ : si elle est inférieure au seuil, l'individu est authentifié. Ce seuil se règle à partir de la distribution des distances entre imposteurs et authentiques, cf. Figure 5.3. Par exemple, fixer une faible valeur pour ce seuil entrainera un taux de faux rejets (TFR) très important, mais garantira un taux de fausses acceptations (TFA) très faible. Au contraire, fixer une valeur élevée donnera un TFR faible avec un TFA élevé. En biométrie, 3 critères d'évaluation de performances sont basés sur ces taux :

- (i) Courbe de caractéristique de fonctionnement du récepteur (ROC).

Cette courbe trace l'évolution du TFA en fonction du TFR pour chaque valeur du seuil de décision λ . Elle permet de caractériser les champs d'applications du système. En effet, une application haute-sécurité doit imposer un TFA très faible alors qu'une application grand-public devra imposer un TFR faible.

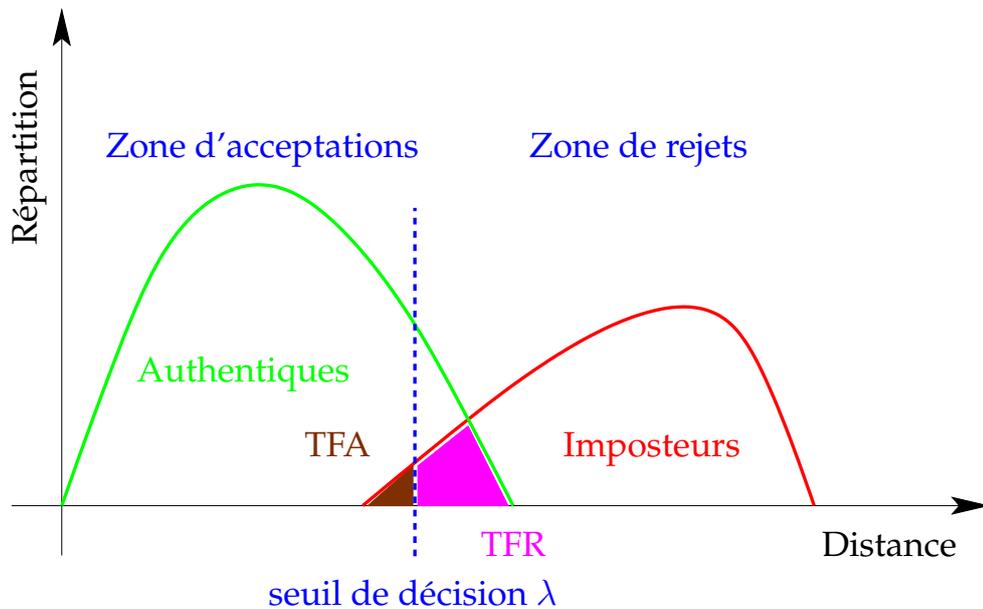


FIGURE 5.3: Répartition des authentiques et des imposteurs.

(ii) Taux d'égalité d'erreur (TEE).

Ce taux correspond au point de la courbe ROC pour lequel $TFA = TFR$. Il est très souvent utilisé pour mesurer les performances des systèmes biométriques.

(iii) Taux de faux rejets pour un taux de fausses acceptations de 10^{-6} ($TFR@TFA=10^{-6}$).

Ce taux correspond au point de la courbe ROC pour lequel $TFA = 10^{-6}$. Il permet de mesurer les performances haute-sécurité du dispositif.

Un autre critère permet aussi de mesurer les performances d'un système biométrique : l'histogramme top-rank. Celui-ci trace les fréquences du classement de la vraie classe de l'individu inconnu selon les distances calculées.

1.2 Modalité vidéo

Les travaux actuels de recherche en biométrie de l'iris s'intéressent maintenant à concevoir des environnements moins contraints et des systèmes capables d'acquérir l'iris en mouvement et à distance. Cela pourra aider, d'une part, à simplifier l'identification de l'utilisateur [33] et, d'autre part, à réduire les problèmes d'occlusion de l'iris. On utilise pour cela non plus des images fixes mais des séquences vidéos provenant de caméras proches infra-rouge.

En plus des problèmes rencontrés par les systèmes d'acquisition d'images photographiques, les systèmes basés sur des vidéos souffrent de problèmes de flou de bouger et d'un manque de résolution. Une première approche consiste à adapter des méthodes de segmentation de l'iris dans le cadre d'images particulièrement dégradées [26, 60]. Or, ces méthodes sont coûteuses en temps, et donc peu compatibles avec la cadence du flux de la vidéo qui est généralement de l'or-

dre de 25 images par seconde (IPS).

L'intérêt principal des systèmes vidéo est d'exploiter les nombreuses vues d'un même iris pour obtenir un iris-code caractéristique de la vidéo, qui serait plus fiable pour la reconnaissance d'iris, que celui obtenu grâce à la meilleure image de la séquence vidéo. Dans la littérature, on peut classer les méthodes dynamiques de fusion en deux grandes familles :

1. La première famille étudie la fusion directe de la texture de l'iris dans le plan polaire pour obtenir une texture super-résolue sur laquelle l'iris-code est calculé [64]. Cette idée est ré-exploitée et évaluée par Nguyen *et al* [50] qui introduit une fusion pondérée par les scores de qualité des différentes textures. Les images doivent être recalées de manière précise pour résoudre d'une part, le problème d'échelle et, d'autre part, les rotations de la tête et de l'oeil. Sinon, d'importantes caractéristiques hautes fréquences de l'iris comme les sillons ou les cryptes sont flouées.
2. La seconde famille de méthodes propose de fusionner les iris-codes obtenus à partir des images d'un même iris pour obtenir un iris-code unique. Dans leur travaux, Hollingsworth *et al* ont montré l'existence de bits fragiles dans l'iris-code, *c.a.d* les bits qui changent régulièrement de valeurs entre deux iris-codes. Ces bits pouvant dégrader les performances de la reconnaissance d'iris, ils proposent dans [30] d'utiliser une distance qui tient compte des bits fragiles mutuels des deux iris-codes comparés. Cette famille de méthodes semble plus prometteuse que la première, puisqu'elle ne souffre pas du mauvais alignement dans le recalage des iris de la première famille.

Dans la suite du chapitre, nous présentons notre algorithme d'enrôlement temps-réel de la vidéo ainsi que son évaluation sur la base de données MBGC.

2 Enrôlement et fusion des iris-codes

Cette section reprend le schéma d'enrôlement de vidéos que nous avons proposé dans [49], synthétisé sur la figure 5.4. La méthode d'enrôlement proposée a été évaluée dans le cadre du concours sur la biométrie de l'iris [22]. Celle-ci se décompose en deux étapes : la première est dédiée à l'enrôlement de l'iris et la seconde est relative à la sélection et la fusion des meilleurs iris-codes depuis une vidéo de l'oeil.

2.1 Enrôlement

L'enrôlement de l'iris est réalisé par une chaîne de traitements assez classique : on cherche d'abord à localiser la pupille, ensuite à segmenter l'iris et enfin à calculer l'iris-code.

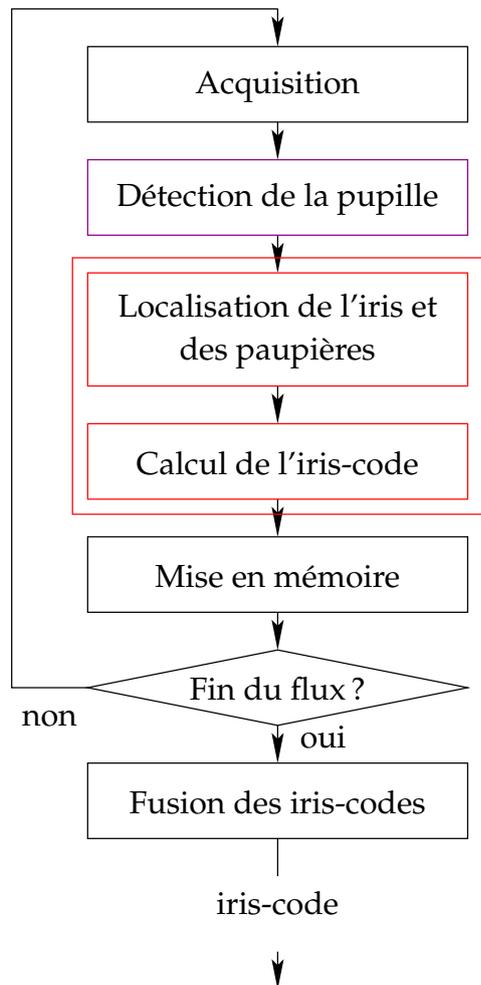


FIGURE 5.4: Schéma d'enrôlement d'une vidéo.

Détection de la pupille

La pupille est généralement considérée comme étant une région sombre et homogène de l'image (autre le spot spéculaire lié à la source lumineuse). Pour réaliser sa segmentation et rejeter les faux candidats, la méthode suivante est appliquée sur une région d'intérêt de l'image (ROI), définie par l'algorithme de suivi de la pupille (voir section 3.1) :

- Comme proposé dans [42], les différents modes de l'histogramme sont détectés en utilisant ses minima locaux, dénotés s_i . De plus, les régions homogènes sont détectées en utilisant une méthode de seuillage adaptatif.
- Pour les 4 minima les plus sombres, s_1, s_2, s_3, s_4 , on réalise les traitements suivants. L'image est binarisée selon s_i . Le masque est alors combiné avec celui des régions homogènes. Dès lors, les objets les plus petits sont supprimés, et les régions étroitement reliées sont séparées par des opérations morphologiques. Chaque composante connexe du masque restante est indexée comme un candidat potentiel de la pupille et leur contour est ré-échantillonné à 64 points. Suivant [23], un algorithme RANSAC est utilisé

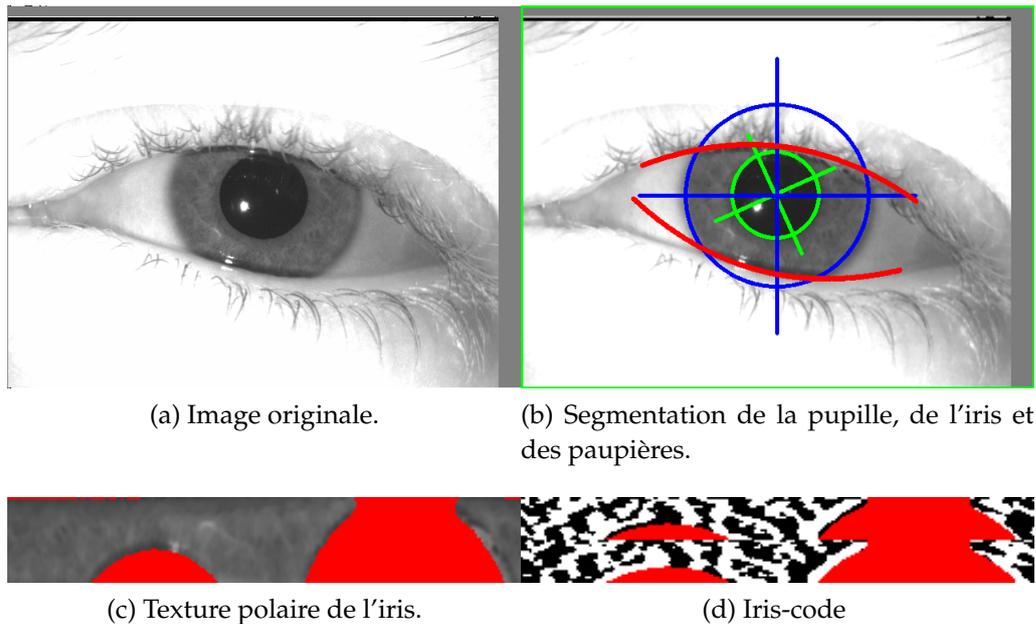


FIGURE 5.5: Illustration des différentes étapes lors de l'enrôlement de l'iris.

pour trouver l'ellipse de paramètres $(x^{pupille}, y^{pupille}, a^{pupille}, b^{pupille}, \theta^{pupille})$ qui lisse le mieux le contour externe à partir des paramètres suivants : des sous-échantillons de 32 points, 50 itérations de l'algorithme, un seuil sur les distances donné par $d_{seuil} = 0.2r^{pupille}^2$ et un seuil de 16 points restants pour que l'ellipse soit considérée comme valide.

- Si elle existe, la meilleure région est sélectionnée comme le dernier candidat potentiel pour la pupille. Si la distance moyenne entre l'ellipse et le contour est trop importante ($d'_{seuil} > 0.1r^{pupille}$), alors la segmentation de la pupille est considérée comme un échec.

Détection de l'iris et des paupières

Pour trouver le cercle de l'iris, nous utilisons l'opérateur intégro-différentiel de J. Daugman [13] dont le résultat est illustré par le cercle bleu dans la figure 5.5b. Dans le but de segmenter les paupières ou les bords de l'iris, nous utilisons un opérateur intégro-différentiel parabolique, décrit dans [34], qui utilise l'équation paramétrique

$$\begin{cases} x(t) = x_i + t \cos \theta + (a(t - x_i)^2 + y_0) \sin \theta \\ y(t) = y_i - t \sin \theta + (a(t - x_i)^2 + y_0) \cos \theta \end{cases} \quad (5.1)$$

où x_i, y_i, r_i sont les coordonnées et le rayon de l'iris, θ varie de -15° à 15° et de 165° à 195° , a évolue de 0 à $(2r_i)^{-1}$ et y_0 évolue de $r^{pupille}$ à r_i pour les paupières

$$2. r^{pupille} = \left[[a^{pupille}]^2 + [b^{pupille}]^2 \right]^{\frac{1}{2}}$$

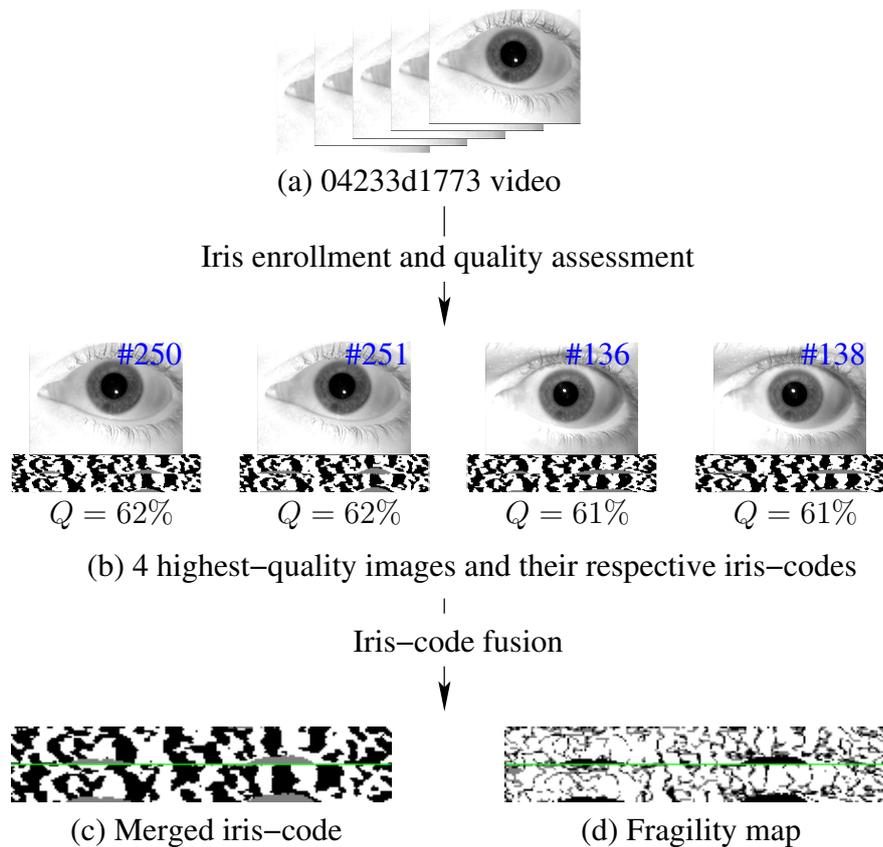


FIGURE 5.6: Schéma de fusion des iris-codes.

supérieure et inférieure. Les cils sont finalement détectés en utilisant le même seuil que celui utilisé pour la détection de la pupille.

Calcul de l'iris code

Finalement, l'iris-code est construit en s'inspirant pour l'essentiel de l'algorithme de Masek [45]. Premièrement, la texture de l'iris (*c.a.d* la région annulaire entre la pupille et la limite extérieure de l'iris) est dépliée dans le domaine polaire avec une correction de l'ellipticité de la pupille. Nous avons choisi une résolution classique de 200 directions angulaires et de 20 rayons *cf.* Figure 5.5c. Ensuite, chaque ligne de la texture polaire est filtrée par des ondelettes 1D dites log-Gabor. En suivant le schéma de Daugman, les données complexes ainsi obtenues sont ensuite quantifiées selon le quadrant dans lequel se trouve la phase pour obtenir l'iris-code. Un exemple d'iris-code est donné sur la figure 5.5d (les régions rouges sont les zones occluses par les paupières, spots ou cils).

2.2 Fusion

Nous décrivons ici comment la fusion des iris-codes extraits à partir des différentes images de la vidéo est réalisée et comment les bits fragiles sont détectés. La technique de fusion utilisée dans ce travail, inspirée de [29], prend plusieurs

images d'iris et leurs iris-codes comme entrées. Elle se décompose en deux étapes principales :

1. **Sélection des images** : Pour chaque image qui passe l'enrôlement avec succès, un score de focalisation Q_{focus} est calculé en utilisant la mesure de Kang et Park [39]. Ce score est ensuite multiplié par le ratio entre la zone utilisable de la texture de l'iris et la surface totale de la texture (cf. [69]), ce qui procure le score Q_{score} . La fusion est réalisée sur les P iris-codes les mieux classés selon Q_{score} . La figure 5.6b montre la sélection des images obtenues sur une vidéo de la base MBGC.
2. **Fusion des iris-codes et détection des bits fragiles** Les iris-codes sont d'abord recalés avec l'iris-code le mieux classé pour tenir compte d'une éventuelle rotation de la tête et la moyenne pondérée par les scores de qualité des iris-codes sélectionnés est ensuite calculée. Nous obtenons un iris-code flou, avec des valeurs flottantes $\hat{I}(r, \theta)$. Finalement, un iris-code classique est obtenu après un seuillage de $\hat{I}(r, \theta)$ par 0.5, et sa carte de fragilité associée est calculée selon

$$f(r, \theta) = 1 - 2 \times \left| \hat{I}(r, \theta) - 0.5 \right| \quad (5.2)$$

L'iris-code et la carte de fragilité pour la vidéo MBGC #04233d1773 sont respectivement représentés sur les figures 5.6c et 5.6d.

3 Enrôlement de l'iris temps-réel

Cette section décrit une version temps-réel de l'algorithme d'enrôlement d'iris décrit dans la section précédente. Trois techniques sont associées pour remplir ce but. Premièrement, un algorithme de poursuite permet de suivre le mouvement de la pupille. Deuxièmement, une sélection intelligente des images permet de réduire le nombre d'images à enrôler. Troisièmement, la parallélisation des différentes tâches permet une réduction du temps de calcul.

3.1 Suivi de la pupille

Cette section est dédiée à l'utilisation du filtre de Kalman couple pour prédire une zone d'intérêt dans laquelle la pupille a une probabilité élevée de se trouver, pour, d'une part, améliorer la détection de la pupille et, d'autre part, réduire son temps de segmentation. Dans notre modèle de poursuite, nous avons considéré comme entrée, la position de la pupille par rapport au centre de l'image. Notons l'observation \mathbf{y}_n à l'instant n ,

$$\mathbf{y}_n = \begin{bmatrix} x_n^{pupille} \\ y_n^{pupille} \end{bmatrix}, \quad (5.3)$$

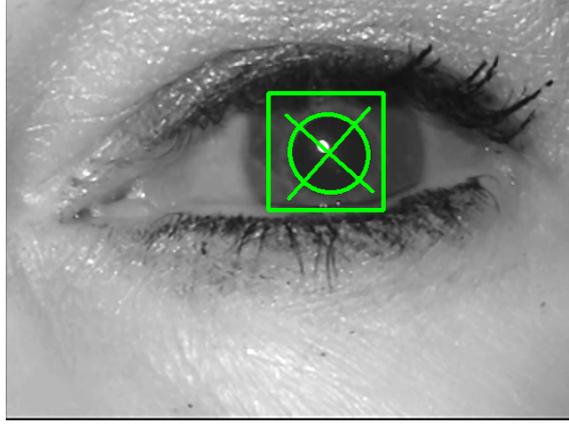


FIGURE 5.7: Exemple de la prédiction de la zone d'intérêt (rectangle).

et un état caché de la forme

$$\mathbf{x}_n = \begin{bmatrix} \mathbf{z}_n \\ \mathbf{r}_n \end{bmatrix}, \quad (5.4)$$

où \mathbf{z}_n désigne les paramètres exacts de la pupille à l'instant n , et \mathbf{r}_n un processus latent. Dès lors, pour imposer la même espérance entre \mathbf{z}_n et \mathbf{y}_n , nous imposons dans le modèle du filtre de Kalman couple, la structure suivante à la matrice de transition :

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}^{z,z} & \mathbf{F}^{z,r} & \mathbf{O}_{3,3} \\ \mathbf{F}^{r,z} & \mathbf{F}^{r,r} & \mathbf{O}_{3,3} \\ \mathbf{I}_3 & \mathbf{O}_{3,n_r} & \mathbf{O}_{3,3} \end{bmatrix}. \quad (5.5)$$

Les paramètres de ce filtre sont soit réglés manuellement, comme pour les filtres de Kalman utilisés pour suivre les mouvements browniens, soit appris par un algorithme EM robuste, détaillé dans le chapitre 1. Pour prédire la zone d'intérêt, nous utilisons la prédiction du filtre de Kalman couple décrite dans le chapitre 1, section 1 et 2 en notant

$$\mathbf{z}_{n|n-1} = \begin{bmatrix} \hat{x}_{n|n-1}^{pupille} \\ \hat{y}_{n|n-1}^{pupille} \end{bmatrix}.$$

La zone d'intérêt, décrite par son centre (x_{ROI}, y_{ROI}) et ses dimensions (w_{ROI}, h_{ROI}) est alors calculée selon

$$\begin{cases} x_{ROI} = \hat{x}_{n|n-1}^{pupille} - \frac{w_{ROI}}{2} \\ y_{ROI} = \hat{y}_{n|n-1}^{pupille} - \frac{h_{ROI}}{2} \\ w_{ROI} = 2\hat{r}_{n-1}^{pupille} + N[\sigma_{n|n-1}(x^{pupille})] \\ h_{ROI} = 2\hat{r}_{n-1}^{pupille} + N[\sigma_{n|n-1}(y^{pupille})] \end{cases}, \quad (5.6)$$

où N est une constante fixée à 4 dans toutes nos expériences.

Lors de l'enrôlement de la base de vidéos, nous avons utilisé un filtre de Kalman supervisé adapté à la poursuite de mouvements browniens *c.a.d* tel que

$$\mathbf{F} = \begin{bmatrix} \mathbf{I}_3 & T_E \mathbf{I}_3 & \mathbf{O}_3 \\ \mathbf{O}_3 & \mathbf{O}_3 & \mathbf{O}_3 \\ \mathbf{O}_3 & \mathbf{O}_3 & \mathbf{O}_3 \end{bmatrix},$$

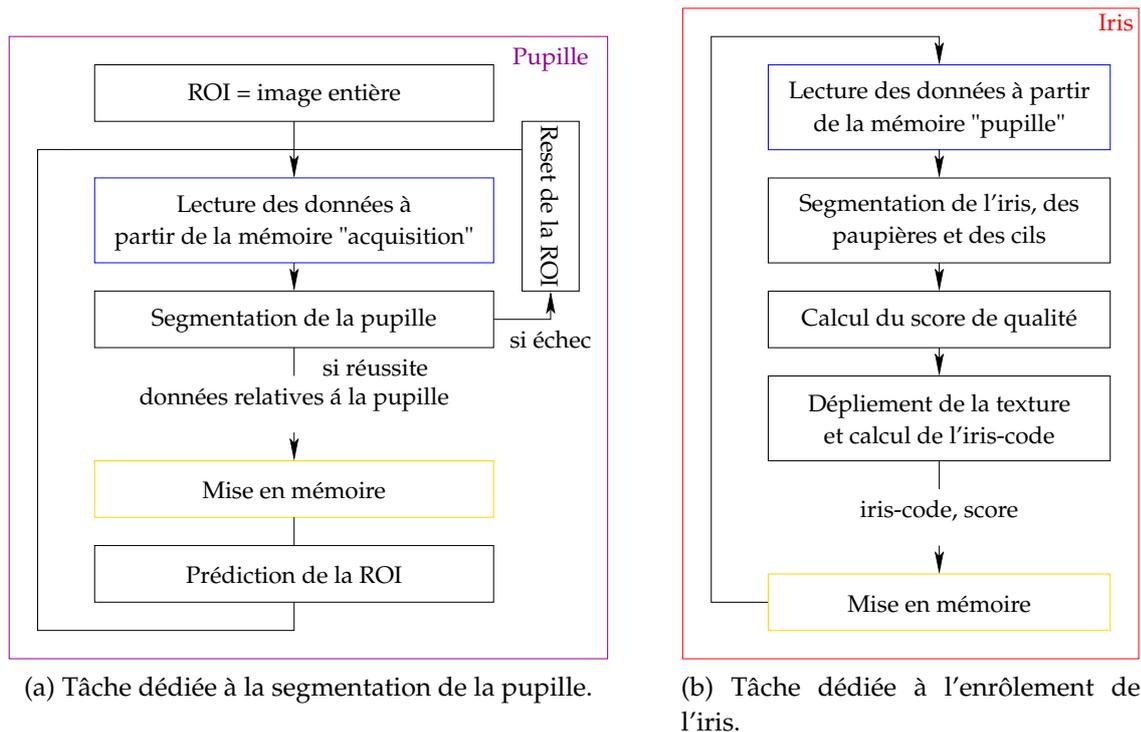


FIGURE 5.8: Organigramme des tâches dédiées à l'enrôlement de l'iris.

où $T_E = 40ms$ est la période entre deux trames.

Cependant, la dernière expérience du chapitre précédent semble montrer que le mouvement de la pupille est caractéristique de l'individu et que le filtre brownien n'est pas forcément le filtre le plus adapté. Or, puisque le mouvement est caractéristique, il est impossible d'utiliser le filtre optimal estimé pour une vidéo sur l'ensemble des vidéos de la base MBGC.

3.2 Sélection à la volée des meilleures images

L'algorithme décrit dans la section 1 peut être parallélisé en 4 tâches :

1. **Acquisition des images** : La première tâche acquiert l'image à partir du flux vidéo, ou d'une caméra infra-rouge, et calcule son score de focalisation Q_{focus} (2.2). Les images acquises sont stockées dans une mémoire tampon pour une exploitation ultérieure dans la seconde tâche. Cette tâche est bridée à 25 images par seconde (IPS).
2. **Segmentation de la pupille** : La seconde tâche, décrite dans la figure 5.8a, récupère l'image mise en mémoire pour segmenter et suivre la pupille. Si la segmentation réussit, l'image et les caractéristiques de la pupille sont stockées pour une exploitation ultérieure dans la troisième tâche.
3. **Enrôlement de l'iris** : La troisième tâche, représentée dans la figure 5.8b, récupère l'image et les données mises en mémoire pour localiser l'iris, les paupières, les cils, les spots, et finalement pour construire l'iris-code. Si le

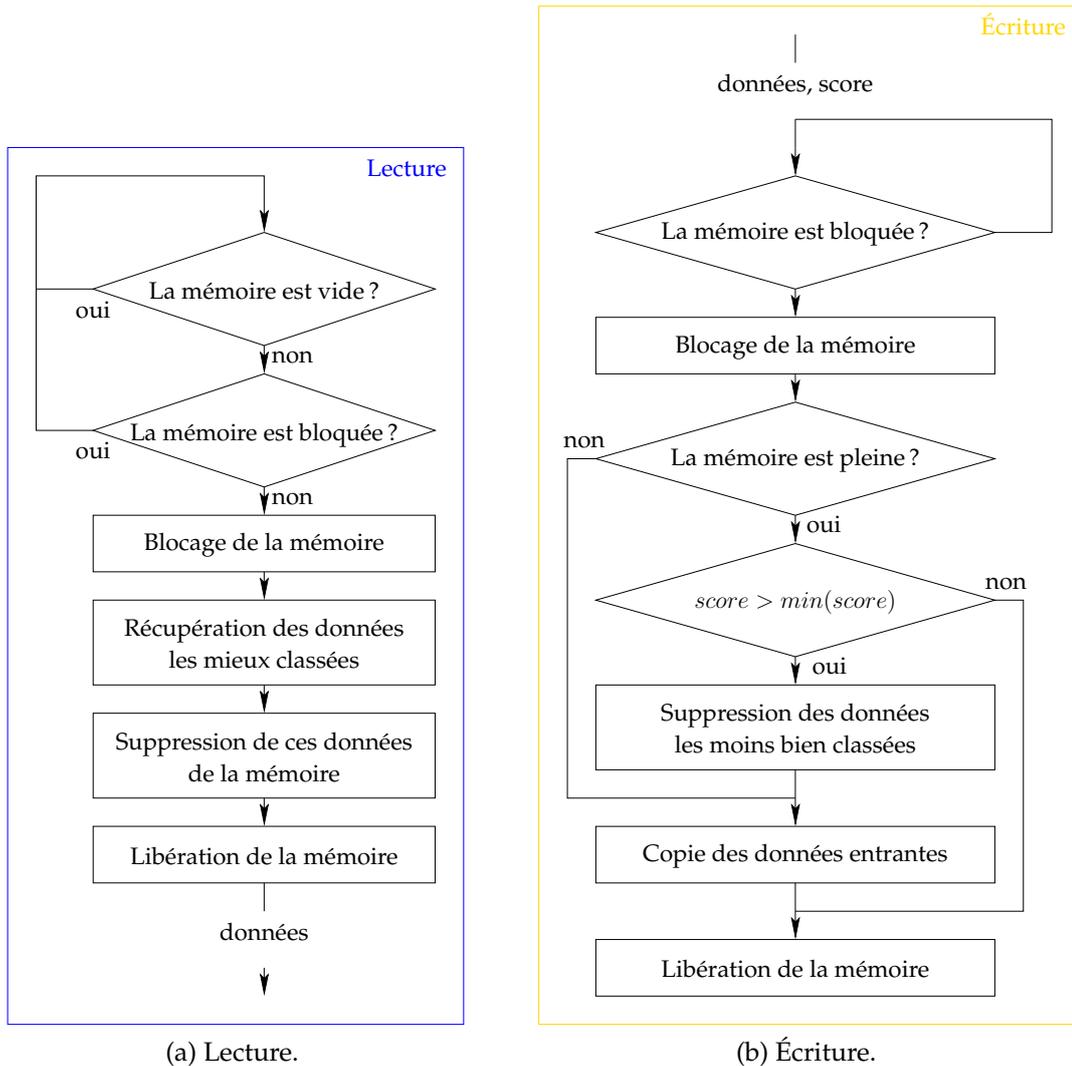


FIGURE 5.9: Différents modes de fonctionnement des mémoires tampons.

score de qualité Q_{score} est suffisant, alors l'iris-code est stocké pour son exploitation par la tâche 4.

- Fusion des iris-codes :** La quatrième tâche réalise la fusion des iris-codes de meilleures qualités dans le but d'obtenir la carte de fragilité 5.6c et l'iris-code 5.6d de la vidéo.

Pour les tâches d'acquisition et de segmentation de la pupille, les tampons de mémoire sont triés au fur à mesure du flux selon le score de focalisation Q_{focus} , tandis que le tampon en sortie de la tâche 3, dédiée à la segmentation de l'iris, est trié selon le score de qualité. Cette stratégie est limitée par la capacité des mémoires tampons : chaque nouvelle entrée est triée selon son score et seulement les meilleures images sont sélectionnées et enrôlées. La figure 5.9 représente les modes de lecture et d'écriture des différentes mémoires-tampons.

Nous avons enregistré les taux d'enrôlement des pupilles et des iris de toute la base de données MBGC (décrite ci-dessous). Quand nous imposons 25 IPS, 75% des pupilles sont détectées et environ 15% des iris sont enrôlés. Si nous utilisons

une méthode séquentielle, seulement 13% des pupilles et des iris sont enrôlés, indépendamment de leur qualité.

4 Évaluation des performances sur la base MBGC

Pour évaluer l'application complète de reconnaissance d'iris, nous avons conduit une série d'expériences sur la base de données MBGC [53]. Cette base de données contient 986 vidéos infra-rouges et 8589 images infra-rouges d'œil issues de 137 individus. Les vidéos et les images ont été segmentées et fusionnées par les algorithmes des sections 1 et 2. La base d'images est utilisée pour enrôler et classer chaque individu tandis que la base de vidéos est utilisée pour tester les performances de reconnaissance de la méthode proposée sur la base d'images.

Le taux de réussite d'enrôlement atteint 99.8% sur les vidéos³ et 100% sur les classes d'iris⁴. À propos de la base d'images, 96.7% des images ont été enrôlées avec succès. Nous avons vérifié que les autres images sont trop sévèrement altérées pour pouvoir être exploitées (occlusion, flou, ...).

Une fois les deux bases de données enrôlées, nous avons comparé les iris-codes de la base des vidéos à ceux de la base des images en utilisant la distance de Hamming⁵. Les résultats de ces expériences sont synthétisés sous la forme de critères de performance utilisés classiquement dans le domaine de la biométrie (TEE, $TFR@TFA=10^{-6}$, courbes ROC et histogrammes top-rank). Étant donné que dans cette base, nous connaissons la vérité terrain, il est possible de calculer les différents critères de performances.

Deux expériences sont réalisées pour estimer l'influence du nombre d'iris-codes fusionnés et pour déterminer l'intérêt de masquer partiellement les bits fragiles sur les taux de reconnaissance. Finalement, ces résultats sont comparés avec 4 méthodes de l'état de l'art : deux basées sur la fusion des textures d'iris et les deux autres sur la fusion des iris-codes.

4.1 Influence du nombre d'iris-codes fusionnés

La première expérience a été conduite pour déterminer (1) si il est intéressant de fusionner des iris-codes, et si oui, (2) s'il existe un nombre optimum d'iris-codes à fusionner avant que les performances commencent à se détériorer. En effet, l'ajout d'un nombre de plus en plus important d'iris-codes (avec des scores de qualité de plus en plus faibles) peut détériorer l'iris-code résultant de la fusion. Pour vérifier ces points, nous avons appliqué les techniques de fusion avec un

3. Seules les vidéos #05344v27 and #05416v25 n'ont pas passé l'enrôlement.

4. Chaque classe d'iris contient au moins un représentant, mais la plupart des classes en contiennent plus de 10.

5. La distance de Hamming est le ratio du nombre de bits différents des deux iris-codes sur le nombre total de bits.

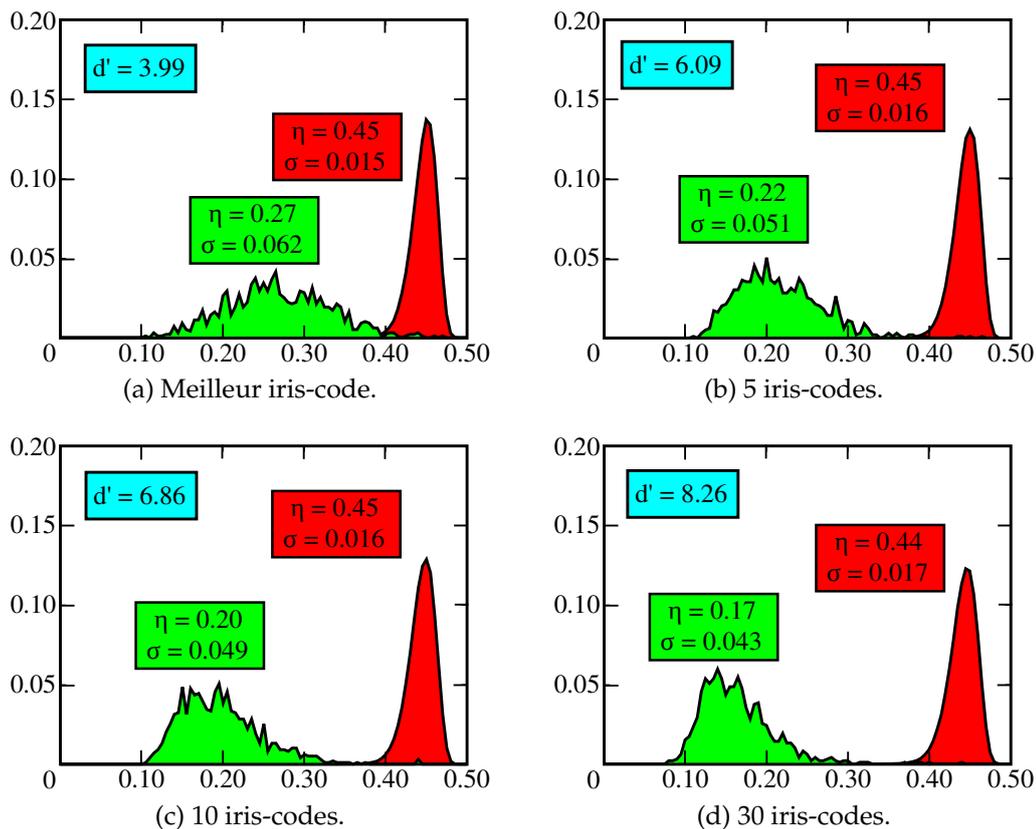


FIGURE 5.10: Répartition des distances pour l'image de plus grande qualité (a), la fusion des 5 (b), des 10 (c) et des 30 (d) meilleurs iris-codes. Les distances des imposteurs sont en rouge, celle des authentiques en vert. η , σ et d' désignent la moyenne, l'écart-type et la distance statistique entre les deux densités $d' = \frac{|\eta_g - \eta_i|}{\sqrt{0.5(\sigma_i^2 + \sigma_g^2)}}$.

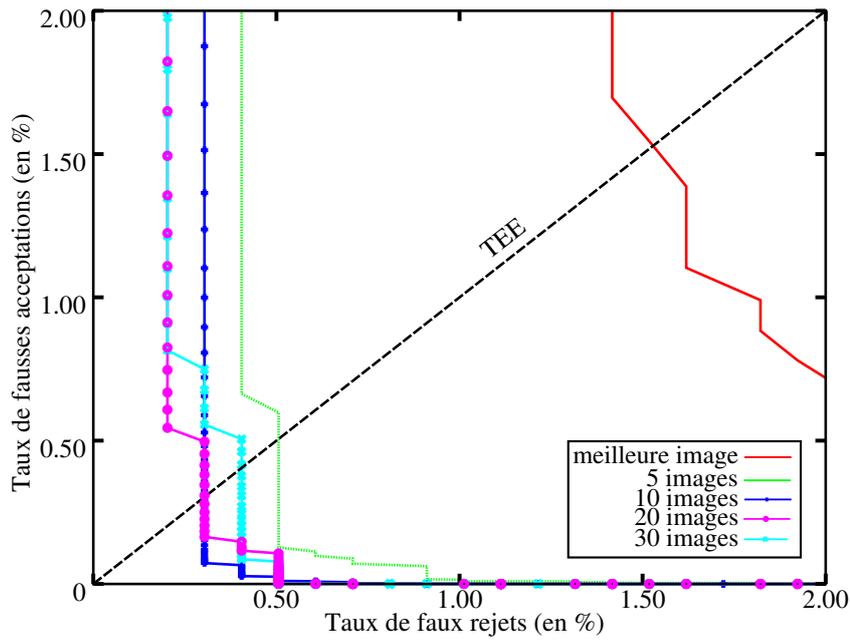


FIGURE 5.11: Courbes ROC : influence du nombre d'iris-codes fusionnés sur les performances de reconnaissance.

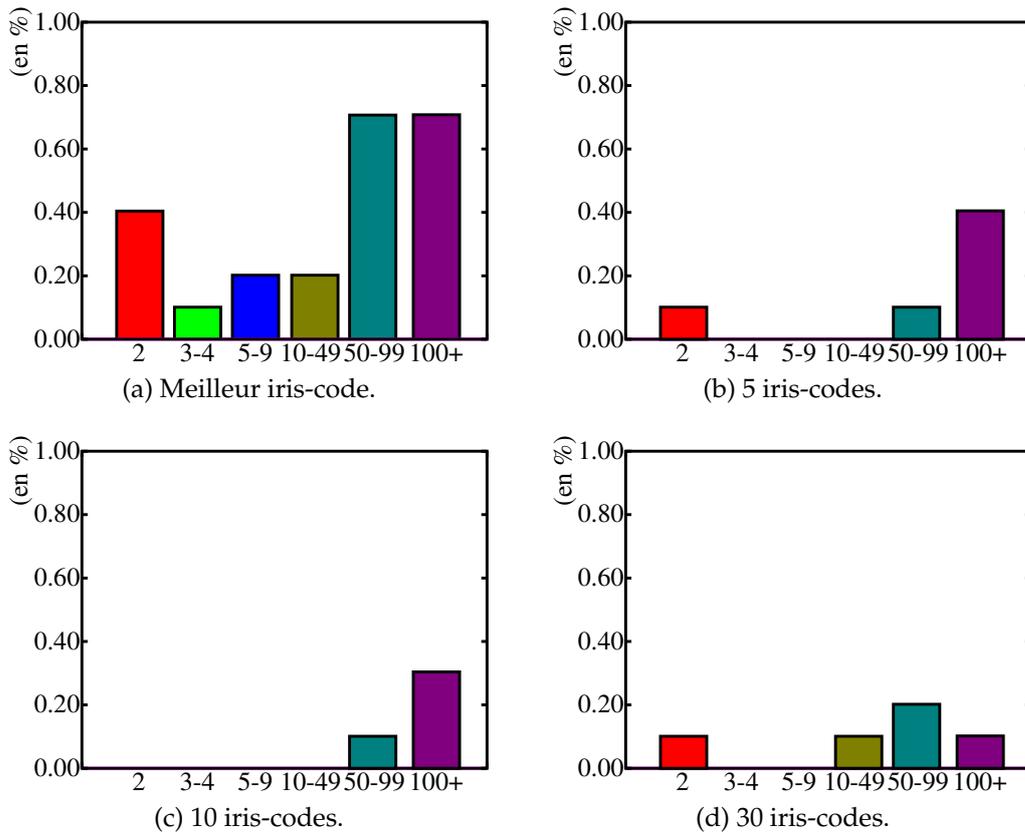


FIGURE 5.12: Rang de la vraie classe quand la classification a échoué pour le meilleur iris-code (a), la fusion des 5 (b), des 10 (c) et des 30 (d) meilleurs iris-codes.

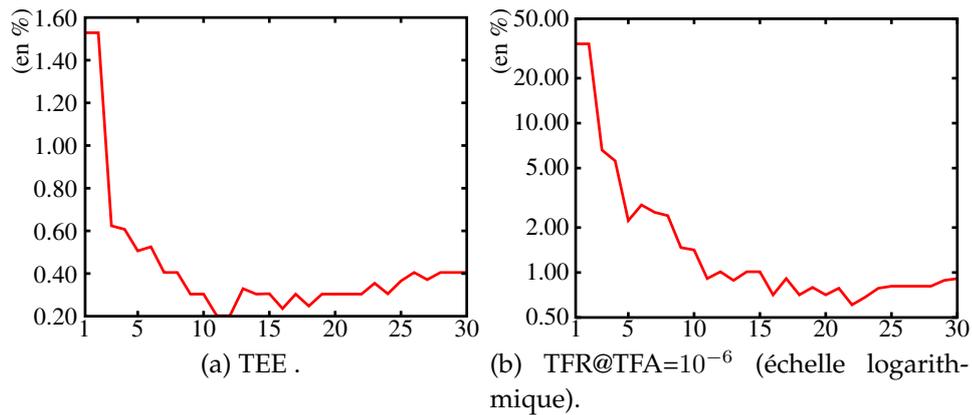


FIGURE 5.13: Évolution du TEE et du $\text{TFR@TFA}=10^{-6}$ en fonction du nombre d'iris-codes fusionnés.

nombre d'iris-codes variant de 1 à 30. La répartition des distances entre authentiques et imposteurs est représentée sur la figure 5.10 pour 1, 5, 10 et 30 iris-codes fusionnés. Les courbes top-rank sont données dans la figure 5.12, la courbe ROC sur la figure 5.11 et les évolutions du TEE et $\text{TFR@TFA}=10^{-6}$ dans la figure 5.13.

Si on observe les répartitions des distances, *cf.* Figure 5.10, la séparation entre authentiques et imposteurs augmente quand le nombre d'iris-codes fusionnés augmente. Comme le montre l'histogramme top-rank, les iris sont de mieux en mieux classés au fur à mesure que le nombre d'iris-codes fusionnés augmente. De plus, le $\text{TFR@TFA}=10^{-6}$ continue de décroître au delà de 10 iris-codes. Finalement, la fusion de plusieurs iris-codes dépasse les performances du meilleur iris-code : le TEE décroît de 1.58% à 0.2% (12 images), et le $\text{TFR@TFA}=10^{-6}$ décroît de 40% à 0.8% (22 images).

Pour conclure sur cette expérience, le point clé est que tous ces indicateurs confirment une amélioration des performances de la reconnaissance quand le nombre d'iris-codes fusionnés va de 1 à 10. De plus, la fusion d'un grand nombre d'iris-codes trouve son intérêt dans le domaine de la reconnaissance haute-sécurité.

4.2 Influence des bits fragiles

La seconde expérience a été conduite pour déterminer si le masquage des bits fragiles pouvait améliorer les performances. Ainsi, nous pouvons examiner les critères de performances lorsque $P\%$ des bits les plus fragiles sont ignorés lors du calcul de la distance de Hamming. L'intérêt de masquer les bits fragiles est de ne pas tenir compte d'une information peu sûre, mais masquer de plus en plus de bits réduit néanmoins l'information sur la texture de l'iris et peut, à terme, détériorer les performances d'identification. La répartition des distances pour des taux de bits fragiles de 0%, 15%, 30% et 45% est représentée sur la figure 5.14

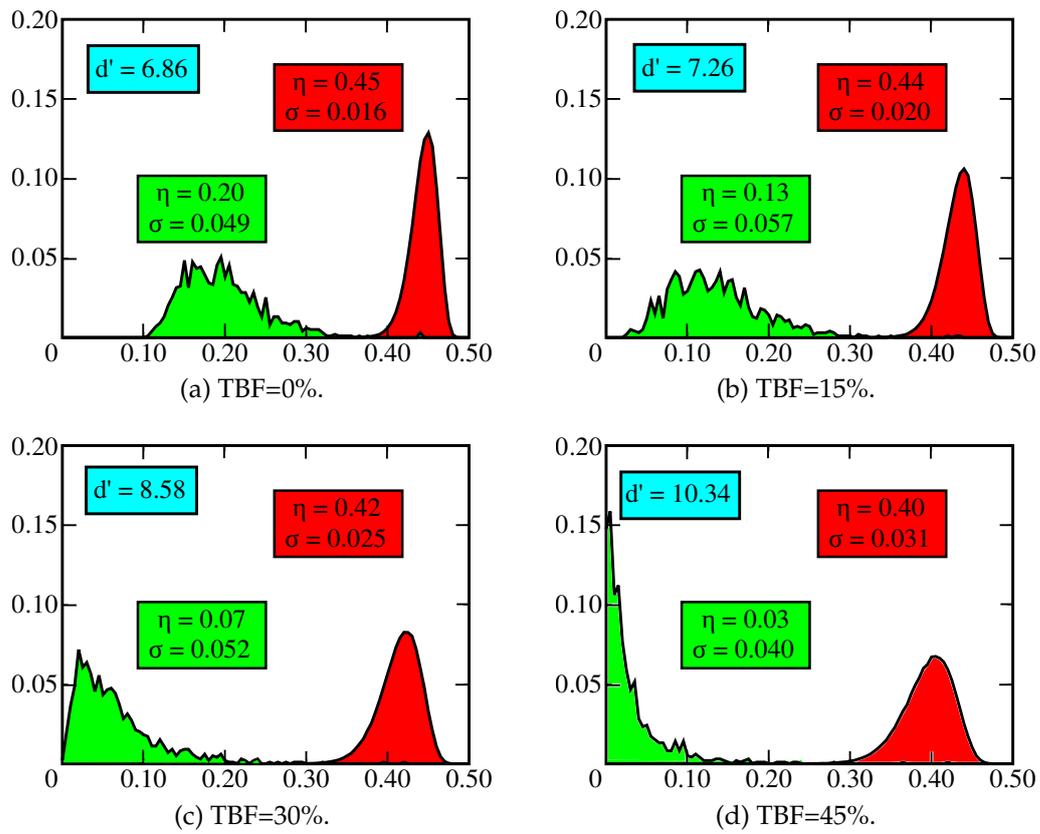


FIGURE 5.14: Répartition des distances entre authentiques et imposteurs pour des taux de bits fragiles différents et pour la fusion des 10 meilleurs iris-codes.

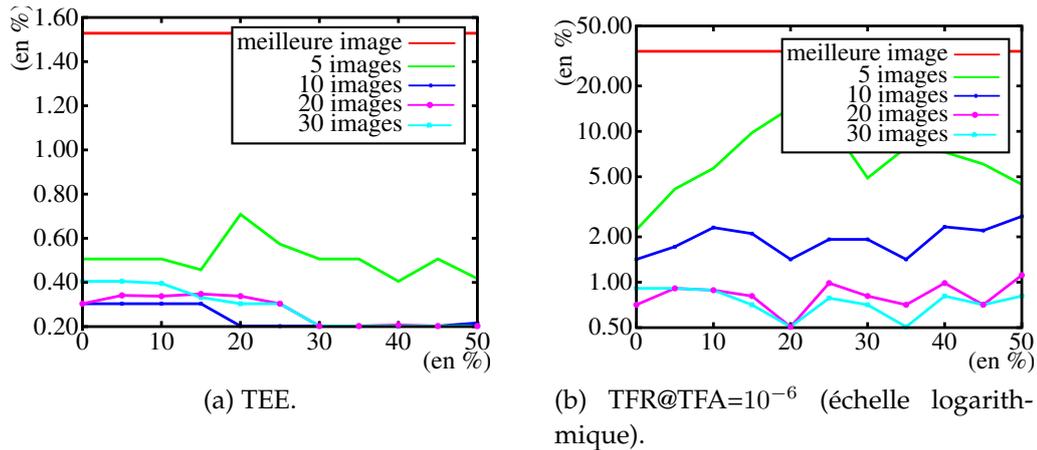


FIGURE 5.15: Évolution du TEE (a) et du $\text{TFR@TFA}=10^{-6}$ (b) en fonction du taux de bits fragiles.

lors de la fusion de 10 iris-codes. L'évolution du TEE et du $\text{TFR@TFA}=10^{-6}$ est représentée sur la figure 5.15 pour différents nombres d'iris-codes fusionnés.

Globalement, la fusion des 10 meilleurs iris-codes donne toujours le meilleur TEE quelque soit le taux de bits fragiles. Le TEE optimal (0.2%) est obtenu quand nous fusionnons 10 iris-codes avec un taux de bits fragiles de 20%, ou quand nous fusionnons 20 iris-codes avec un taux de bits fragiles de 30%. Comme l'expérience précédente, plus nous fusionnons d'iris-codes, plus le $\text{TFR@TFA}=10^{-6}$ diminue quelque soit le taux de bits fragiles. On peut aussi remarquer que le meilleur $\text{TFR@TFA}=10^{-6}$ est obtenu quand nous fusionnons plus de 20 iris-codes pour un taux de bits fragiles de 20%. Si on analyse la répartition des distances entre authentiques et imposteurs, masquer de plus en plus de bits a tendance à réduire la moyenne et l'écart-type des distances entre authentiques, et à augmenter l'écart-type des distances entre imposteurs.

Donc, deux conclusions peuvent être tirées de cette expérience sur la base de données MBGC :

1. Pour les applications tout public, il est conseillé de fusionner environ 10 iris-codes avec un taux de bits fragiles de 20%.
2. Pour les applications haute-sécurité, il est recommandé de fusionner plus de 20 iris-codes avec un taux de bits fragiles de l'ordre de 20%.

4.3 Comparaison avec d'autres méthodes de fusion

La méthode présentée dans ce chapitre a été comparée à quatre autres études

- (1) La première [50] décrit une méthode de super-résolution de la texture de l'iris pour améliorer les performances de reconnaissance.
- (2) La seconde [51] propose une méthode de super-résolution sur la transformée de l'iris par les ondelettes de Gabor 2D.
- (3) La troisième [29] propose de fusionner les iris-codes et d'éliminer les bits fragiles comme nous l'avons fait.

TABLE 5.1: Comparaison du TEE obtenu avec d'autres méthodes sur la base de données MBGC.

(1)	(2)	(3) (20 rayons)	(4) (10 rayons)	Nous (20 rayons)
~ 0.75%	~ 0.87%	~ 0.80%	~ 0.29%	0.20%

(4) La quatrième [30] utilise une distance qui prend en compte la fragilité des bits et qui améliore les résultats de la 3e méthode.

Le tableau 5.1 reporte les TEE pour les meilleurs résultats obtenus par les 4 méthodes ci-dessus et notre algorithme. Ce dernier montre une réduction significative du TEE et s'avère être plus performante que l'état de l'art. De plus, notre méthode est capable de traiter les vidéos en temps-réel, ce qui n'est pas le cas des autres méthodes qui ne reportent pas leur temps de calcul.

5 Conclusion

Dans ce chapitre, nous avons proposé un algorithme temps-réel de fusion d'iris-codes basée sur la qualité des images pour améliorer la reconnaissance d'iris à partir de vidéos centrées sur l'œil. Le traitement temps-réel est obtenu grâce à la décomposition multi-tâches de la chaîne de traitement et à la sélection intelligente des images. Une carte de fragilité a aussi été associée à chaque iris-code, qui donne une mesure de la confiance de chaque bits à être soit 0 ou 1.

L'algorithme a été systématiquement évalué sur la base de données MBGC [53] en fonction du nombre d'iris-codes fusionnés et de l'influence du nombre de bits fragiles dans le calcul de la distance de Hamming. La fusion proposée améliore significativement les performances. De plus, les paramètres optimaux pour des applications tout-publics ou haute-sécurité ont été identifiés. Les performances ont aussi été comparées avec quatre méthodes compétitives et notre algorithme divise le TEE par 1.5.

Les ondelettes log-Gabor 1D ont été choisies pour leur bonne performance et leur faible complexité. Néanmoins, de meilleures performances pourraient être obtenue en utilisant les ondelettes bi-orthogonales [1, 68]. De plus, la distance de Hamming pourrait avantageusement être remplacée par un autre critère, comme celui proposé dans [12]. L'algorithme de type RANSAC utilisé pour segmenter la pupille semble être très efficace ; une comparaison avec la méthode publiée récemment [41] serait intéressante.

Conclusion

Tout au long de cette thèse, nous avons abordé le problème de l'estimation des paramètres par maximum de vraisemblance du modèle de filtre de Kalman couple (FKC) [17]. Ce dernier étend le modèle classique en apportant de nouvelles possibilités d'interactions entre états cachés et observations. Les algorithmes portent soit sur l'estimation de l'ensemble des paramètres, soit sur l'estimation d'une partie de ces paramètres, l'autre étant supposée connue ou liée à un paramètre à estimer. Ce manuscrit présente les différents algorithmes originaux d'estimation issus du principe EM (*Expectation-Maximization*) que nous avons construits et évalués à cet effet.

Dans un premier temps, nous avons proposé une variante robuste de l'algorithme EM, proposé dans [5], pour l'estimation de la totalité des paramètres dans un FKC [48]. Cette variante, qui s'inspire de travaux antérieurs sur le filtre de Kalman [47], utilise des décompositions QR pour propager des racines-carrées des matrices de covariance au lieu des matrices de covariance elles-mêmes. L'apport de cette technique robuste est illustrée au travers d'une expérience sur un modèle couple particulièrement mal-conditionné.

Le chapitre 2 s'est intéressé au problème de l'identifiabilité du FKC selon le critère de la vraisemblance. Nous avons pu montrer, qu'à un jeu d'observations donné, il existait une infinité de jeux de paramètres de même fonction de vraisemblance. Cela se révèle particulièrement problématique pour l'algorithme EM qui converge vers un maximum local de la vraisemblance. Afin de réduire l'ambiguïté, nous avons établi 3 types de jeux de paramètres équivalents qui imposent une structure particulière aux paramètres lors de leur estimation, sans modifier la fonction de vraisemblance. L'utilisation de chacune de ces contraintes est illustrée dans le cadre de la restauration de signaux bruités.

Dans le troisième chapitre, nous avons proposé d'intégrer des contraintes sur la structure des paramètres du FKC dans l'étape de maximisation de l'algorithme EM, sans rechercher à préserver la fonction de vraisemblance comme dans le chapitre précédent. Cependant, n'importe quel type de contrainte ne permet pas de conserver des équations de ré-estimation exactes lors de l'étape M. Aussi, nous avons proposé des contraintes particulières qui s'avèrent pourtant utiles dans la pratique car elles permettent de traiter un grand nombre de problèmes (cas de capteurs identiques par exemple) et généralisent des travaux récents sur le filtre de Kalman [20]. Tous les algorithmes de maximisation issus de nos travaux

sont également proposés en version robuste compatible avec l'étape d'espérance robuste, présentée dans le chapitre 1. L'intérêt de contraindre les paramètres est illustré par des séries d'expériences qui comparent systématiquement les filtres équivalents du chapitre précédent aux filtres contraints.

Le chapitre 4 a permis d'exposer les résultats de comparaisons systématiques entre les différents algorithmes non-supervisés proposés dans cette thèse sur des données théoriques et sur des données ne suivant pas le modèle du FKC. Ces comparaisons ont permis d'établir que

- le FKC permet de restaurer des données suivant le modèle du filtre de Kalman malgré son plus grand nombre de paramètres. Par contre, comme attendu, nous avons vérifié que le filtre de Kalman ne restaure pas correctement des données issues d'un modèle de FKC.
- sur données non-issues de ces deux modèles (signaux sinusoïdaux), le FKC a donné des performances équivalentes de restauration à celles du filtre de Kalman. Par contre, le FKC a montré un réel intérêt sur un signal retardé, *i.e.* où l'acquisition des observations par le capteur n'est pas instantanée.

Dans le dernier chapitre, les résultats de filtrage robuste optimal obtenus dans les chapitres précédents sont exploités pour suivre le mouvement de la pupille dans une application de reconnaissance dynamique de l'iris. La qualité de suivi, ajouté à une sélection intelligente des meilleures images d'iris, et à la parallélisation des différents étapes de segmentation de l'iris, nous a permis, d'une part, d'atteindre des traitements temps-réel sur les vidéos (*i.e.* 25 images par seconde) et, d'autre part, d'obtenir des taux de reconnaissance supérieurs à ceux de l'état de l'art sur la base biométrique MBGC [53].

Nous avons dissocié dans ce manuscrit les jeux de paramètres équivalents et les contraintes sur l'étape de maximisation de l'algorithme EM du FKC. On peut penser combiner ces deux idées pour construire un FKC équivalent et contraint, ce qui permettrait d'élargir l'ensemble des contraintes possibles et donc de s'adapter à un plus large champs d'applications.

Une autre perspective consisterait à imposer, lors de l'estimation, que le signal à restaurer soit stationnaire. Imposer la stationnarité au modèle du FKC revient à contraindre les matrices de transition et de covariance du bruit simultanément, ce qui n'est pas compatible avec les contraintes étudiées dans cette thèse. Sans une étude approfondie, il n'est pas possible d'affirmer qu'il existe un algorithme EM dans ce cas de figure.

Finalement, en ce qui concerne l'application de biométrie dynamique de l'iris, il semble que les paramètres appris par les filtres de Kalman non-supervisés soient propres à chaque individu, *c.a.d* caractéristique du mouvement propre de chaque œil. Aussi, utiliser un jeu de paramètres unique pour un grand nombre d'individus n'a pas été concluant. Il serait donc intéressant de développer une version « en ligne » de ces algorithmes d'apprentissage non-supervisés, en étendant les algorithmes développés récemment dans [11] au FKC contraint pour une

poursuite optimale individualisée de la pupille. D'ailleurs, la dynamique d'un œil serait-elle caractéristique d'un comportement particulier ou d'un individu ? Et les paramètres du FKC seraient-ils des signatures de cet état ou individu ? La question de l'étude du mouvement de l'œil comme trait biométrique fait l'objet d'un travail de recherche depuis quelques années [15,28,57].

Bibliographie

- [1] Aditya Abhyankar and Stephanie Schuckers. Iris quality assessment and bi-orthogonal wavelet based encoding for recognition. *Pattern Recognition*, 42(9) :1878 – 1894, 2009.
- [2] B. Ait-el-Fquih and F. Desbouvries. Bayesian smoothing algorithms in pairwise and triplet Markov chains. In *Proc. of the 2005 IEEE Workshop on Statistical Signal Processing (SSP'05)*, Bordeaux, France, July 2005.
- [3] B. Ait-el-Fquih and F. Desbouvries. Exact and approximate Bayesian smoothing algorithms in partially observed Markov chains. In *Proc. of the IEEE Nonlinear Stat. Sig. Proc. Workshop (NSSPW'06)*, Cambridge, UK, September 2006.
- [4] B. Ait-el-Fquih and F. Desbouvries. Kalman filtering in triplet Markov chains. *IEEE Trans. on Signal Processing*, 54(8) :2957–2963, August 2006.
- [5] B. Ait-el-Fquih and F. Desbouvries. Unsupervised signal restoration in partially observed Markov chains. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'06)*, Toulouse, France, May 2006.
- [6] Gerald J. Bierman. Measurement updating using the u-d factorization. *Automatica*, 12(4) :375 – 382, 1976.
- [7] Gerald J Bierman. *Factorization methods for discrete sequential estimation*. Mathematics in Science and Engineering. Elsevier, Burlington, MA, 1977.
- [8] M. Boutayeb, H. Rafaralahy, and M. Darouach. Convergence analysis of the extended kalman filter used as an observer for nonlinear deterministic discrete-time systems. *Automatic Control, IEEE Transactions on*, 42(4) :581–586, Apr 1997.
- [9] C. Cabal. Les méthodes scientifiques d'identification des personnes à partir de données biométriques et les techniques de mise en oeuvre. Technical report, Office Parlementaire d'Evaluation des Choix Scientifiques et Technologiques, 2003.
- [10] G. Celeux and J. Diebolt. The SEM algorithm : A probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly*, 2 :73–82, 1985.
- [11] Sylvain Le Corff, Gersende Fort, and Eric Moulines. New online EM algorithms for general hidden Markov models. Application to the SLAM

- problem. In Fabian J. Theis, Andrzej Cichocki, Arie Yeredor, and Michael Zibulevsky, editors, *LVA/ICA*, volume 7191 of *Lecture Notes in Computer Science*, pages 131–138. Springer, 2012.
- [12] CiceroFerreiraFernandes Costa Filho, ClaudioFranklinMartins Pinheiro, MarlyGuimaraesFernandes Costa, and WagnerCoelho Albuquerque Pereira. Applying a novelty filter as a matching criterion to iris recognition for binary and real-valued feature vectors. *Signal, Image and Video Processing*, 7(2) :287–296, 2013.
- [13] J. Daugman. How iris recognition works. *IEEE Trans. on Circuits and Systems for Video Technology*, 14(1) :21–30, 2004.
- [14] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1) :1–38, 1977.
- [15] Farzin Deravi and Shivanand P Guness. Gaze trajectory as a biometric modality. In *BIOSIGNALS*, pages 335–341, 2011.
- [16] S. Derrode and W. Pieczynski. Unsupervised signal and image segmentation using pairwise Markov chains. *Signal Processing, IEEE Transactions on*, 52(9) :2477–2489, September 2004.
- [17] F. Desbouvries and W. Pieczynski. Triplet Markov models and Kalman filtering. *Comptes Rendus de l’Académie des Sciences - Mathématique - Série I*, 336(8) :667–670, 2003.
- [18] Arnaud Doucet, Simon Godsill, and Christophe Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *STATISTICS AND COMPUTING*, 10(3) :197–208, 2000.
- [19] G.A. Einicke and L.B. White. Robust extended kalman filtering. *Signal Processing, IEEE Transactions on*, 47(9) :2596–2599, Sep 1999.
- [20] E. Eli Holmes. Derivation of the EM algorithm for constrained and unconstrained MARSS models. Technical report, Northwest Fisheries Science Center, 2013.
- [21] R. J. Elliott, J. B. Moore, and L. Aggoun. *Hidden Markov Models*. Springer, 1995.
- [22] V. Nêmesin et al. The first ICB competition on iris recognition. In *IEEE/IAPR International Joint Conference on Biometrics*, 2014.
- [23] Martin A. Fischler and Robert C. Bolles. Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6) :381–395, June 1981.
- [24] Robert J. Fitzgerald. Divergence of the kalman filter. *Automatic Control, IEEE Transactions on*, 16(6) :736–747, Dec 1971.
- [25] J. De Geeter, H. Van Brussel, J. De Schutter, and M. Decreton. A smoothly constrained Kalman filter. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(10) :1171–1177, 1997.

- [26] Hamed Ghodrati, MohammadJavad Dehghani, and Habibolah Danyali. A new accurate noise-removing approach for non-cooperative iris recognition. *Signal, Image and Video Processing*, 8(1) :1–10, 2014.
- [27] N.J. Gordon, D.J. Salmond, and A. F M Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, 140(2) :107–113, Apr 1993.
- [28] C. Holland and O.V. Komogortsev. Biometric identification via eye movement scanpaths in reading. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–8, 2011.
- [29] K.P. Hollingsworth, E. Ortiz, and K.W. Bowyer. The best bits in an iris code. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31(6) :964–973, 2009.
- [30] K.P. Hollingsworth, E. Ortiz, and K.W. Bowyer. Improved iris recognition through fusion of Hamming distance and fragile bit distance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 33(12) :2465–2476, 2011.
- [31] Salvatore Ingrassia and Roberto Rocci. Monotone constrained EM algorithms for multinormal mixture models. In Sergio Zani, Andrea Cerioli, Marco Riani, and Maurizio Vichi, editors, *Data Analysis, Classification and the Forward Search*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 111–118. Springer Berlin Heidelberg, 2006.
- [32] A.K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Trans. on Circuits and Systems for Video Technology*, 14(1) :4–20, 2004.
- [33] A.K. Jain, A.A.A. Ross, and K. Nandakumar. *Introduction to Biometrics*. SpringerLink : Bücher. Springer US, 2011.
- [34] Y. K. Jang, B. J. Kang, and K. R. Park. A study on eyelid localization considering image focus for iris recognition. *Pattern Recognition Letters*, 29(11) :1698–1704, August 2008.
- [35] S.J. Julier, J.K. Uhlmann, and H.F. Durrant-Whyte. A new approach for filtering nonlinear systems. In *Proc. of the 1995 American Control Conf.*, Seattle, Washington, USA, June 1995.
- [36] R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory. *Transactions of the ASME. Series D, Journal of Basic Engineering*, 83 :95–107, 1961.
- [37] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D) :35–45, 1960.
- [38] P.G. Kaminski, A.E. Bryson, and S.F. Schmidt. Discrete square-root filtering : a survey of current techniques. *IEEE Transactions on Automatic Control*, 16(6) :727–736, December 1971.

- [39] ByungJun Kang and KangRyoung Park. A study on iris image restoration. In Takeo Kanade, Anil Jain, and NaliniK. Ratha, editors, *Audio- and Video-Based Biometric Person Authentication*, volume 3546 of *Lecture Notes in Computer Science*, pages 31–40. Springer Berlin Heidelberg, 2005.
- [40] W. Ketchantang, S. Derrode, S. Bourennane, and L. Martin. Video pupil tracking for iris based identification. In *Proc. of the 7th Int. Conf. on Advanced Concepts for Intelligent Vision Systems (ACIVS'05)*, pages 1–8, Antwerp, Belgium, 2005.
- [41] T. Lefevre, B. Dorizzi, S. Garcia-Salicetti, N. Lemperiere, and S. Belardi. Effective elliptic fitting for iris normalization. *Computer Vision and Image Understanding*, 117(6) :732 – 745, 2013.
- [42] Peihua Li, Xiaomin Liu, Lijuan Xiao, and Qi Song. Robust and accurate iris segmentation in very noisy iris images. *Image and Vision Computing*, 28(2) :246 – 253, 2010.
- [43] R. S. Lipster and A. N. Shiryaev. *Statistics of Random Processes : I. General theory*. Springer, 2nd edition, 2001.
- [44] L. Ljung. Asymptotic behavior of the extended kalman filter as a parameter estimator for linear systems. *Automatic Control, IEEE Transactions on*, 24(1) :36–50, Feb 1979.
- [45] Libor Masek. Recognition of human iris patterns for biometric identification. Technical report, University of Western Australia, 2003.
- [46] Nicholas Metropolis and Stanislaw M. Ulam. The Monte Carlo Method. *Journal of the American Statistical Association*, 44(247) :335–341, September 1949.
- [47] M. Morf and T. Kailath. Square-root algorithms for least-squares estimation. *IEEE Transactions on Automatic Control*, 20(4) :487–497, August 1975.
- [48] V. Nemesin and S. Derrode. Robust Blind Pairwise Kalman Algorithms Using QR Decompositions. *Signal Processing, IEEE Transactions on*, 61(1) :5–9, Jan 2013.
- [49] V. Nemesin, S. Derrode, and A. Benazza-Benyahia. Gradual iris code construction from close-up eye video. In J. Blanc-Talon, W. Philips, D. C. Popescu, P. Scheunders, and P. Zemčík, editors, *14th Int. Conf. on Advanced Concepts for Intelligent Vision Systems (ACIVS'12)*, volume 7517 of *Lecture Notes in Computer Science*, pages 12–23. Springer, 2012.
- [50] K. Nguyen, C. Fookes, S. Sridharan, and S. Denman. Quality-driven super-resolution for less constrained iris recognition at a distance and on the move. *IEEE Trans. on Information Forensics and Security*, 6(4) :1248–1258, 2011.
- [51] Kien Nguyen, Clinton Fookes, Sridha Sridharan, and Simon Denman. Feature-domain super-resolution for iris recognition. *Computer Vision and Image Understanding*, 117(10) :1526 – 1535, 2013.

- [52] D.O. North. An analysis of the factors which determine signal/noise discrimination in pulsed-carrier systems. *Proceedings of the IEEE*, 51(7) :1016–1027, July 1963.
- [53] P. J. Phillips, P. J. Flynn, J. R. Beveridge, W. T. Scruggs, A. J. O’Toole, D. Bolme, K. W. Bowyer, B. A. Draper, G. H. Givens, Y. M. Lui, H. Sahibzada, J. A. Scallan, and S. Weimer. Overview of the Multiple Biometrics Grand Challenge. In *Proc. of the 3rd Int. Conf. on Advances in Biometrics*, (ICB’09), pages 705–714, Berlin, Heidelberg, 2009. Springer-Verlag.
- [54] W. Pieczynski. Pairwise Markov chains. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(5) :634–639, 2003.
- [55] C. Price. An analysis of the divergence problem in the kalman filter. *Automatic Control, IEEE Transactions on*, 13(6) :699–702, Dec 1968.
- [56] S. T. Rachev, J. S. J. Hsu, B. S. Bagasheva, and F. J. Fabozzi. *Bayesian Methods in Finance*. John Wiley and Sons, 2008.
- [57] I Rigas, G. Economou, and S. Fotopoulos. Human eye movements as a trait for biometrical identification. In *Biometrics : Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*, pages 217–222, 2012.
- [58] Branko Ristic, Sanjeev Arulampalam, and Neil Gordon. *Beyond the Kalman Filter : Particle Filters for Tracking Applications*. Artech House, 2004.
- [59] Sam Roweis. Constrained hidden Markov models. In *Advances in Neural Information Processing Systems 12 (NIPS’99)*, pages 782–788. MIT Press, 1999.
- [60] Kaushik Roy, Prabir Bhattacharya, and ChingY. Suen. Iris segmentation using game theory. *Signal, Image and Video Processing*, 6(2) :301–315, 2012.
- [61] Joseph J. K. O. Ruanaidh and William J. Fitzgerald. *Numerical Bayesian methods applied to signal processing*. Springer-Verlag, 1996.
- [62] M.G. Rutten. Square-root unscented filtering and smoothing. In *Intelligent Sensors, Sensor Networks and Information Processing, 2013 IEEE Eighth International Conference on*, pages 294–299, April 2013.
- [63] F. H. Schlee, C. J. STANDISH, and N. F. TODA. Divergence in the kalman filter. *American Institute of Aeronautics and Astronautics Journal*, 5(6) :1114–1120, 1967.
- [64] K. Y. Shin, K. R. Park, B. J. Kang, and S.-J. Park. Super-resolution method based on multiple multi-layer perceptrons for iris recognition. In *Proc. of the 4th Int. Conf. on Ubiquitous Information Technologies Applications (ICUT’09)*, pages 1–5, 2009.
- [65] R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications, second edition*. Springer, 2006.
- [66] R.H. Shumway and D.S. Stoffer. An approach to time series smoothing and forecasting using the EM algorithm. *J. of Time Series Analysis*, 3(4) :253–264, July 1982.

- [67] D. Simon and Tien Li Chia. Kalman filtering with state equality constraints. *IEEE Trans. on Aerospace and Electronic Systems*, 38(1) :128–136, 2002.
- [68] R. Szewczyk, K. Grabowski, M. Napieralska, W. Sankowski, M. Zubert, and A. Napieralski. A reliable iris recognition algorithm based on reverse biorthogonal wavelet transform. *Pattern Recognition Letters*, 33(8) :1019 – 1026, 2012.
- [69] E. Tabassi, P. Grother, and W. Salamon. IREX II - IQCE - Iris Quality Calibration and Evaluation. Technical report, NIST Interagency Report 7820, 2011.
- [70] R. Van der Merwe and E.A. Wan. The square-root unscented kalman filter for state and parameter-estimation. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, volume 6, pages 3461–3464 vol.6, 2001.
- [71] E.A. Wan and R. Van der Merwe. The unscented kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pages 153–158, 2000.
- [72] Max Welling and Markus Weber. A constrained EM algorithm for independent component analysis. *Neural Computation*, 13 :677–689, 2001.
- [73] Norbert Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. The MIT Press, 1964.

Annexe

1 Preuves du chapitre 2

1.1 Démonstration 1

Nous cherchons à montrer que la matrice M définie par (2.8) transforme la matrice de covariance Q en une matrice de la forme (2.7) :

– Prouvons d’abord que $[Q^M]^{x,y} = O$:

$$\begin{aligned} [Q^M]^{x,y} &= M^{x,x}Q^{x,y} + M^{x,y}Q^{y,y}, \\ &= M^{x,x}Q^{x,y} - M^{x,x}Q^{x,y}, \\ &= O_{n_x,n_y}. \end{aligned}$$

– Prouvons ensuite que $[Q^M]^{x,x} = [Q_0^M]^{x,x}$:

$$\begin{aligned} [Q^M]^{x,x} &= M^{x,x}Q^{x,x}[M^{x,x}]^T + M^{x,x}Q^{x,y}[M^{x,y}]^T \\ &\quad + M^{x,y}[Q^{x,y}]^T[M^{x,x}]^T + M^{x,y}Q^{y,y}[M^{x,y}]^T, \\ &= M^{x,x}[Q^{x,x} - Q^{x,y}[Q^{y,y}]^{-1}[Q^{x,y}]^T - Q^{x,y}[Q^{y,y}]^{-1}[Q^{x,y}]^T \\ &\quad + Q^{x,y}[Q^{y,y}]^{-1}[Q^{x,y}]^T][M^{x,x}]^T, \\ &= M^{x,x}\underbrace{[Q^{x,x} - Q^{x,y}[Q^{y,y}]^{-1}[Q^{x,y}]^T]}_{Q_*^{x,x}}[M^{x,x}]^T, \\ &= [Q_0^M]^{x,x}. \end{aligned}$$

1.2 Démonstration 2

Nous cherchons à montrer que la transformation de la matrice de transition F par la matrice M définie par (2.15) permet d’imposer les valeurs de $[F^M]^{y,x}$ et $[F^M]^{y,y}$:

– Prouvons d’abord que $[\mathbf{F}^M]^{y,x} = [\mathbf{F}_0^M]^{y,x}$:

$$\begin{aligned}
[\mathbf{F}^M]^{y,x} &= \mathbf{F}^{y,x} [\mathbf{M}^{x,x}]^{-1} \\
&= \begin{bmatrix} [\mathbf{T}^o]^T & \mathbf{O} \end{bmatrix} [\mathbf{Q}^o]^T [\mathbf{Q}_M^o]^{-T} [\mathbf{Q}_M^+]^T \\
&= \begin{bmatrix} [\mathbf{T}^o]^T & \mathbf{O} \end{bmatrix} [\mathbf{Q}^o]^T \mathbf{Q}^o \begin{bmatrix} [\mathbf{T}^o]^{-T} & \mathbf{O} \\ \mathbf{O} & \mathbf{I} \end{bmatrix} \begin{bmatrix} [\mathbf{T}^+]^T & \mathbf{O} \\ \mathbf{O} & \mathbf{I} \end{bmatrix} [\mathbf{Q}^+]^T \\
&= [\mathbf{I}, \mathbf{O}] \begin{bmatrix} [\mathbf{T}^+]^T & \mathbf{O} \\ \mathbf{O} & \mathbf{I} \end{bmatrix} [\mathbf{Q}^+]^T \\
&= \begin{bmatrix} [\mathbf{T}^+]^T & \mathbf{O} \end{bmatrix} [\mathbf{Q}^+]^T \\
&= [\mathbf{F}_0^M]^{y,x}
\end{aligned}$$

– Prouvons ensuite que $[\mathbf{F}^M]^{y,y} = [\mathbf{F}_0^M]^{y,y}$:

$$\begin{aligned}
[\mathbf{F}^M]^{y,y} &= \mathbf{F}^{y,y} - \mathbf{F}^{y,x} [\mathbf{M}^{x,x}]^{-1} \mathbf{M}^{x,y} \\
&= \mathbf{F}^{y,y} - \begin{bmatrix} [\mathbf{T}^+]^T & \mathbf{O} \end{bmatrix} [\mathbf{Q}^+]^T [\mathbf{Q}_M^+]^{-T} \begin{bmatrix} \mathbf{F}^{y,y} - [\mathbf{F}_0^M]^{y,y} \\ \mathbf{O} \end{bmatrix} \\
&= \mathbf{F}^{y,y} - \begin{bmatrix} [\mathbf{T}^+]^T & \mathbf{O} \end{bmatrix} [\mathbf{Q}^+]^T \mathbf{Q}^+ \begin{bmatrix} [\mathbf{T}^+]^{-T} & \mathbf{O} \\ \mathbf{O} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{F}^{y,y} - [\mathbf{F}_0^M]^{y,y} \\ \mathbf{O} \end{bmatrix} \\
&= \mathbf{F}^{y,y} - [\mathbf{I}, \mathbf{O}] \begin{bmatrix} \mathbf{F}^{y,y} - [\mathbf{F}_0^M]^{y,y} \\ \mathbf{O} \end{bmatrix} \\
&= [\mathbf{F}_0^M]^{y,y}
\end{aligned}$$

2 Preuves du chapitre 3

2.1 Démonstration 3

Nous cherchons à montrer la relation (3.3) :

– \mathbf{Q} est bloc-diagonale, donc son inverse et le logarithme de son déterminant s’écrivent respectivement

$$\begin{aligned}
\mathbf{Q}^{-1} &= \sum_{i=1}^{n_{IN}} \mathbf{P}_i^T \mathbf{Q}_i^{-1} \mathbf{P}_i, \\
\log |\mathbf{Q}| &= \sum_{i=1}^{n_{IN}} \log |\mathbf{Q}_i|.
\end{aligned}$$

– Exprimons la fonction auxiliaire de vraisemblance en fonction de \mathbf{Q}_i et \mathbf{F}_i :

$$\begin{aligned}
g_{\tilde{\mathbf{C}}}(\mathbf{Q}, \mathbf{F}) &= -[N + 1] \log |\mathbf{Q}| - \text{tr} \left[\mathbf{Q}^{-1} \begin{bmatrix} \mathbf{F}^T \\ -\mathbf{I}_{n_t} \end{bmatrix}^T \tilde{\mathbf{C}} \begin{bmatrix} \mathbf{F}^T \\ -\mathbf{I}_{n_t} \end{bmatrix} \right] \\
&= -[N + 1] \sum_{i=1}^{n_{IN}} \log |\mathbf{Q}_i| - \text{tr} \left[\sum_{i=1}^{n_{IN}} \mathbf{P}_i \mathbf{Q}_i^{-1} \mathbf{P}_i^T \begin{bmatrix} \mathbf{F}^T \\ -\mathbf{I}_{n_t} \end{bmatrix}^T \tilde{\mathbf{C}} \begin{bmatrix} \mathbf{F}^T \\ -\mathbf{I}_{n_t} \end{bmatrix} \right] \\
&= -\sum_{i=1}^{n_{IN}} \left[[N + 1] \log |\mathbf{Q}_i| + \text{tr} \left[\mathbf{Q}_i^{-1} \begin{bmatrix} \mathbf{F}_i^T \\ -\mathbf{P}_i^T \end{bmatrix}^T \tilde{\mathbf{C}} \begin{bmatrix} \mathbf{F}_i^T \\ -\mathbf{P}_i^T \end{bmatrix} \right] \right] \\
&= -\sum_{i=1}^{n_{IN}} \left[[N + 1] \log |\mathbf{Q}_i| \right. \\
&\quad \left. + \text{tr} \left[\mathbf{Q}_i^{-1} \begin{bmatrix} \mathbf{F}_i^T \\ -\mathbf{I}_{n_t} \end{bmatrix}^T \begin{bmatrix} \mathbf{I}_{n_t} & \mathbf{O}_{n_t, n_i} \\ \mathbf{O}_{n_t, n_i} & \mathbf{P}_i \end{bmatrix} \tilde{\mathbf{C}} \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{O} & \mathbf{P}_i \end{bmatrix} \begin{bmatrix} \mathbf{F}_i^T \\ -\mathbf{I}_{n_t} \end{bmatrix}^T \right] \right] \\
&= \sum_{i=1}^{n_{IN}} g_{\tilde{\mathbf{C}}_i}(\mathbf{Q}_i, \mathbf{F}_i).
\end{aligned}$$

2.2 Démonstration 4

Nous cherchons à montrer la relation (3.6) :

$$\begin{aligned}
g_{\tilde{\mathbf{C}}_i}(\mathbf{Q}_i, \mathbf{F}_i) &= -[N + 1] \log |\mathbf{Q}_i| - \text{tr} \left[\mathbf{Q}_i^{-1} \begin{bmatrix} \mathbf{F}_i^T \\ -\mathbf{I}_{n_t} \end{bmatrix}^T \tilde{\mathbf{C}}_i \begin{bmatrix} \mathbf{F}_i^T \\ -\mathbf{I}_{n_t} \end{bmatrix} \right] \\
&= -[N + 1] \log |\mathbf{Q}_i| - \text{tr} \left[\mathbf{Q}_i^{-1} \begin{bmatrix} [\mathbf{F}_i^b]^T + [\mathbf{F}_i^0]^T \\ -\mathbf{I}_{n_t} \end{bmatrix}^T \tilde{\mathbf{C}}_i \begin{bmatrix} [\mathbf{F}_i^b]^T + [\mathbf{F}_i^0]^T \\ -\mathbf{I}_{n_t} \end{bmatrix} \right] \\
&= -[N + 1] \log |\mathbf{Q}_i| - \text{tr} \left[\mathbf{Q}_i^{-1} \begin{bmatrix} [\mathbf{F}_i^b]^T \\ -\mathbf{I}_{n_t} \end{bmatrix}^T \tilde{\mathbf{C}}_i^b \begin{bmatrix} [\mathbf{F}_i^b]^T \\ -\mathbf{I}_{n_t} \end{bmatrix} \right] \\
&= g_{\tilde{\mathbf{C}}_i^b}(\mathbf{Q}_i, \mathbf{F}_i).
\end{aligned}$$

2.3 Démonstration 5

Nous cherchons à montrer les relations (3.7) et (3.8) :

– Dérivons d'abord la sous-fonction auxiliaire de vraisemblance $g_{\tilde{\mathbf{C}}_i^b}(\mathbf{Q}_i, \mathbf{F}_i^b)$ par rapport à \mathbf{G}_i :

$$\begin{aligned}
\frac{\partial g_{\tilde{\mathbf{C}}_i^b}(\mathbf{Q}_i, \mathbf{F}_i^b)}{\partial \mathbf{G}_i} &= -\frac{\partial \text{tr} \left[\mathbf{Q}_i^{-1} \begin{bmatrix} [\mathbf{F}_i^b]^T \\ -\mathbf{I}_{n_t} \end{bmatrix}^T \tilde{\mathbf{C}}_i^b \begin{bmatrix} [\mathbf{F}_i^b]^T \\ -\mathbf{I}_{n_t} \end{bmatrix} \right]}{\partial \mathbf{G}_i} \\
&= -2\mathbf{Q}_i^{-1} \begin{bmatrix} \mathbf{M}_i^T \\ \mathbf{O}_{n_t, n_t} \end{bmatrix}^T \tilde{\mathbf{C}}_i^b \begin{bmatrix} \mathbf{M}_i^T \mathbf{G}_i^T \\ -\mathbf{I}_{n_t} \end{bmatrix} \\
&= -2\mathbf{Q}_i^{-1} \left[\mathbf{M}_i \tilde{\mathbf{C}}_i^{b, [0,0]} \mathbf{M}_i^T \mathbf{G}_i^T - \mathbf{M}_i \tilde{\mathbf{C}}_i^{b, [0,1]} \right].
\end{aligned}$$

- Un extremum est trouvé en résolvant $\frac{\partial g_{\tilde{\mathcal{C}}_i^b}(\mathbf{Q}_i, \mathbf{F}_i^b)}{\partial \mathbf{G}_i} = \mathbf{O}_{n_i, n_i^G}$. La matrice

$$\hat{\mathbf{G}}_i = \tilde{\mathcal{C}}_i^{b,[1,0]} \mathbf{M}_i^T \left[\mathbf{M}_i \tilde{\mathcal{C}}_i^{b,[0,0]} \mathbf{M}_i^T \right]^{-1}$$

est un extremum trivial de la fonction auxiliaire. Cependant, \mathbf{M}_i doit être de rang plein pour pouvoir inverser $\mathbf{M}_i^T \tilde{\mathcal{C}}_i^{b,[0,0]} \mathbf{M}_i$. L'extremum trouvé est aussi un maximum global, puisque la fonction $g_{\tilde{\mathcal{C}}_i^b}(\mathbf{Q}_i, \mathbf{F}_i^b)$ est parabolique selon la variable \mathbf{F} .

2.4 Démonstration 6

Nous cherchons à montrer la relation (3.9) et l'existence de la solution

- Dérivons la fonction auxiliaire de vraisemblance $g_{\tilde{\mathcal{C}}_i^b}(\mathbf{Q}_i, \mathbf{F}_i^b(\lambda_i))$ par rapport à λ_i^j :

$$\begin{aligned} \frac{\partial g_{\tilde{\mathcal{C}}_i^b}(\mathbf{Q}_i, \mathbf{F}_i^b(\lambda_i))}{\partial \lambda_i^j} &= - \frac{\partial \text{tr} \left[\mathbf{Q}_i^{-1} \begin{bmatrix} [\mathbf{F}_i^b]^T \\ -\mathbf{I}_{n_i} \end{bmatrix}^T \tilde{\mathcal{C}}_i^b \begin{bmatrix} [\mathbf{F}_i^b]^T \\ -\mathbf{I}_{n_i} \end{bmatrix} \right]}{\partial \lambda_i^j} \\ &= -2 \text{tr} \left[\mathbf{Q}_i^{-1} \begin{bmatrix} [\mathbf{U}_i^j]^T \\ -\mathbf{O}_{n_i, n_i} \end{bmatrix}^T \tilde{\mathcal{C}}_i^b \begin{bmatrix} \sum_{k=1}^{n_i^\lambda} \lambda_i^k [\mathbf{U}_i^k]^T \\ -\mathbf{I}_{n_i} \end{bmatrix} \right] \\ &= -2 \text{tr} \left[\mathbf{Q}_i^{-1} \left[\sum_{k=1}^{n_i^\lambda} \lambda_i^k \mathbf{U}_i^j \tilde{\mathcal{C}}_i^{b,[0,0]} [\mathbf{U}_i^k]^T - \mathbf{U}_i^j \tilde{\mathcal{C}}_i^{b,[0,1]} \right] \right] \end{aligned}$$

- Les extremums sont trouvés en annulant toutes les dérivées par rapport à λ_i^j simultanément :

$$\sum_{k=1}^{n_i^\lambda} \lambda_i^k \underbrace{\text{tr} \left[[\mathbf{Q}_i^0]^{-1} \mathbf{U}_i^j \tilde{\mathcal{C}}_i^{b,[0,0]} [\mathbf{U}_i^k]^T \right]}_{\mathbf{A}_i(j,k)} = \underbrace{\text{tr} \left[[\mathbf{Q}_i^0]^{-1} \mathbf{U}_i^j \tilde{\mathcal{C}}_i^{b,[0,1]} \right]}_{\mathbf{b}_i(j)}.$$

Donc, $\mathbf{A}_i \lambda_i = \mathbf{b}$ et $\lambda_i = \mathbf{A}_i^{-1} \mathbf{b}$. Si, un tel extremum existe, il est automatiquement un maximum puisque la sous-fonction $g_{\tilde{\mathcal{C}}_i^b}(\mathbf{Q}_i, \mathbf{F}_i^b(\lambda_i))$ est parabolique pour la variable \mathbf{F}_i^b .

- Prouvons maintenant l'inversibilité de \mathbf{A}_i . Il est suffisant de montrer qu'elle est définie positive. Selon les résultats classiques de l'algèbre linéaire, une matrice symétrique est définie positive si et seulement si

$$\forall (j, k), \mathbf{M}(j, j) \mathbf{M}(k, k) > [\mathbf{M}(j, k)]^2. \quad (7)$$

- Donc, $f(\mathbf{U}, \mathbf{V}) = \text{tr} \left[[\mathbf{Q}_i^0]^{-1} \mathbf{U} \tilde{\mathcal{C}}_i^{b,[0,0]} \mathbf{V}^T \right]$ est un produit scalaire sur l'espace des matrices réelles de taille $(n_i \times n_i)$. Puisque la matrice $\tilde{\mathcal{C}}_i^{b,[0,0]}$ est

définie positive, et symétrique, alors $g(\mathbf{U}, \mathbf{V}) = \text{tr} \left[\mathbf{U} \tilde{\mathbf{C}}_i^{b,[0,0]} \mathbf{V}^T \right]$ est un produit scalaire. On en déduit donc que

$$f(\mathbf{U}, \mathbf{V}) = \text{tr} \left[\left[[\mathbf{Q}_i^0]^{-1} \right]^{-\frac{T}{2}} \mathbf{U} \tilde{\mathbf{C}}_i^{b,[0,0]} \mathbf{V}^T \left[[\mathbf{Q}_i^0]^{-1} \right]^{-\frac{1}{2}} \right]$$

est un produit scalaire. Par application de l'inégalité de Cauchy-Schwarz dans le cadre d'une famille de matrices indépendantes, nous obtenons :

$$\left[\text{tr} \left[[\mathbf{Q}_i^0]^{-1} \mathbf{U}_i^j \tilde{\mathbf{C}}_i^{b,[0,0]} [\mathbf{U}_i^k]^T \right] \right]^2 < \text{tr} \left[[\mathbf{Q}_i^0]^{-1} \mathbf{U}_i^j \tilde{\mathbf{C}}_i^{b,[0,0]} [\mathbf{U}_i^j]^T \right] \text{tr} \left[[\mathbf{Q}_i^0]^{-1} \mathbf{U}_i^k \tilde{\mathbf{C}}_i^{b,[0,0]} [\mathbf{U}_i^k]^T \right].$$

- Donc, l'inégalité (7) est prouvée. \mathbf{A}_i est définie positive et donc inversible.

2.5 Démonstration 7

Nous cherchons à montrer la relation (3.17) :

- Dérivons la fonction auxiliaire de vraisemblance $g_{\tilde{\mathbf{C}}_i^M}(\mathbf{Q}_i^M, \hat{\mathbf{F}}_i^M)$ par \mathbf{Q}_i^M :

$$\frac{\partial g_{\tilde{\mathbf{C}}_i^M}(\mathbf{Q}_i^M, \hat{\mathbf{F}}_i^M)}{\partial \mathbf{Q}_i^M} = [N+1] [\mathbf{Q}_i^M]^{-1} - [\mathbf{Q}_i^M]^{-1} \begin{bmatrix} [\hat{\mathbf{F}}_i^M]^T \\ -\mathbf{I}_{n_i^M} \end{bmatrix}^T \tilde{\mathbf{C}}_i^M \begin{bmatrix} [\hat{\mathbf{F}}_i^M]^T \\ -\mathbf{I}_{n_i^M} \end{bmatrix} [\mathbf{Q}_i^M]^{-1}$$

- La matrice $\hat{\mathbf{Q}}_i^M$ qui maximise la fonction auxiliaire de vraisemblance est donc donnée par l'équation (3.17).

2.6 Démonstration 8

Cherchons à montrer la relation (3.18) :

- Dérivons la fonction auxiliaire de vraisemblance $g_{\tilde{\mathbf{C}}_i^M}(\mathbf{Q}_i^M(\lambda_i), \hat{\mathbf{F}}_i^M)$ par λ_i :

$$\frac{\partial g_{\tilde{\mathbf{C}}_i^M}(\mathbf{Q}_i^M(\lambda_i), \hat{\mathbf{F}}_i^M)}{\partial \lambda_i} = [N+1] n_i^M \lambda_i^{-1} - \lambda_i^{-2} \text{tr} \left[[\mathbf{Q}_i^0]^{-1} \begin{bmatrix} [\hat{\mathbf{F}}_i^M]^T \\ -\mathbf{I}_{n_i^M} \end{bmatrix}^T \tilde{\mathbf{C}}_i^M \begin{bmatrix} [\hat{\mathbf{F}}_i^M]^T \\ -\mathbf{I}_{n_i^M} \end{bmatrix} \right].$$

- Le scalaire $\hat{\lambda}_i$ qui maximise la fonction auxiliaire de vraisemblance est donc donné par l'équation (3.18).

2.7 Démonstration 9

Nous cherchons à montrer la relation (3.19) :

- Nous cherchons dans un premier temps, à exprimer la fonction auxiliaire de vraisemblance $g_{\tilde{\mathcal{C}}_i^M}(\mathbf{Q}_i^M, \hat{\mathbf{F}}_i^M)$ en fonction de $\hat{\mathbf{F}}_i^M$, $\tilde{\mathcal{C}}_i^M$ et \mathbf{R}_i^M .

Puisque \mathbf{Q}_i^M est semblable à une matrice bloc-diagonale, son inverse et le logarithme de son déterminant s'écrivent respectivement

$$\begin{aligned} [\mathbf{Q}_i^M]^{-1} &= \sum_{j=1}^{n_i^B} [\mathbf{P}_i^j]^T [\mathbf{M}_i^j]^{-T} [\mathbf{R}_i^M]^{-1} [\mathbf{M}_i^j]^{-1} \mathbf{P}_i^j, \\ \log |\mathbf{Q}_i^M| &= \sum_{j=1}^{n_i^B} \log |\mathbf{R}_i^M| + 2 \log |\mathbf{M}_i^j|. \end{aligned}$$

- Donc,

$$\begin{aligned} g_{\tilde{\mathcal{C}}_i^M}(\mathbf{Q}_i^M, \hat{\mathbf{F}}_i^M) &= [N + 1] \sum_{j=1}^{n_i^B} \log |\mathbf{R}_i^M| + 2 \log |\mathbf{M}_i^j| \\ &\quad + \sum_{j=1}^{n_i^B} \text{tr} \left[[\mathbf{R}_i^M]^{-1} [\mathbf{M}_i^j]^{-1} \mathbf{P}_i^j \begin{bmatrix} \hat{\mathbf{F}}_i^M \\ -\mathbf{I}_{n_i^M} \end{bmatrix}^T \right]^T \tilde{\mathcal{C}}_i^M \\ &\quad \begin{bmatrix} \hat{\mathbf{F}}_i^M \\ -\mathbf{I}_{n_i^M} \end{bmatrix}^T [\mathbf{P}_i^j]^T [\mathbf{M}_i^j]^{-T} \right]. \end{aligned}$$

- Dérivons maintenant la fonction auxiliaire de vraisemblance par rapport à \mathbf{R}_i^M

$$\begin{aligned} \frac{\partial g_{\tilde{\mathcal{C}}_i^M}(\mathbf{Q}_i^M, \hat{\mathbf{F}}_i^M)}{\partial \mathbf{R}_i^M} &= [N + 1] n_i^B [\mathbf{R}_i^M]^{-1} \\ &\quad - [\mathbf{R}_i^M]^{-1} [\mathbf{M}_i^j]^{-1} \mathbf{P}_i^j \begin{bmatrix} \hat{\mathbf{F}}_i^M \\ -\mathbf{I}_{n_i^M} \end{bmatrix}^T \tilde{\mathcal{C}}_i^M \\ &\quad \begin{bmatrix} \hat{\mathbf{F}}_i^M \\ -\mathbf{I}_{n_i^M} \end{bmatrix}^T [\mathbf{P}_i^j]^T [\mathbf{M}_i^j]^{-T} [\mathbf{R}_i^M]^{-1}. \end{aligned}$$

- La matrice $\hat{\mathbf{R}}_i^M$ qui maximise la fonction auxiliaire de vraisemblance est donc bien donnée par l'équation (3.19).

3 Filtres de Kalman particuliers

Cette section est dédiée à la construction de filtres de Kalman particuliers exploités dans le chapitre 4. Nous proposons d'imposer la valeur de la matrice d'observation $\mathbf{F}^{y,x}$, notée $\mathbf{F}_0^{y,x}$, soit par le choix d'un filtre équivalent, soit par des contraintes lors de l'étape de maximisation.

3.1 Filtre de Kalman équivalent

Le premier filtre particulier est le filtre de Kalman équivalent (FKE) qui permet d'imposer la valeur de la matrice d'observation $[\mathbf{F}^{y,x}]^M$. Son algorithme EM est obtenu par dérivation de l'algorithme EM du filtre de Kalman couple.

L'estimation de \mathbf{F} est séparée en deux blocs $\mathbf{F}^{x,t}$ et $\mathbf{F}^{y,t}$, donc $\mathbf{Q}^{x,y} = \mathbf{O}$. Sur chacun des blocs de \mathbf{F} , une contrainte de type-2 avec $\mathbf{M} = \mathbf{M}^1 = \mathbf{M}^2$ est imposée, ce qui donne

$$\begin{aligned} \mathbf{F}^{x,t} &= \underbrace{[\mathbf{F}^{x,x}]}_{\mathbf{G}^1} \underbrace{[\mathbf{I} \ \mathbf{O}]}_{\mathbf{M}} \\ \mathbf{F}^{y,t} &= \underbrace{[\mathbf{F}^{y,x}]}_{\mathbf{G}^2} \underbrace{[\mathbf{I} \ \mathbf{O}]}_{\mathbf{M}}. \end{aligned} \quad (8)$$

Les blocs diagonaux de \mathbf{Q} , $\mathbf{Q}^{x,x}$ et $\mathbf{Q}^{y,y}$ sont laissés libres. Finalement, nous utilisons le filtre équivalent de type γ pour imposer la valeur de la matrice d'observation $[\mathbf{F}^M]^{y,x} = \mathbf{F}_0^{y,x}$.

Remarque: Le filtre obtenu est bien un filtre de Kalman puisque $\mathbf{M}^{x,y} = \mathbf{O}$.

3.2 Filtre de Kalman contraint

Le second filtre particulier est le filtre de Kalman contraint (FK contraint) qui permet aussi d'imposer la valeur de la matrice d'observation $[\mathbf{F}^{y,x}]^M$. Son algorithme EM est également obtenu par dérivation de l'algorithme EM du filtre de Kalman couple.

L'estimation de \mathbf{F} est également séparée en deux blocs $\mathbf{F}^{x,t}$ et $\mathbf{F}^{y,t}$, donc $\mathbf{Q}^{x,y} = \mathbf{O}$. Sur le bloc $\mathbf{F}^{x,t}$, une contrainte de type-2 avec $\mathbf{M} = \mathbf{M}^1$ est imposée, ce qui donne

$$\mathbf{F}^{x,t} = \underbrace{[\mathbf{F}^{x,x}]}_{\mathbf{G}^1} \underbrace{[\mathbf{I} \ \mathbf{O}]}_{\mathbf{M}}. \quad (9)$$

Au bloc $\mathbf{F}^{y,t}$, il suffit d'imposer une contrainte de type-0 : $\mathbf{F}^{y,t} = [\mathbf{F}_0^{y,x} \ \mathbf{O}]$.

Remarque: Ce filtre ne doit pas être confondu avec le filtre précédent puisque la maximisation de la vraisemblance est réalisée différemment.

Liste des symboles

d.d.p densité de probabilité

EM *Expectation-Maximisation*

EQM erreur quadratique moyenne

FK filtre de Kalman

FKC filtre de Kalman couple

FKCE filtre de Kalman couple équivalent

FKCO filtre de Kalman couple optimal

FKE filtre de Kalman équivalent

FKO filtre de Kalman optimal

TEE taux d'égle erreur

TFA taux de fausses acceptations

TFR taux de faux rejets

$TFR@TFA=10^{-6}$ TFR pour un TFA= 10^{-6}

v.a variable aléatoire

\mathbf{a} un vecteur \mathbf{a}

n_a dimension du vecteur \mathbf{a}

\mathbf{A} une matrice \mathbf{A}

\mathbf{A}^T transposée de la matrice \mathbf{A}

\mathbf{A}^{-T} inverse de la transposée de la matrice \mathbf{A}

$\mathbf{A}^{\frac{1}{2}}$ racine-carré triangulaire supérieure de \mathbf{A}

$\mathbf{A}^{-\frac{1}{2}}$ inverse de $\mathbf{A}^{\frac{1}{2}}$

$\mathbf{A}^{\frac{T}{2}}$ transposée de $\mathbf{A}^{\frac{1}{2}}$

$\mathbf{A}^{-\frac{T}{2}}$ inverse de la transposée de $\mathbf{A}^{\frac{1}{2}}$

\mathbf{I}_n matrice identité de dimension n

$\mathbf{O}_{n,m}$ matrice nulle de dimensions $(n \times m)$

$\text{vec}(\mathbf{A})$ dé-plierement vectoriel de \mathbf{A}

$[\mathbf{A} \otimes \mathbf{B}]$ produit de Kronecker de \mathbf{A} et \mathbf{B}

\mathbf{Z}^* matrice unitaire inutilisée

$\mathbf{y}_{0:p}$ ensemble des \mathbf{y}_i pour i allant de 0 à p

\hat{a} espérance de la v.a a

\mathbf{x}_n état caché n

\mathbf{y}_n observation n

\mathbf{t}_n vecteur aléatoire $\mathbf{t}_n = [\mathbf{x}_n, \mathbf{y}_{n-1}]^T$

Table des figures

1.1	Algorithme EM pour le filtre de Kalman couple.	15
1.2	Calcul de $\tilde{C}^{\frac{1}{2}}$	22
1.3	Évolution de la plus petite valeur propre de Q au cours des itérations de l'algorithme EM.	24
2.1	Exemple d'un signal de 100 échantillons simulé et de ses restaurations par les différents filtres pour l'expérience de la section 3.1.	35
2.2	Vraisemblances moyennes des différents filtres pour un signal de 1 000 échantillons (section 3.1).	36
2.3	Évolution de l'EQM en fonction du nombre d'échantillons (section 3.1).	36
2.4	Exemple d'un signal de 100 échantillons simulé et de ses restaurations par les différents filtres (section 3.2).	38
2.5	Évolution de l'EQM en fonction du nombre d'échantillons (section 3.2).	39
2.6	Exemple d'un signal de 100 échantillons simulé et de ses restaurations par les différents filtres (section 3.3).	40
2.7	Évolution de l'EQM en fonction du nombre d'échantillons (section 3.3).	41
3.1	Estimation robuste et partielle de F	50
3.2	Estimation robuste et partielle de Q	52
3.3	Vraisemblances moyennes des différents filtres pour un signal de 1 000 échantillons (section 3.1).	55
3.4	Exemple d'un signal de 100 échantillons simulé et de ses restaurations par les différents filtres (section 3.1).	56
3.5	Evolution de l'EQM en fonction du nombre d'échantillons (section 3.1).	57
3.6	Exemple d'un signal de 100 échantillons simulé et de ses restaurations par les différents filtres (section 3.2).	59
3.7	Évolution des EQM pour le lissage en fonction du nombre d'échantillons (section 3.2).	60
3.8	Exemple d'un signal de 100 échantillons simulé et de ses restaurations par les différents filtres (section 3.3).	62

3.9	Évolution de l'EQM en fonction du nombre d'échantillons (section 3.3).	63
3.10	Exemple d'un signal de 100 échantillons simulé et de ses restaurations par les différents filtres (section 3.4).	66
3.11	Vraisemblances moyennes des différents filtres (section 3.4).	67
3.12	Évolution de l'EQM en fonction du nombre d'échantillons (section 3.4).	67
4.1	Évolution de l'EQM en fonction du nombre d'échantillons (section 1.1).	73
4.2	Exemple d'un signal de 100 échantillons simulé et de ses restaurations par les différents filtres (section 1.1).	74
4.3	Évolution de l'EQM en fonction du nombre d'échantillons (section 1.2).	76
4.4	Exemple d'un signal de 100 échantillons simulé et de ses restaurations par les différents filtres (section 1.2).	77
4.5	Exemple de signal généré (section 2.1).	78
4.6	Évolution de l'EQM en fonction du nombre d'échantillons (section 2.1).	80
4.7	EQM instantanée pour les différents filtres (section 2.1).	81
4.8	Évolution de l'EQM en fonction du nombre d'échantillons (section 2.2).	83
4.9	Signal simulé et restaurations (section 2.2).	85
5.1	Schéma de l'œil (source : www.infovisual.com).	92
5.2	Texture de l'iris [9].	92
5.3	Répartition des authentiques et des imposteurs.	93
5.4	Schéma d'enrôlement d'une vidéo.	95
5.5	Illustration des différentes étapes lors de l'enrôlement de l'iris.	96
5.6	Schéma de fusion des iris-codes.	97
5.7	Exemple de la prédiction de la zone d'intérêt (rectangle).	99
5.8	Organigramme des tâches dédiées à l'enrôlement de l'iris.	100
5.9	Différents modes de fonctionnement des mémoires tampons.	101
5.10	Répartition des distances pour l'image de plus grande qualité (a), la fusion des 5 (b), des 10 (c) et des 30 (d) meilleurs iris-codes. Les distances des imposteurs sont en rouge, celle des authentiques en vert. η , σ et d' désignent la moyenne, l'écart-type et la distance statistique entre les deux densités $d' = \frac{ \eta_g - \eta_i }{\sqrt{0.5(\sigma_i^2 + \sigma_g^2)}}$	103
5.11	Courbes ROC : influence du nombre d'iris-codes fusionnés sur les performances de reconnaissance.	104
5.12	Rang de la vraie classe quand la classification a échoué pour le meilleur iris-code (a), la fusion des 5 (b), des 10 (c) et des 30 (d) meilleurs iris-codes.	104

5.13	Évolution du TEE et du $\text{TFR@TFA}=10^{-6}$ en fonction du nombre d'iris-codes fusionnés.	105
5.14	Répartition des distances entre authentiques et imposteurs pour des taux de bits fragiles différents et pour la fusion des 10 meilleurs iris-codes.	106
5.15	Évolution du TEE (a) et du $\text{TFR@TFA}=10^{-6}$ (b) en fonction du taux de bits fragiles.	107

Liste des tableaux

2.1	Paramètres de l'expérience de la section 3.1; Ligne 1 : paramètres de simulation ; Ligne 2 : valeurs d'initialisation de l'algorithme EM; Lignes 3 et 4 : estimation moyenne des paramètres par les différents filtres pour un signal de 1 000 échantillons (100 itérations EM).	35
2.2	Paramètres de l'expérience de la section 3.2. Ligne 1 : paramètres de simulation ; Ligne 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 3 et 4 : estimation moyenne des paramètres par les différents filtres pour un signal de 1 000 échantillons (100 itérations EM).	38
2.3	Paramètres de l'expérience de la section 3.3. Ligne 1 : paramètres de simulation ; Ligne 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 3 et 4 : estimation moyenne des paramètres par les différents filtres pour un signal de 1 000 échantillons (100 itérations EM).	41
3.1	Paramètres de l'expérience de la section 3.1. Ligne 1 : paramètres de simulation ; Ligne 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 3 à 5 : estimation moyenne des paramètres par les différents filtres pour un signal d'une taille de 1 000 échantillons.	55
3.2	Paramètres de l'expérience de la section 3.2. Ligne 1 : paramètres de simulation ; Ligne 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 3 à 5 : estimation moyenne des paramètres par les différents filtres pour un signal d'une taille de 1 000 échantillons (100 itérations EM).	60
3.3	Paramètres de l'expérience de la section 3.3. Ligne 1 : paramètres de simulation ; Ligne 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 3 à 5 : estimation moyenne des paramètres par les différents filtres pour un signal d'une taille de 1 000 échantillons (100 itérations EM).	63

3.4	Paramètres de l'expérience de la section 3.4. Ligne 1 : paramètres de simulation ; Ligne 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 3 à 6 : estimation moyenne des paramètres par les différents filtres pour des signaux d'une taille de 100 échantillons ; Lignes 7 à 10 : estimation moyenne des paramètres pour des signaux d'une taille de 1 000 échantillons (100 itérations EM).	65
4.1	Paramètres de l'expérience de la section 1.1. Ligne 1 : paramètres de simulation ; Ligne 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 3 à 6 : estimation moyenne des paramètres par les différents filtres pour un signal d'une taille de 1 000 échantillons (100 itérations EM).	72
4.2	Paramètres de l'expérience de la section 1.2. Ligne 1 : paramètres de simulation ; Ligne 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 3 à 6 : estimation moyenne des paramètres par les différents filtres pour un signal d'une taille de 1 000 échantillons (100 itérations EM).	75
4.3	Paramètres de l'expérience de la section 2.1. Lignes 1 et 2 : valeurs d'initialisation de l'algorithme EM ; Lignes 4 à 9 : estimation moyenne des paramètres par les différents filtres pour un signal de 100 échantillons ; Lignes 10 à 15 : estimation moyenne des paramètres par les différents filtres pour un signal de 1 000 échantillons (100 ou 600 itérations EM).	79
4.4	Paramètres de l'expérience de la section 2.2. Lignes 1 à 3 : valeurs d'initialisation de l'algorithme EM ; Lignes 4 à 6 : estimation moyenne des paramètres par les différents filtres pour un signal d'une taille de 1 000 échantillons (100 itérations EM).	82
4.5	Valeurs d'initialisation des paramètres de l'expérience de la section 3	86
4.6	Paramètres estimés par le premier filtre non-supervisé pour les différentes vidéos et EQM de prédiction (section 3).	87
4.7	Paramètres estimés par le second filtre non-supervisé pour les différentes vidéos et EQM de prédiction (section 3).	88
4.8	Paramètres estimés par le troisième filtre non-supervisé pour les différentes vidéos et EQM de prédiction (section 3).	89
5.1	Comparaison du TEE obtenu avec d'autres méthodes sur la base de données MBGC.	108

RÉSUMÉ

La famille de modèles dite des filtres de Kalman permet d'estimer les états d'un système dynamique à partir d'une série de mesures incomplètes ou bruitées. Malgré leur relative simplicité de modélisation, ces filtres sont utilisés dans un large spectre scientifique dont le radar, la vision, et les communications. Ce succès repose, pour l'essentiel, sur l'existence d'algorithmes de filtrage et de lissage exacts et rapides, *i.e.* linéaires au nombre d'observations, qui minimisent l'erreur quadratique moyenne.

Dans cette thèse, nous nous sommes intéressés au filtre de Kalman couple. Celui-ci intègre, par rapport au modèle original, de nouvelles possibilités d'interactions entre états cachés et observations, tout en conservant des algorithmes exacts et rapides dans le cas linéaire et gaussien. Nous étudions plus particulièrement le problème de l'estimation non supervisée et robuste des paramètres d'un filtre de Kalman couple à partir d'observations en nombre limité. Le manuscrit décrit ainsi plusieurs algorithmes d'apprentissage par estimation du maximum de vraisemblance selon le principe EM (*Expectation-Maximization*). Ces algorithmes originaux permettent d'intégrer des contraintes a priori sur les paramètres du système étudié, comme expressions de connaissances partielles sur la physique de l'application ou sur le capteur. Ces systèmes contraints réduisent l'ambiguïté liée au problème d'identifiabilité du filtre de Kalman couple lors de l'estimation des paramètres. Ils permettent également de limiter le nombre de maxima locaux de la fonction de vraisemblance en réduisant la dimension de l'espace de recherche, et ainsi évitent parfois le piègeage de l'algorithme EM.

Il est important de noter que l'ensemble des algorithmes proposés dans ce manuscrit s'applique directement au filtre de Kalman original, comme cas particulier du filtre de Kalman couple. Tous les algorithmes sont rendus robustes par la propagation systématique de racines-carrés des matrices de covariance au lieu des matrices de covariance elles-mêmes, permettant ainsi d'éviter les difficultés numériques bien connues liées à la perte de positivité et de symétrie de ces matrices. Ces algorithmes robustes sont finalement évalués et comparés dans le cadre d'une application de biométrie de l'iris à partir de vidéos. Le suivi de la pupille est exploitée pour enrôler et identifier en temps-réel une personne grâce à son iris-code.
