

Centre Énergie Matériaux Télécommunications

SIMPLIFICATIONS DE FILTRE DE KALMAN POUR LE TRAITEMENT DE SIGNAUX

Par

Kais Boussofara

Mémoire présenté pour l'obtention du grade de
Maître ès Sciences (M.Sc.)
en télécommunications

Jury d'évaluation

Président du jury et
examineur interne

Prof. Jacob Benesty
INRS-ÉMT

Examineur externe

Prof. Daniel Massicotte
UQTR

Examineur interne

Prof. Jacob Benesty
INRS-ÉMT

Directeur de recherche

Prof. Leszek Szczecinski
INRS-ÉMT

Remerciements

C'est avec une profonde reconnaissance et un vif estime envers ceux et celles qui ont concouru à la valorisation de mes connaissances, que j'adresse mes remerciements les plus sincères.

Je tiens tout d'abord à exprimer ma profonde gratitude à mon directeur de recherche, le Professeur Leszek Szczecinski, pour son soutien inestimable et sa motivation tout au long de mes études de maîtrise. Je lui suis particulièrement reconnaissant pour ses conseils avisés et ses encouragements constants. Prof. Leszek a toujours fait preuve de patience dans le suivi de mes travaux et d'une grande compréhension à l'égard de ma situation personnelle.

Mes remerciements vont aussi aux membres du jury, Prof. Daniel Massicotte et Prof. Jacob Benesty, qui ont accepté d'évaluer ce mémoire et d'y consacrer une part précieuse de leur temps.

Je souhaite également adresser ma plus sincère gratitude à l'ensemble de l'équipe du Service des études supérieures et de la réussite étudiante de l'INRS, et tout particulièrement à Mme Isabelle Waltzing, dont le rôle a dépassé de loin celui d'une simple orthopédagogue. Par sa bienveillance lumineuse, son écoute généreuse et son accompagnement empreint d'humanité, elle a su m'offrir un appui inestimable à un moment où ma santé rendait chaque pas plus lourd. Sa présence rassurante et son engagement sincère ont laissé une empreinte profonde sur mon parcours.

Enfin, j'adresse ce mémoire à ma famille — à ma mère, lumière de ma vie, à mon père, à mon oncle, à ma sœur, ainsi qu'à tous mes proches. Une mention spéciale à Ayoub, Mohamed, Asma, Omrane et Abdoul Aziz, dont l'affection, la patience et le soutien sans faille m'ont été précieux à chaque étape de ce parcours. Votre générosité et votre présence m'ont porté, et je vous en suis profondément reconnaissant.

Je n'oublie pas toutes celles et ceux qui, par leur générosité, leurs encouragements ou leur aide ponctuelle, ont participé à mon cheminement. Leur soutien, sous toutes ses formes, a été d'une valeur inestimable et m'a offert une source inépuisable de force et de réconfort tout au long de cette aventure.

À vous tous, je dédie ce mémoire, avec toute l'affection que je vous porte.

Résumé

Ce travail présente un cadre bayésien pour l'estimation adaptative en ligne des filtres de Kalman appliqués aux modèles linéaires dans le traitement du signal. La dynamique des paramètres est modélisée par un processus vectoriel autorégressif affecté d'un bruit gaussien, et différentes structures de covariance (standard, scalaire et à variance fixe) sont étudiées et comparées pour analyser leur impact sur la performance du filtre. L'analyse comparative de ces différentes structures de filtre repose sur des simulations appliquées à l'identification de la réponse impulsionnelle acoustique.

Un accent particulier est mis sur plusieurs approches adaptatives d'estimation en ligne des hyper-paramètres, telles que le lissage temporel et l'estimation du maximum de vraisemblance (MLE). Les résultats obtenus montrent que ce cadre bayésien généralise avantageusement les méthodes classiques, en illustrant une robustesse et une adaptabilité supérieures, aussi bien dans des conditions expérimentales variées, des environnements bruités, que dans une grande diversité de scénarios d'application.

Mots-clés Filtre de Kalman, Modèles linéaires, Identification des systèmes, Filtrage bayésien, Traitement du signal, Estimation adaptative.

Abstract

This work introduces a Bayesian framework for online adaptive estimation of Kalman filters applied to linear models in signal processing. The parameter dynamics are modeled by a vector autoregressive process driven by Gaussian noise, and different covariance structures (standard, scalar, and fixed-variance) are studied and compared to assess their impact on filter performance. The comparative analysis of these various filter structures is based on simulations targeting the identification of acoustic impulse responses.

Several online adaptive methods for hyperparameter estimation are specifically explored, including both temporal smoothing and maximum likelihood approaches. The results confirm that the Bayesian framework not only generalizes existing methods but also demonstrates superior robustness and adaptability across diverse experimental settings, noisy contexts, and a broad spectrum of practical applications.

Keywords: Kalman Filter, Bayesian Filtering, Linear Models, System Identification, Signal Processing, Adaptive Estimation.

Table des matières

Remerciements	iii
Résumé	v
Abstract	vii
Table des matières	ix
Liste des figures	xi
Liste des tableaux	xiv
Liste des algorithmes	xvii
Liste des abréviations	xix
1 Introduction et motivation	1
1.1 Objectifs et problématique	2
1.2 Contributions	4
1.3 Positionnement par rapport à la littérature	5
1.4 Structure du document	6
2 Principe du filtre de Kalman	9
2.1 Formulation générique du filtre de Kalman	9
2.2 Approches variationnelles bayésiennes	12
2.3 Différentes variantes du filtre de Kalman	18
2.3.1 Filtre de Kalman : modèle à covariance vectorielle	18
2.3.2 Filtre de Kalman : modèle à covariance scalaire	20
2.3.3 Filtre de kalman simplifié à variance fixe	21
2.3.4 Descente de gradient stochastique	22
2.4 Expériences numériques	23
2.4.1 Identification de la réponse impulsionnelle	23
2.4.2 Méthodologie et résultats	26
2.4.3 Analyse comparative des scénarios simulés	30
2.4.3.1 Impact de la variation du coefficient α	31
2.4.3.2 Impact du changement de SNR	34
2.4.3.3 Impact du changement de l'erreur quadratique moyenne	35

2.5	Conclusion	37
3	Estimation des hyper-paramètres du modèle	39
3.1	Principe général de l'estimation par Maximum Likelihood	40
3.1.1	Formulation du problème	40
3.1.2	Calcul récursif de la densité prédictive	41
3.1.3	Motivation pour l'adoption du modèle dynamique	41
3.1.4	Dynamique stochastique des paramètres du filtre	42
3.1.5	Analyse préliminaire et sensibilité aux paramètres	43
3.1.5.1	Performance du filtre selon différents scénarios de paramétrisation	44
3.1.5.2	Performance du filtre pour $\varepsilon^* = 1 \cdot 10^{-12}$ et $\beta = 1$	48
3.1.5.3	Performance du filtre avec le modèle statique ($\varepsilon^* = 0$ et $\beta = 1$)	53
3.1.6	Optimisation de la variance de transition par descente de gradient	55
3.1.6.1	Expérimentation numérique	58
3.2	Estimation adaptative des hyper-paramètres par lissage temporel.	61
3.2.1	Analyse comparative avec le modèle statique ($\varepsilon^* = 0$ et $\beta = 1$)	64
3.3	Observations	67
3.4	Conclusion	68
4	Conclusion	69
	Références	71

Liste des figures

2.1	Distribution gaussienne multivariée bidimensionnelle.	11
2.2	Signal d'entrée, réponse impulsionnelle, signal bruité et signal convolué. . . .	25
2.3	Erreur quadratique moyenne \bar{m}_t en fonction des hyperparamètres ε et v_0 du filtre de Kalman (KF, Algorithme 2.1), pour $\bar{m}_\infty = -15$ dB, SNR = 5 dB, $\alpha = 0.9$ et $N = 1$ réalisation.	26
2.4	erreur quadratique moyenne m_t du filtre de Kalman simplifié SKF (Algorithme 2.3) pour différents réglages de ε et v_0 , avec $m_\infty = -15$ dB, SNR = 5 dB, $\alpha = 0.9$ et $N = 1$ réalisation.. . . .	27
2.5	Performance du filtre de Kalman simplifié à variance fixe fSKF (Algorithme 2.4) en termes d'erreur quadratique moyenne \bar{m}_t , mesurée pour plusieurs combinaisons de ε et v_0 , dans un scénario défini par $\bar{m}_\infty = -15$ dB, SNR = 5 dB, $\alpha = 0.9$ et $N = 1$ réalisation.	28
2.6	Erreur quadratique moyenne \bar{m}_t obtenue avec la descente de gradient stochastique, SGD (Algorithme 2.5) pour différentes valeurs de μ , avec $\bar{m}_\infty = -15$ dB, SNR = 5 dB, $\alpha = 0.9$ et $N = 1$ réalisation.	29
2.7	Convergence des algorithmes : KF, SKF, fSKF et SGD pour un (SNR) = 5 dB, $\alpha = 0.9$ et erreur quadratique moyenne cible $\bar{m}_\infty = -15$ dB pour 25 réalisations dans la moyenne.	30
2.8	Analyse de la sensibilité de l'erreur quadratique moyenne \bar{m}_t de la descente de Gradient Stochastique aux variations de pas adaptation μ , pour un cas d'étude où $\bar{m}_\infty = -15$ dB, SNR = 5 dB, et $\alpha = 0.99$	32
2.9	Convergence moyenne des algorithmes KF, SKF et SGD sur 25 trajectoires, pour un SNR de 5 dB, un coefficient $\alpha = 0.99$ et une erreur cible de $\bar{m}_\infty = -15$ dB.	33
2.10	Comparaison de la convergence des filtres KF, SKF, fSKF et SGD dans un cadre expérimental où SNR = 10 dB, $m_\infty = -15$ dB et $\alpha = 0.9$, avec des résultats moyens sur 25 simulations.	35
2.11	Convergence moyenne de \bar{m}_t pour KF, SKF, fSKF et SGD, avec SNR = 5 dB, $\alpha = 0.9$, et cible $\bar{m}_\infty = -25$ dB sur 25 trajectoires.	36
3.1	Courbe d'évolution de la moyenne glissante de la log-vraisemblance $\bar{\ell}_t$ obtenue pour différentes valeurs du bruit d'observation v_η dans le cadre du SKF, avec des paramètres initiaux du filtre fixés à $\varepsilon_0 = 10^{-10}$ et $v_0 = 8 \cdot 10^{-6}$. L'évaluation temporelle de \mathbf{h}_t est réalisée en considérant $\varepsilon^* = 10^{-10}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$	44

3.2	Évolution temporelle de l'erreur quadratique moyenne \bar{m}_t pour différentes valeurs du bruit d'observation v_η dans le cadre du SKF, avec des paramètres initiaux du filtre fixés à $\varepsilon_0 = 10^{-10}$ et $v_0 = 8 \cdot 10^{-6}$. L'évaluation temporelle de \mathbf{h}_t est réalisée en considérant $\varepsilon^* = 10^{-10}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$	45
3.3	Évolution de l'erreur de prédiction au carré e_t^2 au cours du temps pour différents v_η dans le cadre du SKF, avec des paramètres initiaux du filtre fixés à $\varepsilon_0 = 10^{-10}$ et $v_0 = 8 \cdot 10^{-6}$. L'évaluation temporelle de \mathbf{h}_t est réalisée en considérant $\varepsilon^* = 10^{-10}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$	46
3.4	Évolution de la moyenne glissante de la log-vraisemblance $\bar{\ell}_t$ pour différentes valeurs de la variance de transition ε , lors de l'évaluation temporelle de \mathbf{h}_t avec $\varepsilon^* = 10^{-10}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$, dans le cadre de l'application du filtre SKF.	47
3.5	Évolution de l'erreur quadratique moyenne \bar{m}_t pour différentes valeurs de la variance de transition ε , illustrant l'impact de ce paramètre sur la performance du filtre SKF, lors de l'évaluation temporelle de \mathbf{h}_t avec $\varepsilon^* = 10^{-10}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$	47
3.6	Évolution de l'erreur de prédiction au carré e_t^2 au cours du temps pour différentes valeurs de ε , lors de l'évaluation temporelle de \mathbf{h}_t avec $\varepsilon^* = 10^{-10}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$, dans le cadre de l'application du filtre SKF.	48
3.7	Évolution de la moyenne glissante de la log-vraisemblance $\bar{\ell}_t$ pour différentes valeurs du bruit d'observation v_η dans le cadre du SKF, avec des paramètres initiaux du filtre fixés à $\varepsilon_0 = 10^{-12}$ et $v_0 = 8 \cdot 10^{-6}$. L'évaluation temporelle de \mathbf{h}_t est réalisée en considérant $\varepsilon^* = 10^{-12}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$	49
3.8	Évolution temporelle de l'erreur quadratique moyenne \bar{m}_t pour différentes valeurs du bruit d'observation v_η dans le cadre du SKF, avec des paramètres initiaux du filtre fixés à $\varepsilon_0 = 10^{-12}$ et $v_0 = 8 \cdot 10^{-6}$. L'évaluation temporelle de \mathbf{h}_t est réalisée en considérant $\varepsilon^* = 10^{-12}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$	50
3.9	Évolution de l'erreur de prédiction au carré e_t^2 au cours du temps pour différents v_η dans le cadre du SKF, avec des paramètres initiaux du filtre fixés à $\varepsilon_0 = 10^{-12}$ et $v_0 = 8 \cdot 10^{-6}$. L'évaluation temporelle de \mathbf{h}_t est réalisée en considérant $\varepsilon^* = 10^{-12}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$	50
3.10	Évolution de la moyenne glissante de la log-vraisemblance $\bar{\ell}_t$ pour différentes valeurs de la variance de transition ε , lors de l'évaluation temporelle de \mathbf{h}_t avec $\varepsilon^* = 1 \cdot 10^{-12}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$, dans le cadre de l'application du filtre SKF.	51
3.11	Effet de la variance de transition ε sur \bar{m}_t lors de l'estimation temporelle de \mathbf{h}_t ($\varepsilon^* = 10^{-12}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$), dans le contexte d'utilisation du filtre SKF.	52
3.12	Évolution de l'erreur de prédiction au carré e_t^2 au cours du temps pour différents ε , lors de l'évaluation temporelle de \mathbf{h}_t avec $\varepsilon^* = 10^{-12}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$, dans le cadre de l'application du filtre SKF.	52
3.13	Évolution de la moyenne glissante de la log-vraisemblance $\bar{\ell}_t$ pour différents v_η dans le cadre du SKF, avec des paramètres initiaux du filtre fixés à $\varepsilon_0 = 10^{-10}$ et $v_0 = 8 \cdot 10^{-6}$, avec des données issues d'un modèle statique ($\varepsilon^* = 0$ et $\beta = 1$).	53

3.14	Évolution temporelle de l'erreur quadratique moyenne \bar{m}_t pour différentes valeurs du bruit d'observation v_η dans le cadre du SKF, avec des paramètres initiaux du filtre fixés à $\varepsilon_0 = 10^{-10}$ et $v_0 = 8 \cdot 10^{-6}$, avec des données issues d'un modèle statique ($\varepsilon^* = 0$ et $\beta = 1$).	54
3.15	Évolution de l'erreur de prédiction au carré e_t^2 au cours du temps pour différents v_η dans le cadre du SKF, avec des paramètres initiaux du filtre fixés à $\varepsilon_0 = 10^{-10}$ et $v_0 = 8 \cdot 10^{-6}$, avec des données issues d'un modèle statique ($\varepsilon^* = 0$ et $\beta = 1$).	54
3.16	Erreur quadratique moyenne \bar{m}_t pour différentes valeurs de μ_ξ lors de l'évaluation temporelle de \mathbf{h}_t avec $\varepsilon^* = 10^{-10}$ et $\beta = 1$, selon trois scénarios initiaux $\varepsilon_0 \in \{10^{-9}, 10^{-10}, 10^{-11}\}$ et $v_0 = 8 \cdot 10^{-6}$, dans le cadre de l'application du filtre SKF.	58
3.17	Évolution de la moyenne glissante de la log-vraisemblance $\bar{\ell}_t$ pour différentes valeurs de μ_ξ lors de l'évaluation temporelle de \mathbf{h}_t avec $\varepsilon^* = 10^{-10}$, selon deux scénarios initiaux $\varepsilon_0 \in \{10^{-9}, 10^{-11}\}$ et $v_0 = 8 \cdot 10^{-6}$	59
3.18	Évolution temporelle de ε_t pour différentes valeurs de μ_ξ lors de l'évaluation temporelle de \mathbf{h}_t avec $\varepsilon^* = 10^{-10}$, selon trois scénarios initiaux $\varepsilon_0 \in \{10^{-9}, 10^{-10}, 10^{-11}\}$ et $v_0 = 8 \cdot 10^{-6}$	59
3.19	Évolution temporelle de ε_t pour différentes valeurs de μ_ξ lors de l'évaluation temporelle de \mathbf{h}_t avec $\varepsilon^* = 10^{-10}$, selon trois scénarios initiaux $\varepsilon_0 \in \{10^{-9}, 10^{-10}, 3 \cdot 10^{-11}\}$ et $v_0 = 8 \cdot 10^{-6}$	60
3.20	Erreur quadratique moyenne \bar{m}_t , en comparant SKF avec différentes valeurs ε fixe et SKF avec ε estimée temporellement (SKF_temp) suivant l'équation (3.28), lors de l'évaluation temporelle de \mathbf{h}_t avec $\varepsilon^* = 10^{-10}$ et $\beta = 1$	62
3.21	Évolution de la moyenne glissante de la log-vraisemblance $\bar{\ell}_t$ pour différentes valeurs de ε fixe et SKF avec ε estimée temporellement (SKF_temp) suivant l'équation (3.28), lors de l'évaluation temporelle de \mathbf{h}_t avec $\varepsilon^* = 10^{-10}$ et $\beta = 1$	62
3.22	Erreur quadratique moyenne \bar{m}_t obtenue en comparant le SKF avec ε constant à une version où ε est estimé par lissage temporel (SKF_temp : équation (3.29), fenêtre $J = 10$), lors de l'évaluation temporelle de \mathbf{h}_t avec $\varepsilon^* = 10^{-10}$ et $\beta = 1$	63
3.23	Erreur quadratique moyenne \bar{m}_t , en comparant le SKF avec ε constant à une version de SKF avec ε estimé par lissage temporel (SKF_temp) selon l'équation (3.29), avec une fenêtre de lissage de taille $J = 100$, dans le cadre du modèle dynamique ($\varepsilon^* = 10^{-10}$ et $\beta = 1$).	64
3.24	Erreur quadratique moyenne \bar{m}_t , en comparant SKF avec différentes valeurs ε fixe et SKF avec ε estimée temporellement (SKF_temp) suivant l'équation (3.28). avec des données issues d'un modèle statique ($\varepsilon^* = 0$ et $\beta = 1$)	65
3.26	Erreur quadratique moyenne \bar{m}_t obtenue en comparant SKF avec ε constant à une version où ε est estimé par lissage temporel (SKF_temp : équation (3.29), fenêtre $J = 10$), avec des données issues d'un modèle statique ($\varepsilon^* = 0$ et $\beta = 1$).	65
3.25	Évolution de la moyenne glissante de la log-vraisemblance $\bar{\ell}_t$ pour différentes valeurs de ε fixe et SKF avec ε estimée temporellement (SKF_temp) suivant l'équation (3.28), avec des données issues d'un modèle statique ($\varepsilon^* = 0$ et $\beta = 1$)	66

3.27	Erreur quadratique moyenne \bar{m}_t comparant le SKF à ε constant et une estimation par lissage temporel SKF_temp ($J = 100$) selon (3.29), pour un modèle statique ($\varepsilon^* = 0$ et $\beta = 1$).	66
------	--	----

Liste des tableaux

2.1	Passage du modèle matriciel au modèle vectoriel diagonal.	19
2.2	Justification de la transition vers le modèle à covariance scalaire.	21
2.3	Paramètres des algorithmes que nous avons utilisé pour obtenir les résultats présentés dans la figure 2.7.	29
2.4	Paramètres des algorithmes que nous avons utilisé pour obtenir les résultats présentés dans la figure 2.9.	33
2.5	Paramètres des algorithmes que nous avons utilisé pour obtenir les résultats présentés dans la figure 2.10.	34
2.6	Paramètres utilisés pour la simulation avec une cible d'erreur $\overline{m}_\infty = -25$ dB.	36

Liste des algorithmes

2.1	Filtre de Kalman : Modèle à covariance matricielle (KF)	18
2.2	filtre de Kalman : modèle à covariance vectorielle (vKF)	19
2.3	filtre de Kalman : modèle à covariance scalaire (SKF)	21
2.4	Filtre de Kalman Simplifié à Variance Fixe (fSKF)	22
2.5	Stochastic Gradient Descent (SGD)	22

Liste des abréviations

KF	Filtre de Kalman
vKF	vector Kalman filter (filtre de Kalman à covariance vectorielle)
SKF	Simplified Kalman Filter (Filtre de Kalman Simplifié)
fSKF	Fixed Variance Simplified KF (Filtre de Kalman Simplifié à Variance Fixe)
SGD	Stochastic Gradient Descent (Descente de Gradient Stochastique)
SNR	Signal-to-Noise Ratio (Rapport Signal sur Bruit)
MAP	Maximum A Posteriori
PDF	Probability Density Function (Fonction de Densité de Probabilité)
KL	Kullback-Leibler (divergence)
ML	Machine Learning
VAR	Vecteur Autorégressif
MLE	Maximum Likelihood Estimation

Chapitre 1

Introduction et motivation

Le filtre de Kalman s'appuie sur la dynamique du système étudié, c'est-à-dire son évolution au cours du temps, afin d'améliorer la qualité des estimations en réduisant l'impact du bruit affectant les mesures. Le développement de méthodes visant à reconstituer l'état réel du système à partir de mesures bruitées représente ainsi un enjeu fondamental pour de nombreuses applications, ce qui a conduit à une production scientifique particulièrement abondante dans le domaine du traitement du signal [1, 2].

Ce travail ne s'inscrit pas dans la perspective d'une question de recherche empirique au sens strict, mais dans celle d'un problème d'estimation d'états au sein de systèmes dynamiques. La problématique centrale peut être formulée ainsi : à partir des mesures disponibles et des connaissances, des hypothèses portant sur la dynamique du système, comment élaborer un schéma d'estimation capable de fournir la meilleure approximation possible de l'état ou du signal recherché ? Dans le cadre de ce mémoire, cette interrogation se décline à travers la conception et l'analyse de versions simplifiées du filtre de Kalman, visant à concilier précision des estimations et réduction de la complexité algorithmique.

Une avancée majeure dans ce domaine a été réalisée en 1960, lorsque R.E. Kalman a publié son article fondateur intitulé "A New Approach to Linear Filtering and Prediction Problems" [3]. Le filtre de Kalman qu'il propose est une méthode mathématique permettant d'estimer de manière optimale l'état d'un système, y compris lorsque la modélisation ou les mesures comportent une part d'incertitude. [1, 2]. Cette approche repose sur l'utilisation de l'état précédent, des commandes appliquées et des observations bruitées pour produire une estimation actualisée. Par ailleurs,

le filtre de Kalman présente une propriété réursive remarquable: il ne nécessite pas l'historique complet des données passées pour calculer l'estimation courante [1,4]. Cette caractéristique le rend particulièrement adapté aux applications en temps réel, dans des environnements où la rapidité de traitement et l'adaptabilité sont essentielles.

Initialement développé pour la navigation aérospatiale [3,4], le filtre de Kalman est aujourd'hui largement utilisé pour le suivi et la prédiction de systèmes dynamiques dans des domaines variés, notamment la finance [5], la robotique [6], et les télécommunications [7,8].

Ce mémoire examine comment simplifier et adapter les filtres de Kalman pour améliorer leur efficacité dans le traitement des signaux à partir de modèles linéaires, avec un accent particulier sur l'amélioration de la précision des estimations et la vitesse de convergence dans des environnements dynamiques.

Les méthodes de filtrage classiques supposent souvent que les paramètres du système, tels que les variances du bruit de processus et du bruit d'observation, sont fixés ou connus à l'avance. Cependant, dans de nombreuses applications réelles, ces paramètres sont incertains et peuvent évoluer au cours du temps. Pour répondre à cette problématique, les filtres de Kalman adaptatifs permettent d'ajuster dynamiquement ces paramètres en fonction des données observées. Cette étude explore différentes formulations du filtre de Kalman, notamment les variantes simplifiées et à variance fixe, et compare leur performance avec des approches d'optimisation alternatives telles que la descente de gradient stochastique (SGD), dont les propriétés de convergence et les variantes pratiques ont été largement étudiées dans la littérature [9].

L'objectif principal de cette recherche est d'évaluer l'efficacité du filtrage de Kalman adaptatif dans les tâches d'identification des systèmes [10], en particulier pour l'estimation de la réponse impulsionnelle dans les modèles linéaires. En analysant différentes stratégies d'estimation des hyperparamètres et leur impact sur la convergence et la précision, cette étude vise à fournir des recommandations sur le choix optimal des techniques de filtrage pour les applications en temps réel.

1.1 Objectifs et problématique

L'estimation dynamique de paramètres dans les modèles linéaires demeure un défi central dans de nombreux domaines scientifiques et techniques, en particulier lorsque ces paramètres évoluent

au cours du temps ou sont soumis à des incertitudes importantes. Parmi les approches les plus efficaces, le **filtre de Kalman** et ses variantes offrent un cadre probabiliste optimal pour la mise à jour récursive des estimations en présence de bruit gaussien [3, 11, 12]. Néanmoins, la performance de ces méthodes dépend fortement du choix et de l'ajustement des hyper-paramètres, tels que les variances des bruits de processus et d'observation.

Dans ce contexte, ce mémoire vise à répondre aux questions suivantes :

- Comment différentes variantes du filtre de Kalman (standard, scalaire et scalaire à variance fixe) se comportent-elles en termes de convergence, de robustesse et de précision pour l'estimation dynamique des paramètres d'un modèle linéaire?
- Comment comparer les performances du filtre de Kalman à celles d'approches classiques d'optimisation, telles que la descente de gradient stochastique (SGD)?
- Quelles stratégies d'estimation bayésienne et d'ajustement adaptatif des hyper-paramètres peuvent être mises en œuvre pour améliorer la performance du filtrage en temps réel?

La problématique de ce travail se situe donc à l'interface entre modélisation probabiliste, estimation des hyper-parameters et optimisation adaptative [13, 14]. L'objectif principal est de proposer une méthodologie rigoureuse pour l'estimation adaptative des paramètres d'un modèle linéaire, en évaluant l'impact du choix des variantes du filtre de Kalman et des stratégies d'ajustement des hyper-paramètres sur la performance globale du modèle. Pour ce faire, nous combinons analyses théoriques et expérimentations numériques, en nous appuyant sur des simulations appliquées à l'identification de la réponse impulsionnelle acoustique, permettant ainsi d'illustrer concrètement les avantages et les limites de chaque approche.

Plus spécifiquement, ce travail cherche à :

- Proposer une formulation probabiliste unifiée du filtre de Kalman permettant d'incorporer l'évolution dynamique des paramètres dans un modèle linéaire à l'aide de différentes structures de covariance (matricielle, vectorielle, scalaire et scalaire à variance fixe) [13, 15].
- Mettre en œuvre et comparer plusieurs variantes du filtre de Kalman: le filtre standard (KF), le filtre simplifié (SKF), et le filtre simplifié à variance fixe (fSKF), en les comparant à une méthode d'optimisation classique, soit la descente de gradient stochastique (SGD) [16, 17].

- Valider expérimentalement l'efficacité des différentes méthodes proposées à travers des simulations appliquées à l'identification de la réponse impulsionnelle acoustique.
- Étudier l'impact du choix et de l'adaptation des hyper-paramètres sur la convergence, la robustesse et la précision de l'estimation dynamique.
- Développer et analyser des approches d'estimation adaptative en ligne des hyper-paramètres, fondées sur des principes bayésiens (Lissage temporelle [18], estimation du maximum de vraisemblance [14]).

Ce mémoire apporte ainsi une contribution à la problématique de l'adaptation dynamique des filtres de Kalman en proposant une méthodologie robuste pour ajuster les hyper-paramètres en fonction des propriétés du modèle. L'étude vise à démontrer que l'introduction d'un mécanisme adaptatif permet d'améliorer la qualité des estimations et la stabilité du filtre.

1.2 Contributions

Ce mémoire propose une approche pour l'optimisation des filtres de Kalman appliqués aux modèles linéaires, en mettant un accent particulier sur l'estimation dynamique des hyper-paramètres. Les principales contributions de cette étude s'articulent autour de plusieurs axes :

- **Analyse approfondie du filtre de Kalman et de ses variantes**

Ce travail propose une approche générique permettant d'intégrer différentes variantes du filtre de Kalman (standard, vectorielle, scalaire, à variance fixe) dans un cadre probabiliste unique [13, 15]. Cette formulation facilite la comparaison théorique et empirique des performances selon la structure de la covariance et le choix des paramètres, , mettant en évidence les conditions dans lesquelles chaque méthode offre une convergence optimale.

- **Comparaison avec des méthodes d'optimisation alternatives**

Afin d'évaluer la pertinence des variantes du filtre de Kalman, une comparaison a été effectuée avec l'algorithme de descente de gradient stochastique (SGD) [19].

- **Validation expérimentale sur des données simulées**

La méthodologie est appliquée à l'identification de la réponse impulsionnelle dans un contexte

acoustique, illustrant la capacité des différentes variantes à s'adapter à divers scénarios et niveaux de bruit.

- **Introduction de stratégies adaptatives pour l'estimation des hyper-paramètres**

Développement et évaluation de méthodes adaptatives en ligne pour l'estimation des hyper-paramètres. Plusieurs approches d'optimisation sont étudiées, telles que le lissage temporel [18] et l'estimation par maximum de vraisemblance (MLE) [14], afin d'identifier les hyper-paramètres optimaux et d'améliorer à la fois la robustesse et la convergence du filtre.

- **Modélisation dynamique des paramètres** Les paramètres du filtre de Kalman sont considérés comme variables dans le temps, ce qui permet de mieux représenter les systèmes réels dont les caractéristiques évoluent.

- **Analyse de l'impact des bruits d'observation et de transition**

Le travail étudie l'effet des différents niveaux de bruit sur la performance du filtre, montrant comment la variance du bruit d'observation et celle du processus influencent la précision et la stabilité du filtrage.

- **Comparaison entre scénarios statiques et dynamiques**

Le mémoire compare la performance du filtre entre des situations où les paramètres restent constants et d'autres où ils évoluent, et montre que les méthodes classiques ne suffisent pas toujours en contexte dynamique.

Les résultats obtenus dans cette étude démontrent que l'intégration d'un mécanisme adaptatif au sein des filtres de Kalman permet d'améliorer la précision des estimations et la robustesse des algorithmes face aux variations des conditions expérimentales. Ces contributions ouvrent la voie à de futures recherches sur l'optimisation des filtres de Kalman dans des applications où la dynamique du système est sujette à des changements non stationnaires.

1.3 Positionnement par rapport à la littérature

Depuis l'article fondateur de Kalman [3], le filtre de Kalman et ses variantes ont fait l'objet d'un développement théorique et d'applications multiples, couvrant des domaines tels que la navigation [4], la robotique [6], la finance [5] ou encore les télécommunications [7,8]. Plusieurs travaux [1,2,

11,12] ont établi le cadre probabiliste optimal du filtrage linéaire gaussien, tandis que des recherches plus récentes [14, 18] se sont concentrées sur l’adaptation dynamique des hyper-paramètres, notamment via le maximum de vraisemblance ou des techniques de lissage temporel.

Cependant, la littérature présente certaines limites :

- Les études comparent rarement de manière systématique les différentes structures de covariance (matricielle, vectorielle, scalaire, scalaire à variance fixe) dans un cadre unifié.
- Les validations expérimentales portent souvent sur des cas spécifiques et ne mesurent pas explicitement l’impact des choix de covariance et de stratégie adaptative sur la convergence et la précision dans des scénarios dynamiques.

Ce travail se distingue de la littérature existante par les contributions suivantes :

- Une formulation unifiée permettant d’intégrer plusieurs variantes du filtre de Kalman dans un même cadre probabiliste, afin de faciliter leur comparaison théorique et numérique.
- Une analyse expérimentale comparative entre ces variantes et la descente de gradient stochastique sur un problème d’identification de réponse impulsionnelle acoustique.
- Le développement et l’évaluation de stratégies adaptatives en ligne pour l’estimation des hyper-paramètres, avec une étude approfondie de leur impact sur la précision, la robustesse et la vitesse de convergence.

Ce positionnement permet de distinguer clairement notre contribution par rapport aux travaux existants, en offrant une vision comparative et intégrée des méthodes de filtrage et d’optimisation dans un contexte d’estimation adaptative.

1.4 Structure du document

Ce mémoire est structuré en quatre chapitres principaux, chacun abordant un aspect spécifique du problème étudié. Il suit une progression logique allant de la présentation du contexte et des concepts fondamentaux à l’analyse des résultats obtenus et à la conclusion.

- **Chapitre 1 : Introduction**

Ce premier chapitre introduit le contexte général du travail de recherche. Il expose la problématique et les motivations qui sous-tendent l'étude, ainsi que les objectifs poursuivis. Enfin, il présente un aperçu des contributions de ce travail et de la méthodologie employée.

- **Chapitre 2 : Principe du filtre de Kalman**

Ce chapitre présente la formulation générale du filtre de Kalman appliquée à l'identification des systèmes linéaires [15]. Il détaille les principales variantes du filtre, en soulignant leurs propriétés, leurs limites, ainsi que leur capacité d'adaptation à différents contextes et niveaux de bruit. Une attention particulière est portée au choix des paramètres et à leur influence sur la qualité des estimations.

- **Chapitre 3 : Estimation des hyper-paramètres du modèle**

Ce chapitre traite des méthodes d'estimation dynamique des hyper-paramètres du filtre de Kalman, notamment la variance du bruit de processus et celle du bruit d'observation [14, 18]. Il explore plusieurs approches adaptatives visant à améliorer la stabilité et la performance du filtre. L'évaluation de ces méthodes est réalisée à travers des simulations numériques.

- **Chapitre 4 : Conclusion et perspectives**

Le dernier chapitre synthétise les principaux résultats obtenus et met en perspective les contributions du travail. Il propose également des pistes de recherche future visant à améliorer les méthodes présentées et à explorer de nouvelles applications du filtre de Kalman dans d'autres contextes.

L'ensemble de ce document vise à démontrer l'importance d'une estimation adaptative des hyper-paramètres du filtre de Kalman et son impact sur la précision des estimations dans des modèles linéaires sous bruit gaussien dans le traitement du signal.

Chapitre 2

Principe du filtre de Kalman

Ce chapitre présente le cadre théorique du filtre de Kalman pour l'identification des systèmes linéaires bruités. Après avoir exposé la formulation générale et les hypothèses probabilistes associées, il décrit les principales variantes de l'algorithme et analyse leurs propriétés, avantages et limites. Une attention particulière est portée à l'influence des paramètres sur la précision des estimations, ainsi qu'aux liens avec certaines méthodes d'optimisation stochastique. L'application à l'estimation d'une réponse impulsionnelle acoustique illustre enfin la performance comparative de ces approches.

2.1 Formulation générique du filtre de Kalman

Dans les modèles linéaires en identification des systèmes, l'évolution des poids au cours du temps est souvent représentée par un processus auto-régressif vectoriel. Ce processus permet de capturer les relations temporelles entre les poids en fonction de leurs valeurs passées. Lorsqu'il est appliqué avec le filtre de Kalman, il offre une estimation optimale des états du système, en présence de bruit gaussien [20]. Cette approche permet donc de suivre et d'ajuster les paramètres en fonction des données observées, tout en minimisant l'erreur de prédiction:

$$\mathbf{w}_t = \mathbf{w}_{t-1} + \mathbf{u}_t \sqrt{\varepsilon_t}, \quad (2.1)$$

$$y_t = \mathbf{x}_t^\top \mathbf{w}_t + \eta_t, \quad (2.2)$$

où $\mathbf{w}_t \in \mathbb{R}^M$ désigne le vecteur des poids (ou paramètres) du système à l'instant t et \mathbf{u}_t est un vecteur de variables aléatoires gaussiennes indépendantes, de moyenne nulle et de variance unitaire. Ainsi, ε_t décrit la variance de la fluctuation aléatoire des poids entre les instants $t - 1$ et t . De plus, $y_t \in \mathbb{R}$ désigne l'observation réalisée au temps t , c'est-à-dire la valeur mesurée à cet instant et définie par l'équation (2.2) comme la somme du produit scalaire $\mathbf{x}_t^\top \mathbf{w}_t$ et du bruit d'observation η_t . Ce bruit η_t modélise l'incertitude ou les erreurs de mesure affectant y_t , et il est supposé suivre une loi normale centrée, $\eta_t \sim \mathcal{N}(0, v_\eta^*)$, où v_η^* désigne la variance du bruit d'observation.

La relation (2.1) peut être présentée comme suit :

$$f(\mathbf{w}_t | \mathbf{w}_{t-1}) = \mathcal{N}(\mathbf{w}_t; \mathbf{w}_{t-1}, \mathbf{I}\varepsilon_t), \quad (2.3)$$

La notation $\mathcal{N}(\mathbf{w}_t; \mathbf{w}_{t-1}, \mathbf{I}\varepsilon_t)$ désigne une loi normale multivariée, présentée en détail dans la section 2.1. Sa densité s'écrit :

$$\mathcal{N}(\mathbf{w}_t; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^M |\boldsymbol{\Sigma}|}} \exp \left(-\frac{1}{2} (\mathbf{w}_t - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{w}_t - \boldsymbol{\mu}) \right), \quad (2.4)$$

où, dans ce contexte, $\boldsymbol{\mu} = \mathbf{w}_{t-1}$ et $\boldsymbol{\Sigma} = \mathbf{I}\varepsilon_t$, avec \mathbf{I} la matrice identité. Cette écriture met en évidence que la dynamique des poids suit une loi normale multivariée conditionnelle, centrée sur \mathbf{w}_{t-1} et caractérisée par une covariance isotrope. Comme l'indique [1], souligne que dans les modèles linéaires adaptatifs, la matrice de covariance du bruit est fréquemment considérée comme une identité multipliée par un scalaire. Cette hypothèse traduit une variation indépendante et identiquement distribuée dans toutes les directions.¹

Afin de formaliser cette représentation. et d'en faciliter la manipulation dans les développements ultérieurs, il est utile de rappeler la définition générale de la distribution gaussienne multivariée.

Distribution gaussienne multivariée

En théorie des probabilités, la distribution gaussienne multivariée (ou normale multivariée) constitue un outil fondamental dans la modélisation des incertitudes. Elle est définie comme suit :

¹ Dans de nombreux cas, la covariance du bruit du processus est supposée isotrope, c'est-à-dire proportionnelle à la matrice identité, afin de simplifier les calculs et de refléter une incertitude uniforme dans toutes les dimensions. [11]

Définition 2.1 (Distribution gaussienne multivariée [21, Sec. 8.4]). La fonction de densité de probabilité d'une loi normale multivariée de moyenne $\boldsymbol{\mu} \in \mathbb{R}^M$ et de matrice de covariance $\mathbf{V} \in \mathbb{R}^{M \times M}$, notée $\mathcal{N}(\mathbf{w}; \boldsymbol{\mu}, \mathbf{V})$, est donnée par :

$$\mathcal{N}(\mathbf{w}; \boldsymbol{\mu}, \mathbf{V}) = \frac{1}{\sqrt{(2\pi)^M \det(\mathbf{V})}} \exp\left(-\frac{1}{2}(\mathbf{w} - \boldsymbol{\mu})^\top \mathbf{V}^{-1}(\mathbf{w} - \boldsymbol{\mu})\right). \quad (2.5)$$

Cette distribution est centrée autour du vecteur moyen $\boldsymbol{\mu}$, et la matrice de covariance \mathbf{V} régit la dispersion des probabilités autour de ce centre.

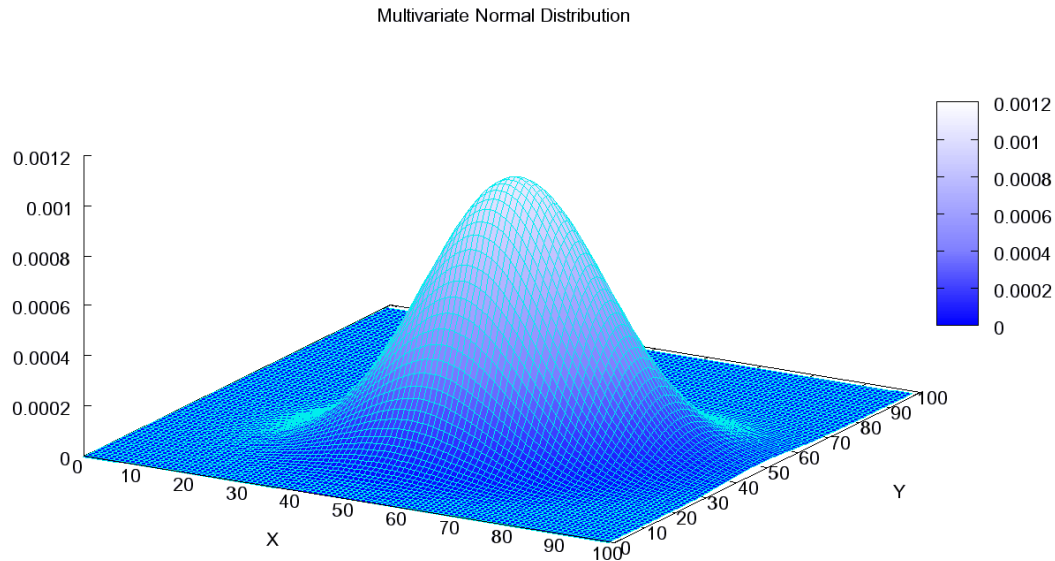


Figure 2.1: Distribution gaussienne multivariée bidimensionnelle.

Source: [22]

Comme illustré à la figure 2.1, la courbe est symétrique et sa forme (ellipsoïdale ou sphérique) dépend de la structure de \mathbf{V} . Une augmentation de la variance entraîne un aplatissement de la distribution sur les axes correspondants, alors qu'une faible variance concentre la masse de probabilité autour de $\boldsymbol{\mu}$. Dans tous les cas, l'intégrale de la densité sur \mathbb{R}^M est égale à 1.

Cette loi est omniprésente dans les modèles linéaires dynamiques. Par exemple, dans les modèles d'état utilisés en filtrage de Kalman, on suppose généralement que :

- La distribution a priori inconditionnelle de l'état initial \mathbf{w}_0 : $f(\mathbf{w}_0) = \mathcal{N}(\boldsymbol{\mu}_0, \mathbf{V}_0)$.
- Le bruit de prédiction est modélisé par $\mathbf{u}_t \sim \mathcal{N}(0, \varepsilon_t \mathbf{I})$.
- Le bruit d'observation est une variable aléatoire gaussienne : $\eta_t \sim \mathcal{N}(0, v_\eta^*)$.

où v_η^* représente la variance du bruit d'observation, c'est-à-dire l'intensité moyenne des erreurs de mesure affectant y_t . Ces hypothèses permettent des mises à jour récursives optimales des états via les équations du filtre de Kalman, tout en maintenant une forme analytique fermée pour les densités de probabilité à chaque instant t .

Les équations (2.1)-(2.2) sont couramment utilisées dans le domaine du traitement du signal, car il est fréquent de supposer que les entrées observées du système, \mathbf{x}_t , influencent la sortie y_t par une combinaison linéaire avec les variables d'état \mathbf{w}_t .

2.2 Approches variationnelles bayésiennes

En approche bayésienne, on vise à déterminer la distribution des poids \mathbf{w}_t conditionnée aux sorties y_t , en prenant en compte la dynamique temporelle des variables d'état modélisée par l'équation auto-régressive (2.1). L'objectif final est de calculer la distribution a posteriori de \mathbf{w}_t en fonction de toutes les observations disponibles jusqu'à l'instant t , soit $\{y_t\} = [y_1, \dots, y_{t-2}, y_{t-1}, y_t]$ [20, Ch. 12]

$$f(\mathbf{w}_t | \{y_t\}) = f(\mathbf{w}_t | \{y_{t-1}\}, y_t), \quad (2.6)$$

$$\propto \Pr\{y_t | \mathbf{w}_t\} \int f(\mathbf{w}_t, \mathbf{w}_{t-1} | \{y_{t-1}\}) d\mathbf{w}_{t-1}, \quad (2.7)$$

$$= \Pr\{y_t | \mathbf{w}_t\} \int f(\mathbf{w}_t | \mathbf{w}_{t-1}) f(\mathbf{w}_{t-1} | \{y_{t-1}\}) d\mathbf{w}_{t-1}. \quad (2.8)$$

Nous avons exploité la propriété markovienne du modèle [23], selon laquelle la connaissance de \mathbf{w}_{t-1} suffit à caractériser la distribution de \mathbf{w}_t . La relation (2.8) permet ainsi de mettre à jour récursivement la distribution $f(\mathbf{w}_t | \{y_t\})$ à partir de $f(\mathbf{w}_{t-1} | \{y_{t-1}\})$. Ce mécanisme constitue le fondement même de l'estimation des paramètres: à chaque nouvelle observation, la distribution des paramètres est réajustée en s'appuyant sur les résultats des étapes précédentes. C'est précisément

sur ce principe de mise à jour récursive de la distribution a posteriori ² que repose le **filtrage de Kalman**.

Pour $t = 1$, la récurrence (2.8) nécessite une distribution a priori, que nous supposons gaussienne.

$$f(\mathbf{w}_0|\{y_0\}) = f(\mathbf{w}_0) = \mathcal{N}(\mathbf{w}_0; \mathbf{0}, v_0 \mathbf{I}), \quad (2.9)$$

où v_0 , la variance a priori doit être ajustée aux données.

Afin de traiter (2.8), nous nous appuyons sur une représentation paramétrique gaussienne de $f(\mathbf{w}_t|\{y_t\})$, c'est-à-dire $\tilde{f}(\mathbf{w}_t|\{y_t\}) = \mathcal{N}(\mathbf{w}_t; \boldsymbol{\mu}_t, \mathbf{V}_t)$, ce qui nous permet de mettre en œuvre une approximation définie comme suit ³ :

$$\tilde{f}(\mathbf{w}_t|\{y_{t-1}\}) = \int f(\mathbf{w}_t|\mathbf{w}_{t-1}) \tilde{f}(\mathbf{w}_{t-1}|\{y_{t-1}\}) d\mathbf{w}_{t-1}, \quad (2.10)$$

$$\hat{f}(\mathbf{w}_t|\{y_t\}) \propto \Pr\{y_t|\mathbf{w}_t\} \tilde{f}(\mathbf{w}_t|\{y_{t-1}\}), \quad (2.11)$$

$$\tilde{f}(\mathbf{w}_t|\{y_t\}) = \mathcal{P} \left[\hat{f}(\mathbf{w}_t|\{y_t\}) \right], \quad (2.12)$$

où $\mathcal{P}[f(\mathbf{w})]$ est l'opérateur de projection $f(\mathbf{w})$ sur l'espace des distributions gaussiennes, parmi lesquelles nous considérons les formes possibles suivantes avec différents degrés de simplification :

$$\tilde{f}(\mathbf{w}_t|\{y_t\}) = \begin{cases} \mathcal{N}(\mathbf{w}_t, \boldsymbol{\mu}_t, \mathbf{V}_t) & \text{Modèle à covariance matricielle} \\ \mathcal{N}(\mathbf{w}_t, \boldsymbol{\mu}_t, \text{diag}_{\text{mat}}(\mathbf{v}_t)) & \text{Modèle à covariance vectorielle,} \\ \mathcal{N}(\mathbf{w}_t, \boldsymbol{\mu}_t, v_t \mathbf{I}) & \text{Modèle à covariance scalaire} \end{cases} \quad (2.13)$$

Où $\text{diag}_{\text{mat}}(\mathbf{v})$ est la matrice diagonale dont les éléments diagonaux sont rassemblés dans le vecteur \mathbf{v} .

La projection dans (2.12) consistera à trouver $\tilde{f}(\mathbf{w}_t|\{y_t\})$ qui minimise la divergence de Kullback-Leibler (KL) [25] par rapport à l'argument de projection $\hat{f}(\mathbf{w}_t|\{y_t\})$; ceci est réalisé en utilisant ce qui suit.

²L'estimation récursive présentée en (2.8) est dérivée sans formuler d'hypothèses particulières sur la nature des fonctions de densité de probabilité (PDF). Cette approche, bien reconnue dans la littérature ([20], Sec. 12.6), conduit au filtre de Kalman linéaire dans le cas où les relations entre les variables aléatoires sont linéaires et que l'ensemble des distributions considérées sont gaussiennes ([20], Sec. 13.2)

³Nous utilisons la relation $\mathcal{N}(\mathbf{w}; \boldsymbol{\mu}_1, \mathbf{V}_1) \mathcal{N}(\mathbf{w}; \boldsymbol{\mu}_2, \mathbf{V}_2) = \mathcal{N}(\mathbf{w}; \boldsymbol{\mu}_3, \mathbf{V}_3) \mathcal{N}(\boldsymbol{\mu}_1; \boldsymbol{\mu}_2, \mathbf{V}_1 + \mathbf{V}_2)$, [24, Ch. 8.4].

Proposition 1. La distribution $\tilde{f}(\mathbf{w}_t \mid \{y_t\})$, définie dans (2.13), est la plus proche de $\hat{f}(\mathbf{w}_t \mid \{y_t\})$ au sens de la divergence de [KL] [25], lorsque ses paramètres sont définis comme suit :

$$\boldsymbol{\mu}_t = \mathbb{E}[\mathbf{w}_t \mid \{y_t\}], \quad (2.14)$$

$$\mathbf{V}_t = \mathbb{E}[(\mathbf{w}_t - \boldsymbol{\mu}_t)(\mathbf{w}_t - \boldsymbol{\mu}_t)^\top \mid \{y_t\}], \quad (2.15)$$

$$\mathbf{v}_t = \text{diag}_{\text{ext}}(\mathbf{V}_t), \quad (2.16)$$

$$v_t = \frac{\text{Tr}(\mathbf{V}_t)}{M} = \frac{1}{M} \mathbf{1}^\top \mathbf{v}_t, \quad (2.17)$$

où $\text{diag}_{\text{ext}}(\mathbf{V})$ représente la diagonale de la matrice \mathbf{V} , et v_t dans (2.17) correspond à la moyenne arithmétique des éléments de \mathbf{v}_t , M désigne la dimension du vecteur d'état et $\text{Tr}(\cdot)$ représente la trace de la matrice.

Preuve : voir [13, Annexe A].

La récurrence approximative définie dans (2.10)–(2.12) s'explique par le fait que $\tilde{f}(\mathbf{w}_t \mid \{y_t\})$ est une loi gaussienne, telle que définie dans (2.13). En introduisant une variable gaussienne supplémentaire, comme indiqué dans l'équation (2.1), on obtient alors :

$$f(\mathbf{w}_t \mid \{y_{t-1}\}) = \mathcal{N}(\mathbf{w}_t; \boldsymbol{\mu}_{t,t-1}, \mathbf{V}_{t,t-1}). \quad (2.18)$$

Où

$$\boldsymbol{\mu}_{t,t-1} = \boldsymbol{\mu}_{t-1}, \quad (2.19)$$

$$\mathbf{V}_{t,t-1} = \mathbf{V}_{t-1} + \varepsilon_t \mathbf{I}. \quad (2.20)$$

Dans notre cas d'étude, tout le développement sera réalisé dans le domaine logarithmique, où les distributions gaussiennes sont représentées par des formes quadratiques de l'argument :

$$-\log \mathcal{N}(\mathbf{w}_t; \boldsymbol{\mu}_{t,t-1}, \mathbf{V}_{t,t-1}) \propto \frac{1}{2} (\mathbf{w}_t - \boldsymbol{\mu}_{t,t-1})^\top \mathbf{V}_{t,t-1}^{-1} (\mathbf{w}_t - \boldsymbol{\mu}_{t,t-1}). \quad (2.21)$$

Ainsi, nous procédons en deux étapes : (i) nous approchons $\ln \hat{f}(\mathbf{w}_t \mid \{y_t\})$ en utilisant une forme quadratique, dans laquelle (ii) nous identifions ensuite les termes correspondant à la moyenne et à la covariance.

Les étapes de l'approximation sont effectuées comme suit :

$$Q(\mathbf{w}) = -\ln \hat{f}(\mathbf{w}_t | \{y_t\}), \quad (2.22)$$

$$= \ell(\mathbf{x}_t^\top \mathbf{w}; y_t) - \ln \tilde{f}(\mathbf{w}_t | \{y_{t-1}\}), \quad (2.23)$$

$$\propto \ell(\mathbf{x}_t^\top \mathbf{w}; y_t) + \frac{1}{2}(\mathbf{w}_t - \boldsymbol{\mu}_{t,t-1})^\top \mathbf{V}_{t,t-1}^{-1}(\mathbf{w}_t - \boldsymbol{\mu}_{t,t-1}), \quad (2.24)$$

$$\approx \tilde{\ell}(\mathbf{x}_t^\top \mathbf{w}; y_t) + \frac{1}{2}(\mathbf{w}_t - \boldsymbol{\mu}_{t,t-1})^\top \mathbf{V}_{t,t-1}^{-1}(\mathbf{w}_t - \boldsymbol{\mu}_{t,t-1}), \quad (2.25)$$

$$\propto -\ln \tilde{f}(\mathbf{w}_t | \{y_t\}). \quad (2.26)$$

Cette approximation est obtenue en développant $\ell(\mathbf{x}_t^\top \mathbf{w}; y_t)$ en une série de Taylor tronquée autour de $\boldsymbol{\mu}_{t,t-1}$ (en se basant sur $\{y_t - 1\}$ [21, 24], on obtient l'estimation MAP ⁴ de \mathbf{w})

$$\tilde{\ell}(\mathbf{x}_t^\top \mathbf{w}; y_t) = \ell(\mathbf{x}_t^\top \boldsymbol{\mu}_{t,t-1}; y_t) + g_t \mathbf{x}_t^\top (\mathbf{w} - \boldsymbol{\mu}_{t,t-1}) + h_t (\mathbf{w} - \boldsymbol{\mu}_{t,t-1})^\top \mathbf{x}_t \mathbf{x}_t^\top (\mathbf{w} - \boldsymbol{\mu}_{t,t-1}), \quad (2.27)$$

où g_t et h_t sont respectivement les premières et deuxièmes dérivées de $\ell(z; y_t)$ par rapport à z ou $z = \mathbf{x}_t^\top \boldsymbol{\mu}_{t,t-1}$;

$$g_t = g(z; y_t) = \frac{d}{dz} \ell(z; y_t). \quad (2.28)$$

$$h_t = h(z; y_t) = \frac{d^2}{dz^2} \ell(z; y_t). \quad (2.29)$$

La convexité de $\ell(z; y_t)$ en z est assurée par le fait que $h_t > 0$.

À partir de cette approximation quadratique, nous identifions les paramètres de la distribution approchée en déterminant le mode et la matrice de covariance associée.

Le mode de la distribution a posteriori est obtenu en minimisant la fonction $Q(\mathbf{w})$, définie comme l'opposée du logarithme de la densité. Ce mode est donné par :

$$\boldsymbol{\mu}_t = \arg \min_{\mathbf{w}} Q(\mathbf{w}), \quad (2.30)$$

tandis que la matrice de covariance \mathbf{V}_t est obtenue en prenant l'inverse de la Hessienne de $Q(\mathbf{w})$, évaluée au point $\boldsymbol{\mu}_t$:

$$\mathbf{V}_t = \left[\nabla_{\mathbf{w}}^2 Q(\mathbf{w}) \right]^{-1} \Big|_{\mathbf{w}=\boldsymbol{\mu}_t}. \quad (2.31)$$

⁴Le MAP (Maximum A Posteriori) est une méthode d'estimation qui maximise la probabilité *a posteriori* basée sur les données observées et les connaissances *a priori* [21, 24].

Ainsi, la densité approchée $\hat{f}(\mathbf{w}_t|\mathbf{y}_t)$ peut être projetée sur la famille des distributions normales \mathcal{N} sous la forme :

$$\mathcal{P}[\hat{f}(\mathbf{w}_t|\mathbf{y}_t)] = \mathcal{N}(\mathbf{w}_t; \boldsymbol{\mu}_t, \mathbf{V}_t). \quad (2.32)$$

En remplaçant $\ell(\cdot; y_t)$ par $\tilde{\ell}(\cdot; y_t)$ dans $Q(\mathbf{w})$, le mode $\boldsymbol{\mu}_t$ est obtenu en annulant le gradient :

$$\nabla_{\mathbf{w}} Q(\mathbf{w})|_{\mathbf{w}=\boldsymbol{\mu}_t} = g_t \mathbf{x}_t + h_t \mathbf{x}_t \mathbf{x}_t^\top (\boldsymbol{\mu}_t - \mathbf{w}_0) + \mathbf{V}_t^{-1} (\boldsymbol{\mu}_t - \beta_t \boldsymbol{\mu}_{t-1}) = 0. \quad (2.33)$$

L'équation est résolue par :

$$\boldsymbol{\mu}_t = \mathbf{V}_t \left[\mathbf{x}_t \left(g(\mathbf{x}_t^\top \mathbf{w}_0; y_t) + h(\mathbf{x}_t^\top \mathbf{w}_0; y_t) \mathbf{x}_t^\top \mathbf{w}_0 \right) + \mathbf{V}_{t,t-1}^{-1} \boldsymbol{\mu}_{t-1} \right], \quad (2.34)$$

où :

$$\mathbf{V}_t = \left[h(\mathbf{x}_t^\top \mathbf{w}_0; y_t) \mathbf{x}_t \mathbf{x}_t^\top + \mathbf{V}_{t,t-1}^{-1} \right]^{-1}, \quad (2.35)$$

ce qui, en utilisant le lemme d'inversion de matrice [20, Sec. 4.11], donne :

$$\mathbf{V}_t = \mathbf{V}_{t,t-1} - \mathbf{V}_{t,t-1} \mathbf{x}_t \mathbf{x}_t^\top \mathbf{V}_{t,t-1} \frac{h(\mathbf{x}_t^\top \mathbf{w}_0; y_t)}{1 + h(\mathbf{x}_t^\top \mathbf{w}_0; y_t) \omega_t}. \quad (2.36)$$

En combinant (2.34) avec (2.36), on obtient :

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} - \mathbf{V}_{t,t-1} \mathbf{x}_t \frac{g_t + h_t \mathbf{x}_t^\top (\mathbf{w}_0 - \boldsymbol{\mu}_{t-1})}{1 + h(\mathbf{x}_t^\top \mathbf{w}_0; y_t) \omega_t}. \quad (2.37)$$

Une amélioration supplémentaire peut être envisagée en alternant l'équation (2.37) avec la réinitialisation $\mathbf{w}_0 \leftarrow \boldsymbol{\mu}_t$. Néanmoins, une seule itération avec $\mathbf{w}_0 = \boldsymbol{\mu}_{t-1}$ permet d'obtenir une mise à jour explicite de la moyenne et de la covariance des paramètres, constituant ainsi l'estimateur du filtre de Kalman. Donc, les mises à jour de la moyenne $\boldsymbol{\mu}_t$ et de la covariance \mathbf{V}_t s'écrivent :

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t,t-1} - \mathbf{V}_{t,t-1} \mathbf{x}_t \frac{g_t}{1 + h_t \omega_t}, \quad (2.38)$$

$$\mathbf{V}_t = \mathbf{V}_{t,t-1} - \mathbf{V}_{t,t-1} \mathbf{x}_t \mathbf{x}_t^\top \mathbf{V}_{t,t-1} \frac{h_t}{1 + h_t \omega_t}, \quad (2.39)$$

où $\omega_t = \mathbf{x}_t^\top \mathbf{V}_{t,t-1} \mathbf{x}_t$.

Étant donné \mathbf{w} et \mathbf{x}_t , la vraisemblance d'observer y_t sous l'hypothèse d'un bruit gaussien est :

$$p(y_t \mid \mathbf{w}, \mathbf{x}_t) = \mathcal{N}(y_t \mid \mathbf{w}^\top \mathbf{x}_t, v_\eta^*), \quad (2.40)$$

où la distribution normale \mathcal{N} est définie par :

$$p(y_t \mid \mathbf{w}, \mathbf{x}_t) = \frac{1}{\sqrt{2\pi v_\eta}} \exp \left(-\frac{(y_t - \mathbf{w}^\top \mathbf{x}_t)^2}{2v_\eta} \right). \quad (2.41)$$

En prenant le logarithme de la fonction de likelihood (2.41) et en utilisant la forme standard de la distribution gaussienne uni-variée, la fonction de perte $\ell(z, y_t)$ est alors donnée par :

$$\ell(z, y_t) = \frac{(\mathbf{w}^\top \mathbf{x}_t - y_t)^2}{2v_\eta}, \quad (2.42)$$

où y_t représente les cibles (sorties), \mathbf{x}_t sont les données d'entrée, et \mathbf{w} sont les poids du modèle linéaire.

- Première dérivée de la fonction scalaire $\ell(z, y_t)$

$$g_t = \frac{\partial}{\partial z_t} (\ell(z, y_t)) = \frac{(z_t - y_t)}{v_\eta} = \frac{-e_t}{v_\eta}, \quad (2.43)$$

où l'erreur de prédiction définie par $e_t = y_t - \mathbf{x}_t^\top \mathbf{w}_{t-1}$.

- Deuxième dérivée de la fonction scalaire $\ell(z, y_t)$

$$h_t = \frac{\partial^2}{\partial z_t^2} (\ell(z, y_t)) = \frac{1}{v_\eta}. \quad (2.44)$$

Algorithme 2.1. Filtre de Kalman : Modèle à covariance matricielle (KF)

```

1: Initialisation:  $\mu_0 = 0$ ,  $V_0 \leftarrow V_0 I$ 
2: for  $t = 1$  à  $T$  do
3:    $V_{t,t-1} \leftarrow V_{t-1} + \varepsilon_t I$  (Mise à jour de la covariance de transition)
4:    $\omega_t \leftarrow \mathbf{x}_t^\top V_{t,t-1} \mathbf{x}_t$  (Mise à jour de oméga  $\omega_t$ )
5:    $\mu_t \leftarrow \mu_{t-1} + \frac{V_{t,t-1} \mathbf{x}_t e_t}{v_\eta + \omega_t}$  (Mise à jour de la moyenne)
6:    $V_t \leftarrow V_{t,t-1} - \frac{V_{t,t-1} \mathbf{x}_t \mathbf{x}_t^\top V_{t,t-1}}{v_\eta + \omega_t}$  (Mise à jour de la covariance)
7: end for
8: return  $\mu$ 

```

2.3 Différentes variantes du filtre de Kalman

2.3.1 Filtre de Kalman : modèle à covariance vectorielle

Les lignes 2 à 7 de l'algorithme de Kalman à covariance matricielle peuvent être reformulées sous l'hypothèse que la matrice de covariance V_t est diagonale. Cette simplification consiste à représenter la covariance non plus comme une matrice pleine, mais comme un vecteur contenant uniquement les éléments diagonaux.

Cette hypothèse de diagonalité permet de remplacer toutes les opérations matricielles coûteuses (produits, inversions, etc.) par des opérations élément par élément (Hadamard), réduisant ainsi considérablement la complexité algorithmique et la charge de calcul. Pour rappel :

- $\mathbf{v}_t = \text{diag}_{\text{ext}}(V_t)$, où $\mathbf{v}_t \in \mathbb{R}^M$ regroupe les variances de chaque composante de \mathbf{w}_t ;
- $\text{diag}_{\text{mat}}(\mathbf{v}_t)$ désigne la matrice diagonale formée à partir du vecteur \mathbf{v}_t .

Le tableau ci-dessous présente les équations clés du modèle vectoriel (l'Algorithme 2.2) obtenues à partir du modèle matriciel (l'Algorithme 2.1).

Formule vectorielle	Passage matriciel
$\mathbf{v}_{t,t-1} = \mathbf{v}_{t-1} + \varepsilon_t \mathbf{1}$	$\text{diag}_{\text{mat}}(\mathbf{v}_{t-1} + \varepsilon_t \mathbf{1}) = \text{diag}_{\text{mat}}(\mathbf{v}_{t-1}) + \varepsilon_t \mathbf{I}$
$\omega_t = \mathbf{v}_{t,t-1}^\top (\mathbf{x}_t^2)$	$\mathbf{x}_t^\top \cdot \text{diag}_{\text{mat}}(\mathbf{v}_{t,t-1}) \cdot \mathbf{x}_t = \sum_j v_{t,t-1}^{(j)} x_t^{(j)^2}$
$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \frac{\mathbf{v}_{t,t-1} \odot \mathbf{x}_t \cdot e_t}{v_\eta + \omega_t}$	$\text{diag}_{\text{mat}}(\mathbf{v}_{t,t-1}) \cdot \mathbf{x}_t = \mathbf{v}_{t,t-1} \odot \mathbf{x}_t$
$\mathbf{v}_t = \mathbf{v}_{t,t-1} \odot \left(1 - \frac{\mathbf{v}_{t,t-1} \odot \mathbf{x}_t^2}{v_\eta + \omega_t}\right)$	$\left[\text{diag}_{\text{mat}}(\mathbf{v}_{t,t-1}) - \frac{\text{diag}_{\text{mat}}(\mathbf{v}_{t,t-1}) \mathbf{x}_t \mathbf{x}_t^\top \text{diag}_{\text{mat}}(\mathbf{v}_{t,t-1})}{v_\eta + \omega_t} \right]_{jj}$ $= v_{t,t-1}^{(j)} \left(1 - \frac{v_{t,t-1}^{(j)} x_t^{(j)^2}}{v_\eta + \omega_t}\right)$

Note : \odot désigne le produit élément par élément (Hadamard), \mathbf{x}_t^2 le vecteur des carrés des composantes de \mathbf{x}_t , $v_{t,t-1}^{(j)}$ le j -ième élément de $\mathbf{v}_{t,t-1}$.

Table 2.1: Passage du modèle matriciel au modèle vectoriel diagonal.

L'initialisation s'effectue selon : $\mathbf{v}_0 \leftarrow v_0 \cdot \mathbf{1}$.

Algorithme 2.2. filtre de Kalman : modèle à covariance vectorielle (vKF)

```

1: Initialisation:  $\mathbf{v}_0 \leftarrow v_0 \cdot \mathbf{1}$  ;  $\varepsilon \leftarrow \varepsilon \cdot \mathbf{1}$  ;  $\boldsymbol{\mu}_0 \leftarrow \mathbf{0}$ 
2: for  $t = 1$  à  $T$  do
3:    $\mathbf{v}_{t,t-1} \leftarrow \mathbf{v}_{t-1} + \varepsilon_t \mathbf{1}$  (Mise à jour de la covariance de transition)
4:    $\omega_t \leftarrow \mathbf{v}_{t,t-1}^\top (\mathbf{x}_t^2)$  (Mise à jour de oméga  $\omega$ )
5:    $\boldsymbol{\mu}_t \leftarrow \boldsymbol{\mu}_{t-1} + \mathbf{v}_{t,t-1} \odot \mathbf{x}_t \cdot \frac{e_t}{v_\eta + \omega_t}$  (Mise à jour de la moyenne)
6:    $\mathbf{v}_t \leftarrow \mathbf{v}_{t,t-1} \odot \left(1 - \frac{\mathbf{v}_{t,t-1} \odot |\mathbf{x}_t|^2}{v_\eta + \omega_t}\right)$  (Mise à jour de la covariance)
7: end for
8: return  $\mu$ 

```

En remplaçant la matrice \mathbf{V}_t par un vecteur \mathbf{v}_t , on obtient une version simplifiée du filtre de Kalman, qui :

- réduit la complexité de $\mathcal{O}(M^2)$ à $\mathcal{O}(M)$,
- facilite l'implémentation numérique.
- permet une interprétation élément par élément de la mise à jour des incertitudes.

L'algorithme vectoriel ainsi obtenu (voir Algo 2.2) offre un compromis efficace entre précision de l'estimation et coût computationnel, tout en conservant l'essentiel des propriétés du filtre matriciel. Cet algorithme est exactement équivalent à celui proposé dans [18].

2.3.2 Filtre de Kalman : modèle à covariance scalaire

Nous passons d'un modèle à covariance vectorielle $\mathbf{v}_t = \text{diag}_{\text{ext}}(\mathbf{V}_t)$ à un modèle scalaire en supposant que toutes les composantes de la variance sont identiques et égales à une moyenne v_t . Celle-ci est définie comme la moyenne des éléments diagonaux de la matrice \mathbf{V}_t :

$$v_t = \frac{\text{Tr}(\mathbf{V}_t)}{M} = \frac{1}{M} \mathbf{1}^\top \mathbf{v}_t. \quad (2.45)$$

On suppose que la covariance diagonale est homogène, c'est-à-dire que chaque composante de \mathbf{v}_t est identique :

$$v_t^{(j)} = v_t \quad \text{pour tout } j = 1, \dots, M \quad (2.46)$$

Alors, la moyenne des composantes de \mathbf{v}_t est :

$$\frac{1}{M} \sum_{j=1}^M v_t^{(j)} = \frac{1}{M} \sum_{j=1}^M v_t = \frac{1}{M} \cdot M \cdot v_t = v_t. \quad (2.47)$$

Ce résultat justifie la supposition :

$$\mathbf{v}_t = \frac{1}{M} \mathbf{1}^\top \mathbf{v}_t, \quad (2.48)$$

où $\mathbf{1}$ est un vecteur de M uns. Cette moyenne permet de remplacer le vecteur \mathbf{v}_t par le scalaire v_t dans les équations du filtre.

Ainsi, dans les équations du filtre de Kalman à covariance vectorielle, on remplace chaque vecteur de variances $\mathbf{v}_{t,t-1}$ par un vecteur homogène $v_{t,t-1} \cdot \mathbf{1}$, en supposant que toutes les composantes de la variance sont égales. Cette simplification permet de dériver les équations du modèle à covariance scalaire à partir de la version vectorielle.

Le tableau suivant résume les principales étapes de cette transition, en détaillant les équivalences formelles entre les deux modèles :

Formule scalaire	Passage depuis le modèle vectoriel
$v_t = \frac{1}{M} \mathbf{1}^\top \mathbf{v}_t$	On approxime $\mathbf{v}_t = \text{diag}_{\text{ext}}(\mathbf{V}_t)$ par une moyenne : $v_t = \frac{1}{M} \text{Tr}(\mathbf{V}_t)$ [13]
$\mathbf{v}_{t,t-1} = v_{t,t-1} \cdot \mathbf{1}$	Supposition d'homogénéité de la variance sur toutes les dimensions
$\omega_t = v_{t,t-1} \cdot \ \mathbf{x}_t\ ^2$	$\mathbf{v}_{t,t-1}^\top \mathbf{x}_t^2 = v_{t,t-1} \cdot \sum_j x_t^{(j)2}$
$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + v_{t,t-1} \mathbf{x}_t \cdot \frac{e_t}{v_\eta + \omega_t}$	$\mathbf{v}_{t,t-1} \odot \mathbf{x}_t = v_{t,t-1} \cdot \mathbf{x}_t$
$v_t = v_{t,t-1} \left(1 - \frac{1}{M} \cdot \frac{\omega_t}{v_\eta + \omega_t}\right)$	Moyenne des $v_t^{(j)}$ si tous égaux à v_t : $\frac{1}{M} \sum_j v_t^{(j)} = v_t$

Table 2.2: Justification de la transition vers le modèle à covariance scalaire.

En appliquant ces équations, on obtient l'algorithme du filtre de Kalman à covariance scalaire suivant, qui repose uniquement sur une mise à jour scalaire de la variance commune à toutes les dimensions :

Algorithme 2.3. filtre de Kalman : modèle à covariance scalaire (SKF)

```

1: Initialisation :  $\boldsymbol{\mu} \leftarrow \mathbf{0}$ ,  $v_{t,t-1} \leftarrow v_0$ 
2: for  $t = 1$  à  $T$  do
3:    $v_{t,t-1} \leftarrow v_{t-1} + \varepsilon$  (Mise à jour de la covariance de transition)
4:    $\omega_t = \|\mathbf{x}_t\|^2 v_{t,t-1}$  (Mise à jour de oméga  $\omega$ )
5:    $\boldsymbol{\mu}_t \leftarrow \boldsymbol{\mu}_{t-1} + v_{t,t-1} \mathbf{x}_t \cdot \frac{e_t}{v_\eta + \omega_t}$  (Mise à jour de la moyenne)
6:    $v_t \leftarrow v_{t,t-1} \left(1 - \frac{1}{M} \cdot \frac{\omega_t}{v_\eta + \omega_t}\right)$  (Mise à jour de la covariance)
7: end for
8: return  $\boldsymbol{\mu}$ 

```

2.3.3 Filtre de kalman simplifié à variance fixe

Une autre simplification est obtenue si l'on suppose que la variance $v_{t,t-1}$ est constante dans le temps t , c'est-à-dire $v_{t,t-1} \equiv \bar{v}$. Nous obtenons alors l'algorithme du filtre de Kalman simplifié à

variance fixe (fSKF) :

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \bar{v} \mathbf{x}_t \cdot \frac{e_t}{v_\eta + \bar{\omega}_t}, \quad (2.49)$$

Où l'initialisation nécessite de définir \bar{v} .

Algorithme 2.4. Filtre de Kalman Simplifié à Variance Fixe (fSKF)

```

1: Initialisation :  $\boldsymbol{\mu} \leftarrow \mathbf{0}$  ,  $\bar{v} = cte$ 
2: for  $t = 1$  à  $T$  do
3:    $\bar{\omega}_t \leftarrow \|\mathbf{x}\|^2 \bar{v}$  (Calcul de oméga  $\bar{\omega}$ )
4:    $\boldsymbol{\mu}_t \leftarrow \boldsymbol{\mu}_{t-1} + \bar{v} \mathbf{x}_t \cdot \frac{e_t}{v_\eta + \bar{\omega}_t}$  (Mise à jour de la moyenne)
5: end for
6: return  $\boldsymbol{\mu}$ 

```

2.3.4 Descente de gradient stochastique

Enfin, en supposant que $h_t \equiv 0$, (2.49) puisse être réécrit comme

$$\mu_t \leftarrow \mu_{t-1} - \bar{v} x_t g_t. \quad (2.50)$$

ce qui est équivalent à l'algorithme de gradient stochastique (SGD)⁵ avec un pas d'adaptation égal à la variance a posteriori \bar{v} .

Algorithme 2.5. Stochastic Gradient Descent (SGD)

```

1: Initialisation : Pas d'adaptation  $\bar{v}$ 
2: for  $t = 1$  à  $T$  do
3:    $\boldsymbol{\mu}_t \leftarrow \boldsymbol{\mu}_{t-1} + \bar{v} \mathbf{x}_t \cdot \frac{e_t}{v_\eta}$  (Mise à jour de la moyenne)
4: end for
5: return  $\boldsymbol{\mu}$  (Estimation finale du paramètre)

```

⁵Dans l'approche du gradient stochastique (SGD), le gradient de $\ell(\mu_{t-1}^\top x_t; y_t)$ est calculé, multiplié par l'étape d'adaptation et soustrait de la solution disponible actuellement μ_{t-1} ([24] Ch. 3.1.); c'est ce qui est fait dans 2.50.

2.4 Expériences numériques

2.4.1 Identification de la réponse impulsionnelle

Dans ce mémoire, nous nous intéressons à la problématique du filtrage linéaire appliqué à l'identification des systèmes acoustiques [10]. Contrairement aux méthodes traditionnelles qui visent sur la reconstruction directe du signal de sortie $y_t \in \mathbb{R}$, notre approche se concentre sur l'estimation de la *réponse impulsionnelle* du système \mathbf{h} . Cette réponse constitue une description fondamentale du système, puisqu'elle encode la manière dont celui-ci réagit à un signal d'entrée $\mathbf{x}_t \in \mathbb{R}^M$ à travers un vecteur de paramètres optimaux $\mathbf{w} \in \mathbb{R}^M$. Autrement dit, elle permet de caractériser dynamiquement le système en modélisant la relation linéaire entre les entrées et les sorties dans un cadre bruité.

Cette estimation repose sur des méthodes adaptatives permettant de suivre l'évolution temporelle des paramètres du système. En effet, dans un contexte acoustique, les propriétés du canal peuvent varier au cours du temps. L'identification de la réponse impulsionnelle s'inscrit alors dans une approche paramétrique, où les poids \mathbf{w} sont ajustés en temps réel à partir des observations disponibles, tout en tenant compte du bruit présent dans les données. Cette stratégie permet d'optimiser la précision des estimations et de mieux comprendre le comportement du système dans un environnement réel. Cette approche s'inspire de travaux classiques en filtrage adaptatif [16] [15] et en estimation bayésienne appliquée à l'identification acoustique [26], tout en intégrant des formulations récentes basées sur le filtre de Kalman [13] [15].

Plus précisément, nous considérons le problème de l'identification de la réponse impulsionnelle acoustique, visant à caractériser la réponse d'un système acoustique. Dans ce contexte, le signal d'entrée x_t est généré selon le processus auto-régressif suivant :

$$x_t = \alpha x_{t-1} + u_t, \quad (2.51)$$

où α est le coefficient d'atténuation, fixé à 0.9, et u_t représente un bruit blanc gaussien de moyenne nulle et de variance unitaire.

La réponse impulsionnelle $\mathbf{h} = [h_0, h_1 \dots, h_{M-1}]^\top$ avec une longueur $M = 600$, montrée dans la Fig 2.2, est calculée à l'aide du logiciel "audiolabs/rir-generator" [27] pour une pièce de dimensions

(5, 4, 6) m, la source étant en position (2, 3.5, 2) m, le récepteur en position (2, 1.5, 1) m, avec une fréquence d'échantillonnage de 8 kHz et un temps de réverbération de 225 ms [28]. La sortie désirée est obtenue comme

$$y_t = \mathbf{h}^\top \mathbf{x}_t + \eta_t, \quad (2.52)$$

où η_t est un bruit gaussien de moyenne nulle avec une variance v_η^* , et

$$\mathbf{x}_t = \begin{bmatrix} x_t, x_{t-1}, \dots, x_{t-M+1} \end{bmatrix}^\top.$$

Nous définissons le rapport signal sur bruit (SNR) comme

$$\text{SNR} = 10 \log_{10} \left(\frac{\mathbb{E} [|\mathbf{h}^\top \mathbf{x}_t|^2]}{v_\eta^*} \right) \text{ [dB]}, \quad (2.53)$$

où $\mathbb{E}[\cdot]$ désigne l'espérance prise par rapport à la distribution de \mathbf{x}_t . Dès lors, pour une valeur donnée du SNR et de l'espérance $\mathbb{E}[|\mathbf{h}^\top \mathbf{x}_t|^2]$, il est possible de déterminer v_η^* en inversant l'équation (2.53)

Notre objectif est d'identifier la réponse impulsionnelle \mathbf{h} . Comme l'équation d'observation (2.2) correspond exactement au modèle (2.52), nous interprétons le résultat du filtrage adaptatif \mathbf{w}_t comme une estimation de \mathbf{h} , sous l'hypothèse que les dimensions de \mathbf{h} et de \mathbf{w} sont identiques. La qualité de cette estimation est mesurée par une erreur quadratique moyenne :

$$m_t = \frac{\|\mathbf{w}_t - \mathbf{h}\|^2}{\|\mathbf{h}\|^2}, \quad (2.54)$$

ou par sa moyenne :

$$\bar{m}_t = \mathbb{E}[m_t], \quad (2.55)$$

où, en pratique, l'espérance est calculée en moyennant m_t sur N réalisations indépendantes du bruit η_t et du signal d'entrée \mathbf{x}_t ; et $\|\cdot\|^2$ désigne la norme euclidienne au carré

Comme indiqué dans l'équation (2.1), l'ensemble des algorithmes étudiés vise à estimer l'évolution temporelle de l'état \mathbf{w}_t . Dans notre contexte, la réponse impulsionnelle \mathbf{h} est supposée constante, c'est-à-dire invariante dans le temps, de sorte que $\mathbf{h}_t = \mathbf{h}$ pour tout t . Ainsi, pour des choix de paramètres appropriés, l'erreur quadratique moyenne converge vers une valeur faible, notée \bar{m}_∞ .

Le choix des paramètres des algorithmes implique un compromis entre la rapidité de convergence et la valeur limite de l'erreur quadratique moyenne, \bar{m}_∞ . Afin de comparer objectivement les différentes méthodes, la convergence est évaluée en fixant un seuil cible pour l'erreur quadratique moyenne. Plus précisément, nous réalisons des simulations sur une grille de paramètres et sélectionnons ceux permettant d'atteindre la cible fixée, $\bar{m}_\infty = -15$ dB.

Enfin, si l'on considère \mathbf{w}_t comme une estimation de \mathbf{h} , on a alors $\mathbb{E}[\mathbf{w}_t] = \mathbf{h}$

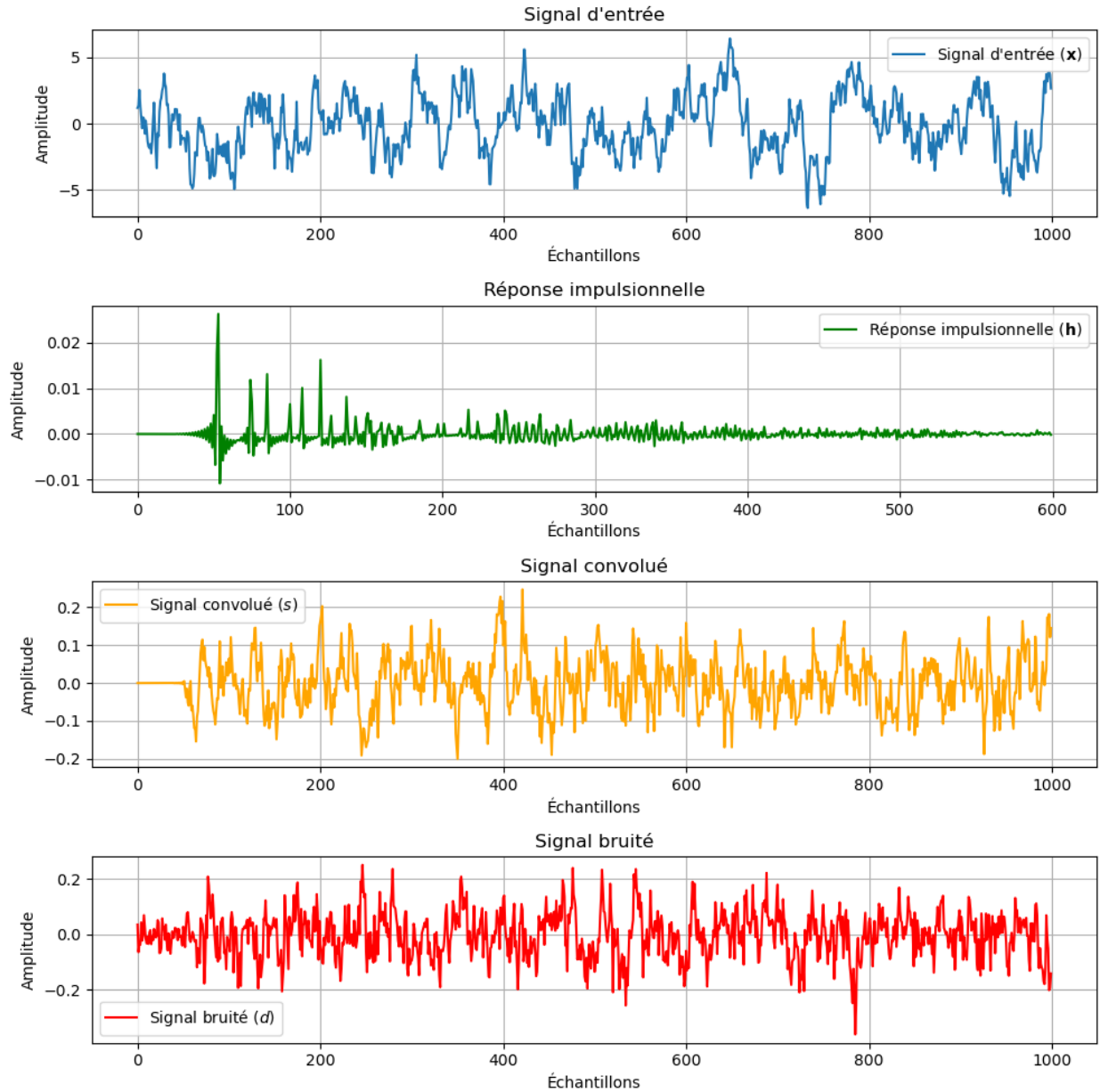


Figure 2.2: Signal d'entrée, réponse impulsionnelle, signal bruité et signal convolué.

2.4.2 Méthodologie et résultats

Dans cette étude, nous avons appliqué les différentes variantes du filtre de Kalman pour estimer la réponse impulsionnelle dans un modèle linéaire avec un bruit gaussien.

L'efficacité du filtre a été évaluée dans un premier temps en fonction des paramètres principaux propres à chaque algorithme, du coefficient d'atténuation α défini à l'équation (2.50), de l'erreur quadratique moyenne, du rapport signal sur bruit (SNR) et du nombre d'échantillons (T). Les simulations sont réalisées initialement avec un rapport signal sur bruit (SNR) fixé à 5 dB, un coefficient d'atténuation $\alpha = 0.9$ et une erreur quadratique moyenne cible $\bar{m}_\infty = -15$ dB. Chaque méthode est étudiée sous différents scénarios en variant systématiquement ses paramètres spécifiques.

1. Filtre de Kalman (KF)

Dans cette analyse, nous évaluons l'impact de la variation des paramètres ε et v_0 (variance initiale) sur la performance d'un filtre de Kalman appliqué à l'estimation de la réponse impulsionnelle dans un modèle linéaire.

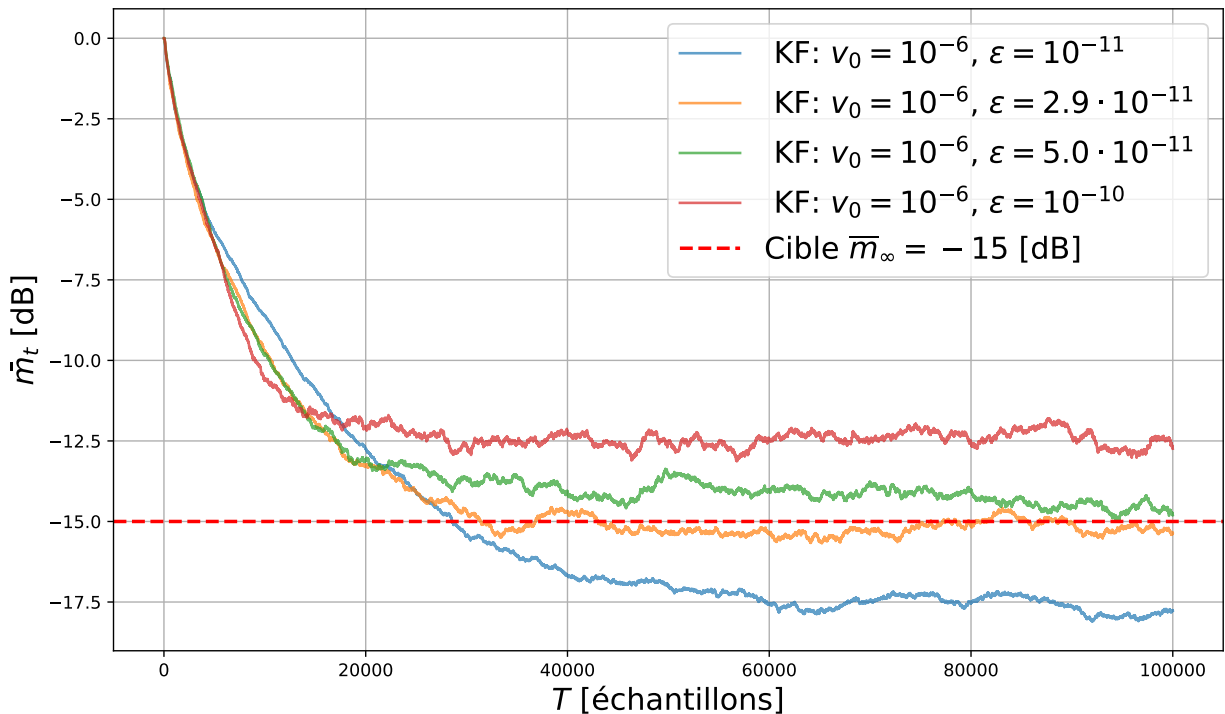


Figure 2.3: Erreur quadratique moyenne \bar{m}_t en fonction des hyperparamètres ε et v_0 du filtre de Kalman (KF, Algorithme 2.1), pour $\bar{m}_\infty = -15$ dB, SNR = 5 dB, $\alpha = 0.9$ et $N = 1$ réalisation.

2. Filtre de Kalman Simplifié (SKF)

Ensuite, nous analysons l'effet des variations des paramètres ε et v_0 (variance initiale) sur les performances d'un filtre de Kalman simplifié utilisé pour l'estimation de la réponse impulsionnelle dans un modèle linéaire.

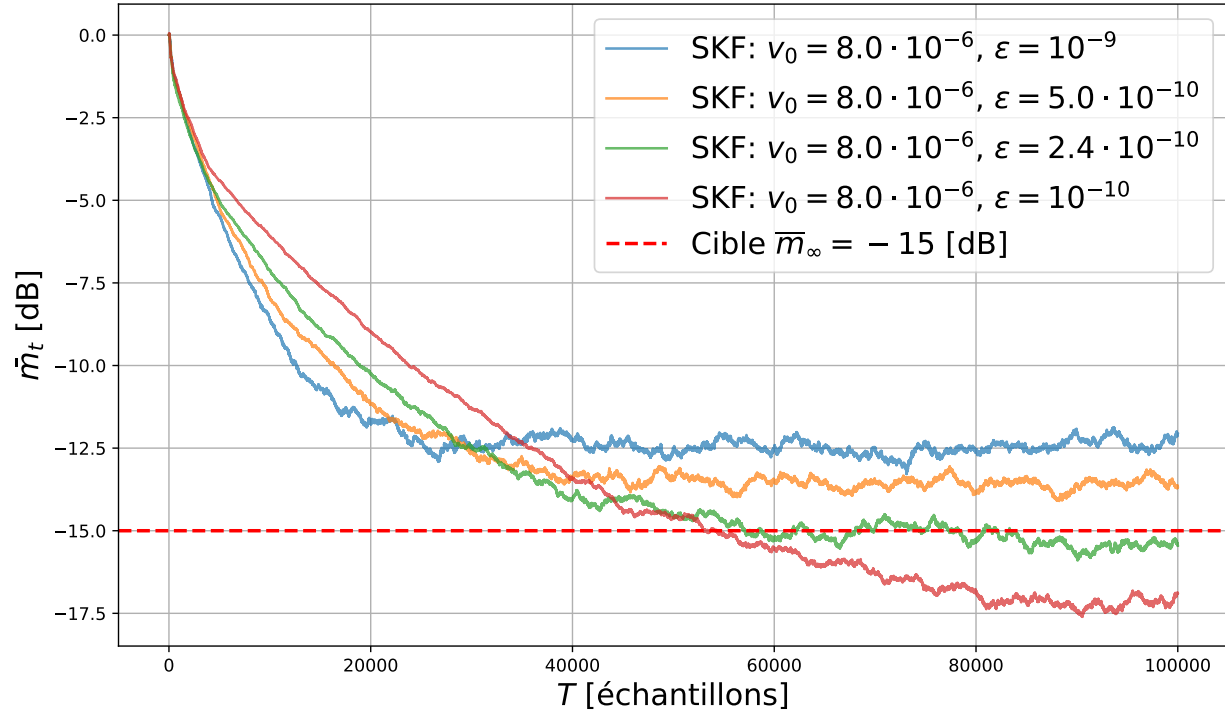


Figure 2.4: erreur quadratique moyenne m_t du filtre de Kalman simplifié SKF (Algorithme 2.3) pour différents réglages de ε et v_0 , avec $m_\infty = -15$ dB, SNR = 5 dB, $\alpha = 0.9$ et $N = 1$ réalisation..

3. Filtre de Kalman Simplifié à Variance Fixe (fSKF)

Nous analysons ensuite l'effet des variations du paramètre \bar{v} (variance fixe) sur les performances d'un filtre de Kalman simplifié à variance fixe utilisé pour l'estimation de la réponse impulsionnelle dans un modèle linéaire.

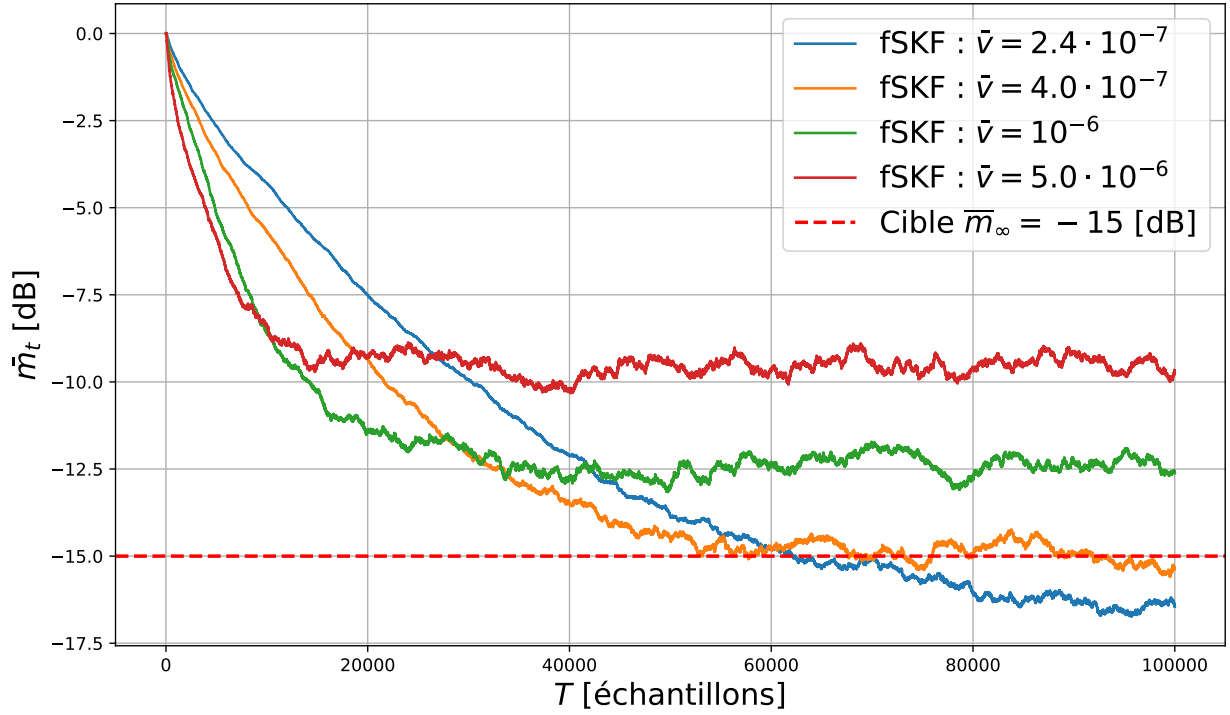


Figure 2.5: Performance du filtre de Kalman simplifié à variance fixe fSKF (Algorithme 2.4) en termes d'erreur quadratique moyenne \bar{m}_t , mesurée pour plusieurs combinaisons de \tilde{v} et v_0 , dans un scénario défini par $\bar{m}_\infty = -15$ dB, $\text{SNR} = 5$ dB, $\alpha = 0.9$ et $N = 1$ réalisation.

4. Descente de Gradient Stochastique (SGD)

Cette section évalue l'impact des variations du paramètre $\bar{v} \equiv \mu$ sur l'efficacité de la descente de gradient stochastique appliquée à l'estimation de la réponse impulsionnelle pour un modèle linéaire.

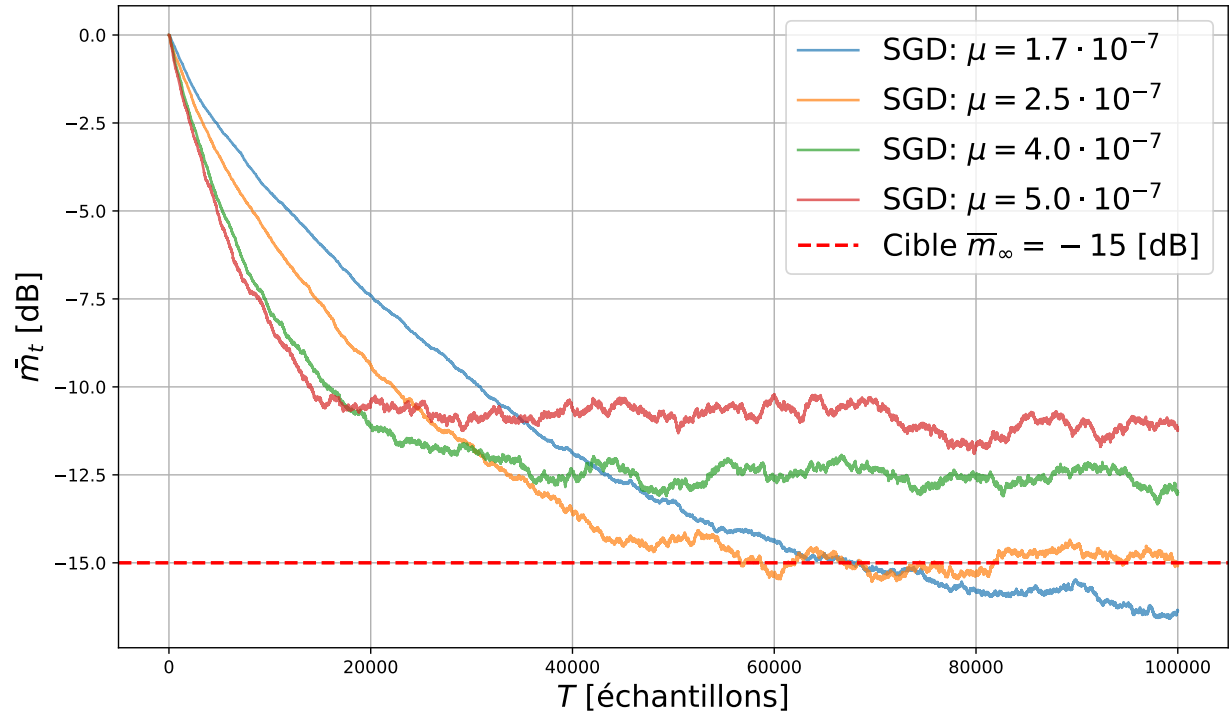


Figure 2.6: Erreur quadratique moyenne \bar{m}_t obtenue avec la descente de gradient stochastique, SGD (Algorithme 2.5) pour différentes valeurs de μ , avec $\bar{m}_\infty = -15$ dB, $\text{SNR} = 5$ dB, $\alpha = 0.9$ et $N = 1$ réalisation.

Les résultats pour les différents algorithmes sont illustrés dans la Fig. 2.7. Ainsi que, les paramètres utilisés pour obtenir ces courbes sont indiqués dans le Tableau 2.3 à des fins de reproductibilité.

Algorithmes	Paramètres
SGD	$\mu = 2.5 \cdot 10^{-7}$
KF	$\varepsilon = 2.9 \cdot 10^{-11}$ $v_0 = 1.10^{-6}$
SKF	$\varepsilon = 2.4 \cdot 10^{-10}$ $v_0 = 1.10^{-6}$
fSKF	$\bar{v} = 4 \cdot 10^{-7}$

Table 2.3: Paramètres des algorithmes que nous avons utilisé pour obtenir les résultats présentés dans la figure 2.7.

La figure 2.7 illustre la convergence des différents algorithmes en fonction de l'évolution de l'erreur quadratique moyenne \bar{m}_t .

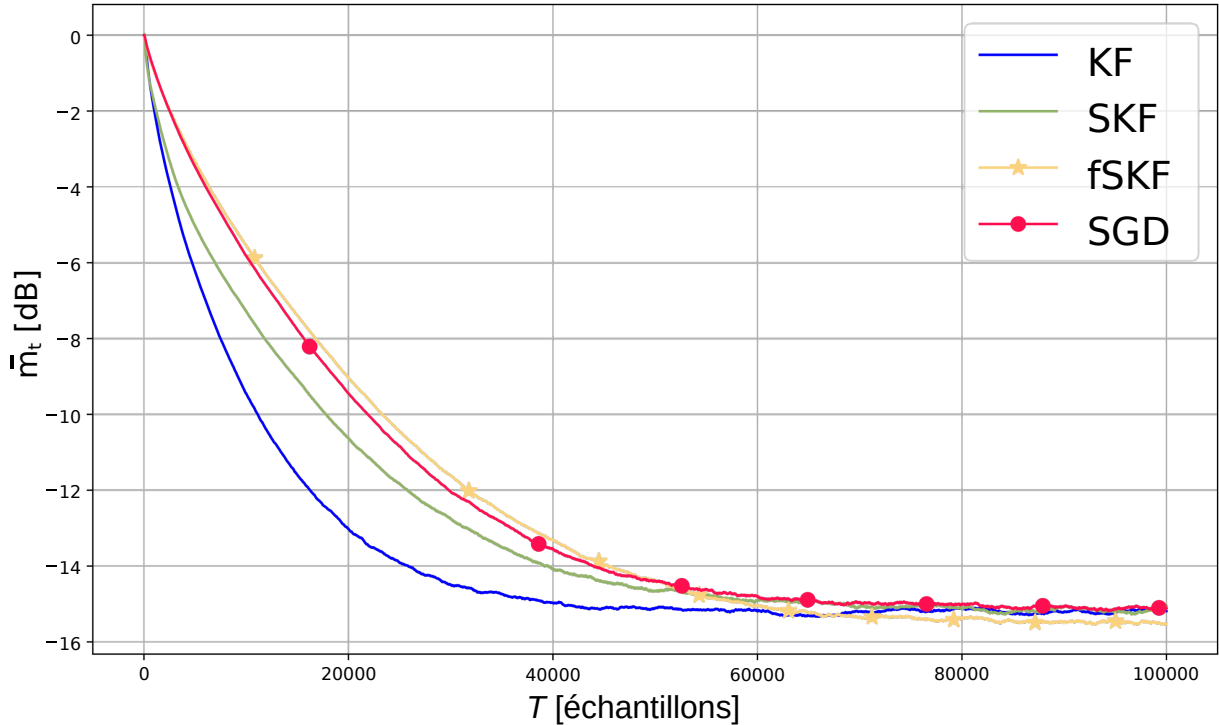


Figure 2.7: Convergence des algorithmes : KF, SKF, fSKF et SGD pour un $(\text{SNR}) = 5$ dB, $\alpha = 0.9$ et erreur quadratique moyenne cible $\bar{m}_\infty = -15$ dB pour 25 réalisations dans la moyenne.

Les analyses sont réalisées avec un nombre d'échantillons fixé à $T = 100\,000$ pour l'ensemble des méthodes, et une réponse impulsionnelle de taille $M = 600$, où T désigne le nombre d'échantillons et M la longueur mémoire du système. Le rapport signal sur bruit (SNR) est fixé à 5 dB, avec un coefficient d'atténuation $\alpha = 0.9$, et une erreur quadratique moyenne cible de $m_\infty = -15$ dB. Les résultats présentés correspondent à une moyenne calculée sur $N = 25$ réalisations indépendantes.

2.4.3 Analyse comparative des scénarios simulés

À la suite des expérimentations initiales, nous avons réalisé une série supplémentaire de simulations afin d'analyser rigoureusement l'effet de la variation de certains paramètres environnementaux clés sur les performances des algorithmes étudiés. Ces paramètres incluent notamment le coefficient d'atténuation α , signal sur bruit (SNR) et l'erreur quadratique moyenne cible \bar{m}_∞ . L'objectif principal de ces simulations est d'évaluer l'impact de ces variations sur les métriques de performance

considérées.

Les simulations initiales, précédemment définies dans la section 2.4.2, avaient été effectuées en utilisant les valeurs suivantes : un coefficient d'atténuation $\alpha = 0.9$, signal sur bruit (SNR) de 5 dB et une erreur quadratique moyenne cible de -15 dB. Dans cette nouvelle phase, notre étude se focalise sur la manière dont la modification de ces paramètres influence le comportement global du système.

Afin d'approfondir notre compréhension des réactions des différents algorithmes sous des conditions variées, nous proposons une analyse comparative structurée autour de trois scénarios spécifiques :

- Le premier scénario étudie les effets induits par la variation du coefficient d'atténuation α .
- Le deuxième scénario se concentre sur l'analyse des impacts du rapport signal sur bruit (SNR) sur les performances globales des algorithmes.
- Le troisième scénario examine l'influence d'une modification de la cible d'erreur quadratique moyenne \bar{m}_∞ sur les résultats obtenus.

Cette méthodologie comparative permettra d'obtenir une vision complète et détaillée de l'influence de ces paramètres sur les performances des algorithmes étudiés.

2.4.3.1 Impact de la variation du coefficient α

En faisant varier le coefficient, initialement fixé à 0.9 dans l'équation (2.51), nous procédons à une nouvelle série de simulations pour ajuster les paramètres de chaque algorithme. La valeur testée dans cette étude est de 0.99, afin d'observer précisément son influence sur la convergence et la stabilité des algorithmes.

Nous suivons le processus méthodologique décrit dans la section précédente, qui consiste à déterminer les ensembles de paramètres permettant aux algorithmes d'atteindre l'erreur quadratique moyenne cible de -15 dB avec un SNR fixé à 5 dB. Une fois ces paramètres identifiés, nous comparons la vitesse de convergence afin d'évaluer les performances respectives des algorithmes.

Lors du lancement des simulations utilisant l'algorithme SGD, nous n'avons pas réussi à atteindre l'erreur quadratique moyenne cible de -15 dB. Cette dynamique est illustrée dans la figure 2.8.

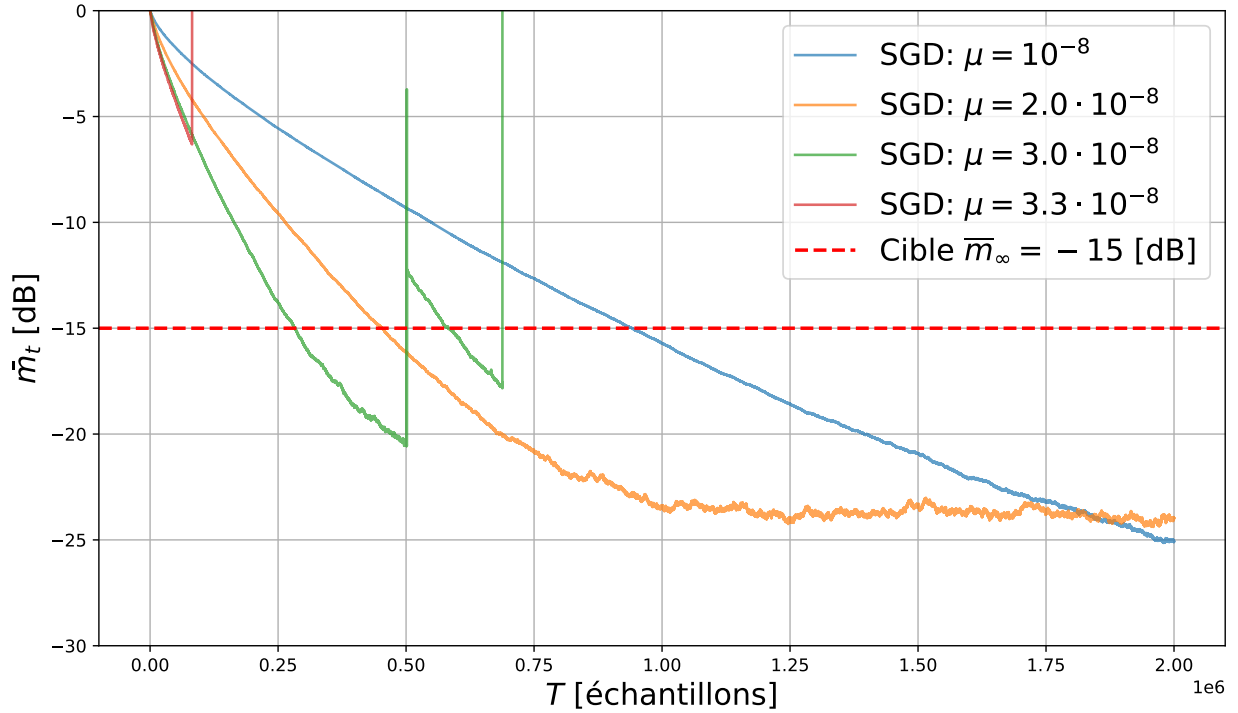


Figure 2.8: Analyse de la sensibilité de l’erreur quadratique moyenne \bar{m}_t de la descente de Gradient Stochastique aux variations de pas adaptation μ , pour un cas d’étude où $\bar{m}_\infty = -15$ dB, SNR = 5 dB, et $\alpha = 0.99$.

Nous observons que la convergence est atteinte pour une erreur instantanée avoisinant -25 dB. Cependant, il n’a pas été possible d’obtenir une convergence pour des erreurs plus faibles ($\bar{m}_\infty = -15$ dB). Cela est dû au fait qu’en augmentant le pas d’adaptation (step size μ), l’algorithme devient instable et diverge, comme le montre la courbe verte. Ce comportement met en évidence un compromis entre la rapidité de convergence et la stabilité de l’algorithme utilisé.

Les résultats pour les différents algorithmes sont illustrés dans la Fig. 2.9. Les paramètres optimaux identifiés sont regroupés dans le tableau 2.4, assurant ainsi la reproductibilité de nos résultats.

Algorithmes	Paramètres
KF	$\varepsilon = 1.5 \cdot 10^{-11}$ $v_0 = 1.10^{-6}$
SKF	$\varepsilon = 3 \cdot 10^{-9}$ $v_0 = 8.10^{-6}$
fSKF	$\bar{v} = 1.4 \cdot 10^{-6}$

Table 2.4: Paramètres des algorithmes que nous avons utilisé pour obtenir les résultats présentés dans la figure 2.9.

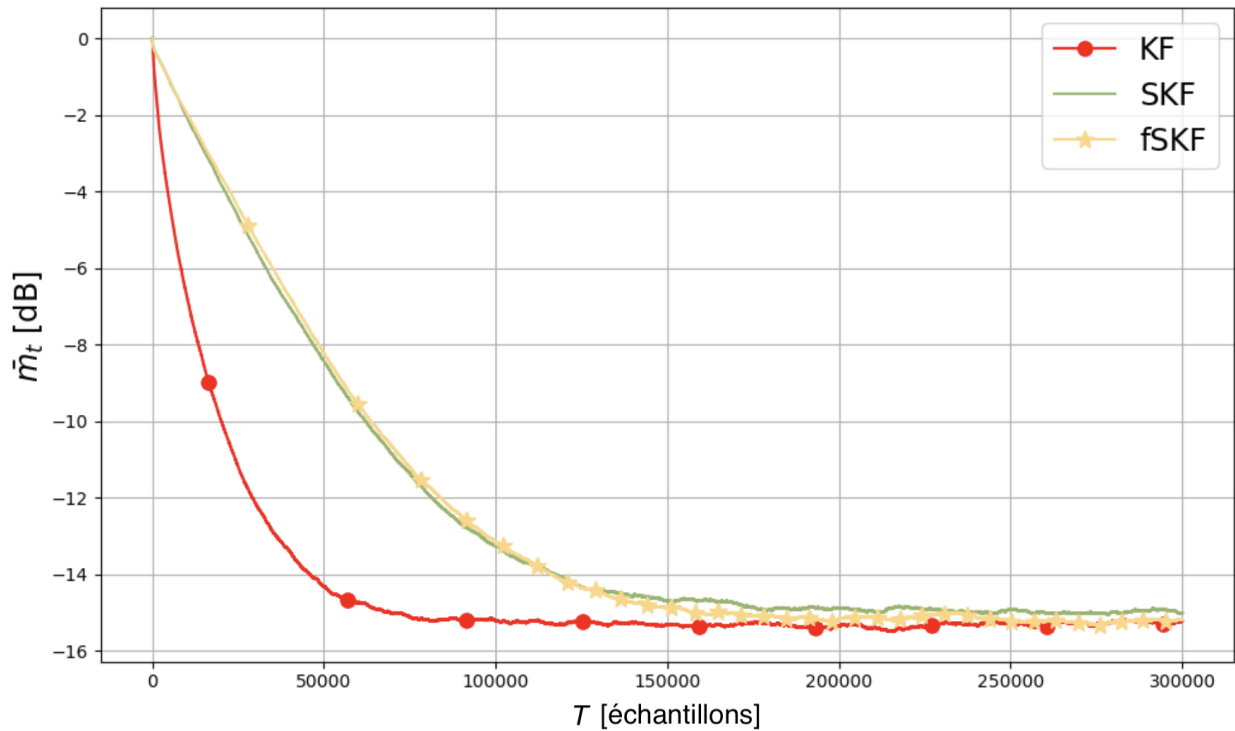


Figure 2.9: Convergence moyenne des algorithmes KF, SKF et SGD sur 25 trajectoires, pour un SNR de 5 dB, un coefficient $\alpha = 0.99$ et une erreur cible de $\bar{m}_\infty = -15$ dB.

La figure 2.9 présente la convergence des différents algorithmes à travers l'évolution de \bar{m}_t . Pour garantir la stabilité des résultats, les simulations ont été effectuées sur $T = 300000$ échantillons, en considérant une réponse impulsionnelle de longueur $M = 600$. Le rapport signal sur bruit (SNR) est fixé à 5 dB, avec un coefficient d'atténuation $\alpha = 0.99$ et une cible d'erreur quadratique moyenne $m_\infty = -15$ dB. Les résultats sont moyennés sur $N = 25$ réalisations indépendantes.

2.4.3.2 Impact du changement de SNR

Dans cette série de simulations, nous examinons l'effet d'une augmentation du rapport signal sur bruit (SNR) de 5 dB à 10 dB sur les performances des différents algorithmes étudiés. L'objectif principal est d'évaluer comment cette amélioration du SNR influence la précision et la convergence vers une erreur quadratique moyenne cible de -15 dB.

Afin de garantir la cohérence des résultats, nous conservons les paramètres environnementaux précédemment définis dans la section 2.4.2, initialement calibrés pour un rapport signal sur bruit (SNR) de 5 dB, $\alpha = 0.9$ et $\bar{m}_\infty = -15$ dB.

Les résultats obtenus sont présentés à la figure 2.10, tandis que les paramètres utilisés pour chaque algorithme figurent dans le tableau 2.5.

Algorithmes	Paramètres
SGD	$\mu = 1.8 \cdot 10^{-7}$
KF	$\varepsilon = 1.1 \cdot 10^{-10} \quad v_0 = 1.10^{-6}$
SKF	$\varepsilon = 1.68 \cdot 10^{-9} \quad v_0 = 8.10^{-6}$
fSKF	$\bar{v} = 1 \cdot 10^{-6}$

Table 2.5: Paramètres des algorithmes que nous avons utilisé pour obtenir les résultats présentés dans la figure 2.10.

La figure 2.10 illustre la convergence des différents algorithmes en fonction de l'évolution de \bar{m}_t , pour un rapport signal sur bruit (SNR) fixé à 10 dB et un coefficient d'autocorrélation $\alpha = 0.9$. L'erreur quadratique moyenne cible est de $\bar{m}_\infty = -15$ dB, moyennée sur $N = 25$ réalisations indépendantes. Pour garantir la convergence tout en limitant le coût computationnel, le nombre d'échantillons a été restreint à $T = 60000$.

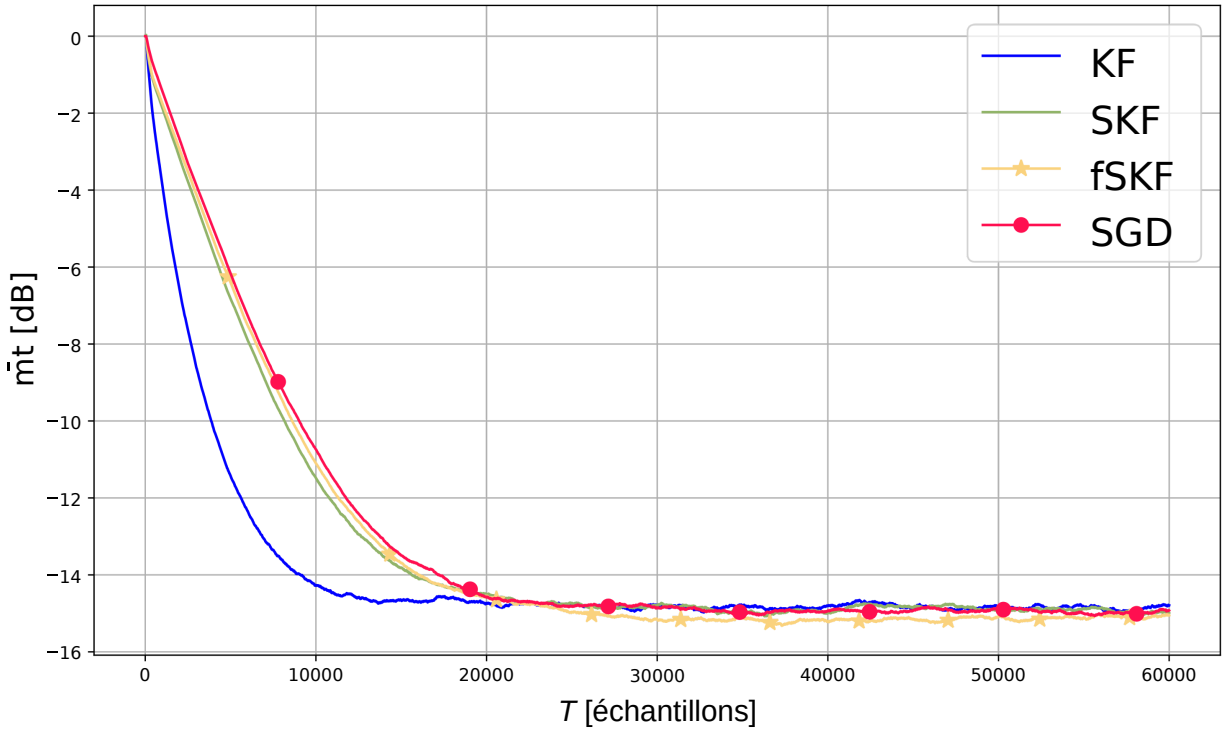


Figure 2.10: Comparaison de la convergence des filtres KF, SKF, fSKF et SGD dans un cadre expérimental où $\text{SNR} = 10 \text{ dB}$, $m_\infty = -15 \text{ dB}$ et $\alpha = 0.9$, avec des résultats moyens sur 25 simulations.

L'amélioration observée s'explique par le fait qu'une augmentation du rapport signal sur bruit (SNR) réduit l'incertitude des mesures, ce qui accélère la convergence du filtre vers l'erreur quadratique moyenne cible. Cette rapidité de convergence est particulièrement manifeste pour $\text{SNR} = 10 \text{ dB}$, où les filtres atteignent le même niveau d'erreur quadratique moyenne en un nombre d'itérations plus faible. Cela illustre clairement l'influence bénéfique d'un environnement moins bruité sur la vitesse d'adaptation des filtres.

2.4.3.3 Impact du changement de l'erreur quadratique moyenne

Cette série d'expériences vise à analyser l'impact d'un renforcement des exigences de performance, en abaissant la cible d'erreur quadratique moyenne de -15 dB à -25 dB . L'objectif est d'évaluer la capacité des différents algorithmes à converger vers un seuil de précision plus strict, sans modification des paramètres d'apprentissage environnementaux précédemment identifiés dans la section 2.4.2, à savoir un SNR de 15 dB , un coefficient d'atténuation $\alpha = 0.9$ et une erreur quadratique moyenne cible de -15 dB . Pour assurer la comparabilité des résultats, les simula-

tions ont été réalisées en conservant les paramètres $\alpha = 0.9$ et un rapport signal sur bruit (SNR) de 5 dB, initialement calibrés pour une erreur cible $\bar{m}_\infty = -15$ dB..

Les performances obtenues sont présentées dans la figure 2.11, tandis que les valeurs des paramètres utilisés figurent dans le tableau 2.6.

Algorithmes	Paramètres
SGD	$\mu = 3.3 \cdot 10^{-8}$
KF	$\varepsilon = 2.7 \cdot 10^{-13} \quad v_0 = 1 \cdot 10^{-6}$
SKF	$\varepsilon = 3 \cdot 10^{-12} \quad v_0 = 8 \cdot 10^{-6}$
fSKF	$\bar{v} = 3.5 \cdot 10^{-8}$

Table 2.6: Paramètres utilisés pour la simulation avec une cible d'erreur $\bar{m}_\infty = -25$ dB.

La figure 2.11 illustre la convergence des différents algorithmes en fonction de l'évolution de \bar{m}_t .

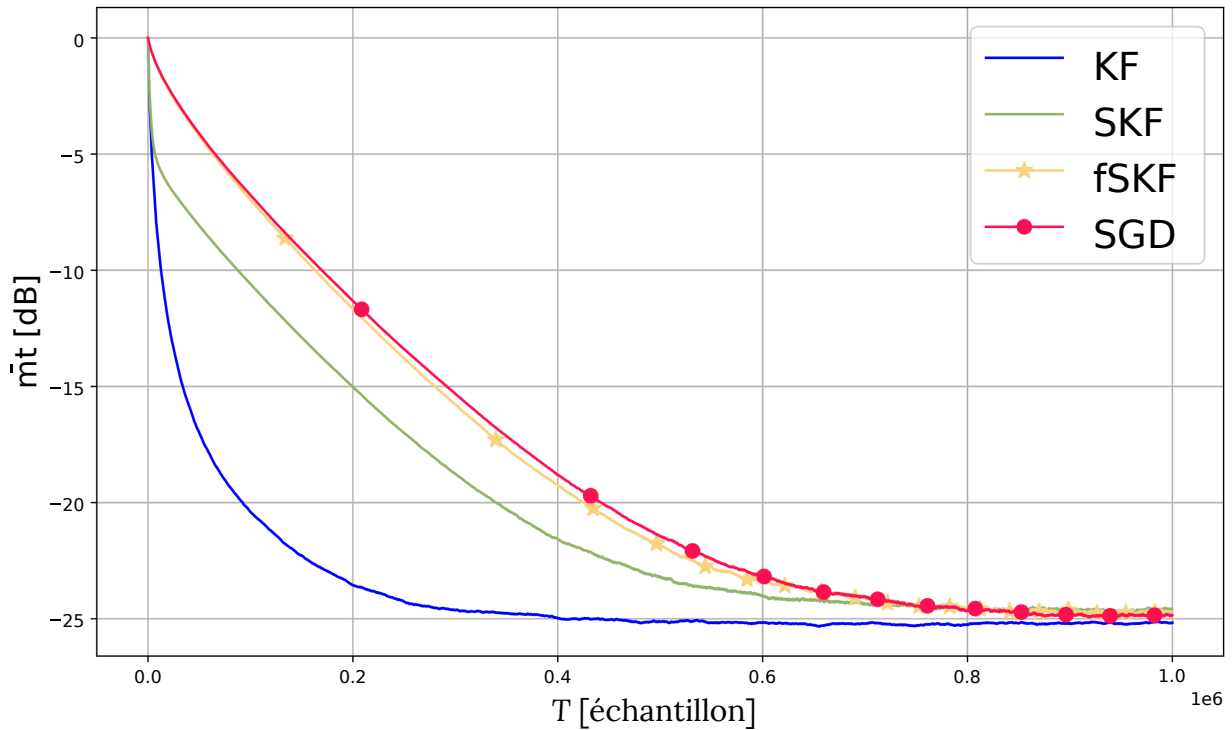


Figure 2.11: Convergence moyenne de \bar{m}_t pour KF, SKF, fSKF et SGD, avec SNR = 5 dB, $\alpha = 0.9$, et cible $\bar{m}_\infty = -25$ dB sur 25 trajectoires.

L'interprétation de cette amélioration réside dans l'ajustement des paramètres d'apprentissage afin de répondre à une exigence de précision plus stricte. En particulier, l'utilisation de valeurs d' ϵ plus faibles permet de réduire l'intensité du bruit de processus. En addition, l'augmentation du nombre d'échantillons T sur lesquels la moyenne est effectuée améliore la robustesse des estimations instantanées. Ces ajustements favorisent ainsi une convergence plus progressive mais plus fiable vers une erreur moyenne cible de -25 dB. Ils illustrent la nécessité de configurations paramétriques adaptées pour maintenir la stabilité et la performance dans des conditions d'exigence accrues.

2.5 Conclusion

Ce chapitre a posé les fondements théoriques du filtre de Kalman appliqué aux modèles linéaires à bruit gaussien, en détaillant sa formulation probabiliste, son interprétation bayésienne, et ses différentes variantes selon la structure de la matrice de covariance [13]. À partir d'une modélisation autorégressive des poids, nous avons vu comment le filtre permet une estimation récursive et optimale de ces derniers, tout en intégrant l'information issue des observations successives.

L'approche a posteriori adoptée repose sur une représentation gaussienne paramétrique, projetée à chaque instant dans l'espace des distributions normales via la minimisation de la divergence de Kullback-Leibler [25]. Cette projection donne lieu à différentes implémentations du filtre, du modèle complet à covariance matricielle jusqu'aux versions simplifiées vectorielles et scalaires, voire fixes. En parallèle, nous avons montré comment le filtre de Kalman s'apparente à la descente de gradient stochastique (SGD), avec une interprétation bayésienne du pas d'apprentissage.

Les formulations introduites ici fournissent un cadre rigoureux et modulaire pour l'estimation en ligne dans des environnements bruités. Le chapitre suivant s'attarde sur l'estimation des hyperparamètres du modèle, qui constituent des leviers essentiels pour optimiser la qualité des prédictions et assurer la robustesse du filtrage dans des environnements dynamiques.

Chapitre 3

Estimation des hyper-paramètres du modèle

À la suite de la présentation théorique du filtre de Kalman et de ses fondements probabilistes, ce chapitre s'attache à l'étude empirique de ses performances dans un contexte applicatif d'identification de réponse impulsionnelle acoustique. En particulier, l'accent est mis sur l'estimation des hyper-paramètres du modèle, notamment la variance du bruit de processus ε_t et celle du bruit de mesure v_η qui conditionnent fortement la qualité des prédictions. Plusieurs stratégies d'estimation sont ainsi comparées afin d'évaluer leur impact sur la convergence, la précision des estimations et la robustesse face aux variations dynamiques. Ces investigations visent à déterminer les configurations optimales pour un suivi adaptatif précis, tout en mettant en évidence les limites des approches classiques à paramètres fixes.

Afin de guider cette analyse, il est essentiel de comprendre le rôle structurel de ces hyper-paramètres au sein des équations du filtre de Kalman. Leur estimation conditionne directement la qualité du processus de prédiction et de mise à jour, qui constitue le cœur du filtrage bayésien séquentiel.

Dans le cadre des filtres de Kalman, l'estimation des hyper-paramètres, à savoir la variance du bruit de processus ε_t et celle du bruit de mesure v_η , joue un rôle crucial dans la performance du modèle. Une estimation incorrecte de ces paramètres peut nuire à la convergence et à la précision du modèle. Ces hyper-paramètres régulent les deux étapes fondamentales des modèles de Kalman : *(i)* la prédiction de l'état du système basée sur son évolution dynamique (2.1) et *(ii)* la mise à jour

de cet état en fonction des observations bruitées (2.2).

L'équation (2.1) modélise la dynamique du système, en supposant que l'état actuel \mathbf{w}_t dépend de l'état précédent \mathbf{w}_{t-1} avec une perturbation aléatoire contrôlée par ε_t . En parallèle, l'équation (2.2) relie les observations y_t aux entrées \mathbf{x}_t via une combinaison linéaire avec l'état \mathbf{w}_t , et est perturbée par le bruit d'observation η_t .

3.1 Principe général de l'estimation par Maximum Likelihood

Cette section expose le principe fondamental de l'estimation par maximum de vraisemblance (MLE) dans le cadre spécifique du filtre de Kalman adaptatif. L'objectif central consiste à déterminer les valeurs optimales des paramètres de variance de transition ε et du bruit d'observation v_η en maximisant la vraisemblance associée aux observations $y_{1:T}$, compte tenu du modèle dynamique considéré [14].

3.1.1 Formulation du problème

Le problème d'estimation consiste à identifier les valeurs de ε et v_η qui maximisent la vraisemblance des données observées. Formellement, cela revient à minimiser la log-vraisemblance négative définie comme suit :

$$\hat{\varepsilon}, \hat{v}_\eta = \arg \max_{\varepsilon, v_\eta} p(y_{1:T} | \varepsilon, v_\eta) \quad (3.1)$$

$$= \arg \min_{\varepsilon, v_\eta} \sum_{t=1}^T \ell_t. \quad (3.2)$$

$$\ell_t = -\log p(y_t | y_{1:t-1}, \varepsilon, v_\eta), \quad (3.3)$$

où $p(y_t | y_{1:t-1}, \varepsilon, v_\eta)$ désigne la densité prédictive de l'observation y_t , conditionnée aux observations précédentes et aux paramètres ε et v_η .

3.1.2 Calcul récursif de la densité prédictive

Afin d'obtenir une formulation opérationnelle, il est nécessaire d'établir explicitement la densité prédictive. Celle-ci se calcule récursivement selon :

$$p(y_t|y_{1:t-1}) = \int p(y_t|\mathbf{w}_t)p(\mathbf{w}_t|y_{1:t-1})d\mathbf{w}_t, \quad (3.4)$$

$$= \int \mathcal{N}(y_t; \mathbf{x}_t^\top \mathbf{w}_t, v_\eta) \mathcal{N}(\mathbf{w}_t; \boldsymbol{\mu}_{t-1}, (v_{t-1} + \varepsilon_t)\mathbf{I})d\mathbf{w}_t, \quad (3.5)$$

$$= \int \mathcal{N}(y_t; z_t, v_\eta) \mathcal{N}(z_t; \mathbf{x}_t^\top \boldsymbol{\mu}_{t-1}, \omega_t)dz_t, \quad (3.6)$$

$$= \mathcal{N}(y_t; \mathbf{x}_t^\top \boldsymbol{\mu}_{t-1}, \omega_t + v_\eta), \quad (3.7)$$

où $\omega_t = \|\mathbf{x}_t\|^2(v_{t-1} + \varepsilon_t)$.

Cette expression permet ainsi d'expliciter la log-vraisemblance négative instantanée :

$$\ell_t = -\log p(y_t|y_{1:t-1}) = \frac{e_t^2}{2(\omega_t + v_\eta)} + \frac{1}{2} \log(\omega_t + v_\eta) + \text{Const}, \quad (3.8)$$

où l'erreur de prédiction e_t est définie par $e_t = y_t - \mathbf{x}_t^\top \boldsymbol{\mu}_{t-1}$ et $\text{Const} = \frac{1}{2} \log(2\pi)$ désigne un terme constant indépendant de t , issu de la normalisation de la densité gaussienne..

3.1.3 Motivation pour l'adoption du modèle dynamique

Dans de nombreuses applications pratiques, il est irréaliste de supposer que les paramètres du modèle restent constants au cours du temps. Des phénomènes tels que les changements structurels, l'évolution du système ou la présence de dynamiques non observées exigent une approche plus flexible. Ainsi, bien que le modèle statique présenté précédemment parte du principe que les coefficients demeurent fixes durant toute la période d'observation, cette hypothèse peut s'avérer trop restrictive face à des données réelles, dont les relations statistiques évoluent dans le temps.

Cependant, cette hypothèse empêche le modèle de s'ajuster à de nouveaux schémas et limite sa capacité à saisir les vraies variations du système. En se basant sur ces constats, il est naturel de généraliser le problème en considérant un cadre plus large où les paramètres du filtre évoluent dynamiquement, comme cela sera exposé dans ce qui suit.

3.1.4 Dynamique stochastique des paramètres du filtre

Dans ce modèle, les coefficients du filtre (\mathbf{h}_t) sont autorisés à évoluer au cours du temps selon un processus de marche aléatoire. Plus précisément, à chaque instant t , le vecteur des poids $\mathbf{h}_t \in \mathbb{R}^M$ est mis à jour selon la relation suivante :

$$\mathbf{h}_t = \beta \mathbf{h}_{t-1} + \mathbf{u}_t \sqrt{\varepsilon^*}, \quad (3.9)$$

où \mathbf{u}_t est un vecteur de variables aléatoires gaussiennes indépendantes, de moyenne nulle et de variance unitaire, ε^* représente la variance du bruit de processus qui contrôle l'ampleur des fluctuations aléatoires des coefficients entre deux instants successifs. Et, $0 < \beta \leq 1$ est un paramètre d'auto-régression qui module la mémoire du système:

- lorsque $\beta = 1$, le modèle correspond à une marche aléatoire classique, sans perte d'information, comme souvent supposé dans la littérature [15];
- en revanche, pour $\beta < 1$, le modèle introduit une décroissance progressive des coefficients, reflétant une éventuelle perte ou évolution des paramètres au fil du temps [29].

Cette formulation permet de modéliser les variations des paramètres du système dues à des dynamiques ou à des incertitudes de modélisation.

L'observation associée à chaque instant t , notée $y_t \in \mathbb{R}$, est alors donnée par :

$$y_t = \mathbf{x}_t^\top \mathbf{h}_t + \eta_t, \quad (3.10)$$

où $\mathbf{x}_t \in \mathbb{R}^M$ désigne le vecteur d'entrée à l'instant t , et η_t correspond à un bruit d'observation gaussien, de moyenne nulle et de variance v_η^* . Cette variance est déterminée en fonction du rapport signal-sur-bruit (SNR).

Cette équation traduit la mesure bruitée de la sortie du système à partir de l'état courant des coefficients du filtre.

Ce cadre évolutif permet au filtre de suivre et d'ajuster continuellement les paramètres au fur et à mesure de l'arrivée de nouvelles données, ce qui s'avère particulièrement pertinent dans un environnement dynamique.

Intérêt du modèle dynamique

Le recours au modèle dynamique présente plusieurs avantages majeurs :

- **Adaptation aux changements** : Il permet au filtre de s'ajuster en temps réel à des données dont les propriétés évoluent, ce qui est fondamental pour garantir des prédictions fiables dans un environnement non stationnaire.
- **Généralisation** : Le modèle dynamique englobe le modèle statique comme cas particulier ($\varepsilon^* = 0$ et $\beta = 1$), tout en offrant une plus grande flexibilité de modélisation.

Ce choix est particulièrement pertinent dès lors que l'on cherche à estimer de manière optimale la variance du bruit de processus $\hat{\varepsilon}$ ainsi que celle du bruit d'observation \hat{v}_η par maximisation de la vraisemblance des données observées.

3.1.5 Analyse préliminaire et sensibilité aux paramètres

Avant d'aborder l'optimisation des paramètres, il est pertinent d'examiner la sensibilité du modèle à différents choix de ε et v_η . Une première étape consiste à analyser la moyenne glissante de la log-vraisemblance, définie par :

$$\bar{\ell}_t = \frac{1}{L} \sum_{l=0}^{L-1} \ell_{t-l}. \quad (3.11)$$

Cette exploration préliminaire permet d'appréhender qualitativement l'impact de ε et v_η sur la performance prédictive du modèle, et de mieux comprendre leur rôle avant l'introduction de stratégies adaptatives.

Afin de compléter cette étude, nous menons une série de simulations visant à évaluer l'influence de ces paramètres sur l'erreur quadratique moyen \bar{m}_t et sur la log-vraisemblance moyenne $\bar{\ell}_t$. Notons que le cas du modèle statique ($\varepsilon^*=0$) s'inscrit naturellement comme une situation particulière de ce cadre dynamique, et fera l'objet d'une comparaison dédiée en fin de section.

3.1.5.1 Performance du filtre selon différents scénarios de paramétrisation

Pour illustrer et approfondir ces analyses, nous présentons ci-après des visualisations graphiques des principales métriques de performance du filtre. Ces représentations permettent de saisir la dynamique d'apprentissage et la capacité d'adaptation de ce dernier face aux variations des paramètres du système.

Les figures suivantes synthétisent ces résultats et apportent un éclairage visuel sur le comportement du filtre selon les différents scénarios étudiés:

1. Effet de la variation de v_η (bruit d'observation)

Ces deux figures examinent l'influence de différents niveaux de bruit d'observation sur la performance du filtre :

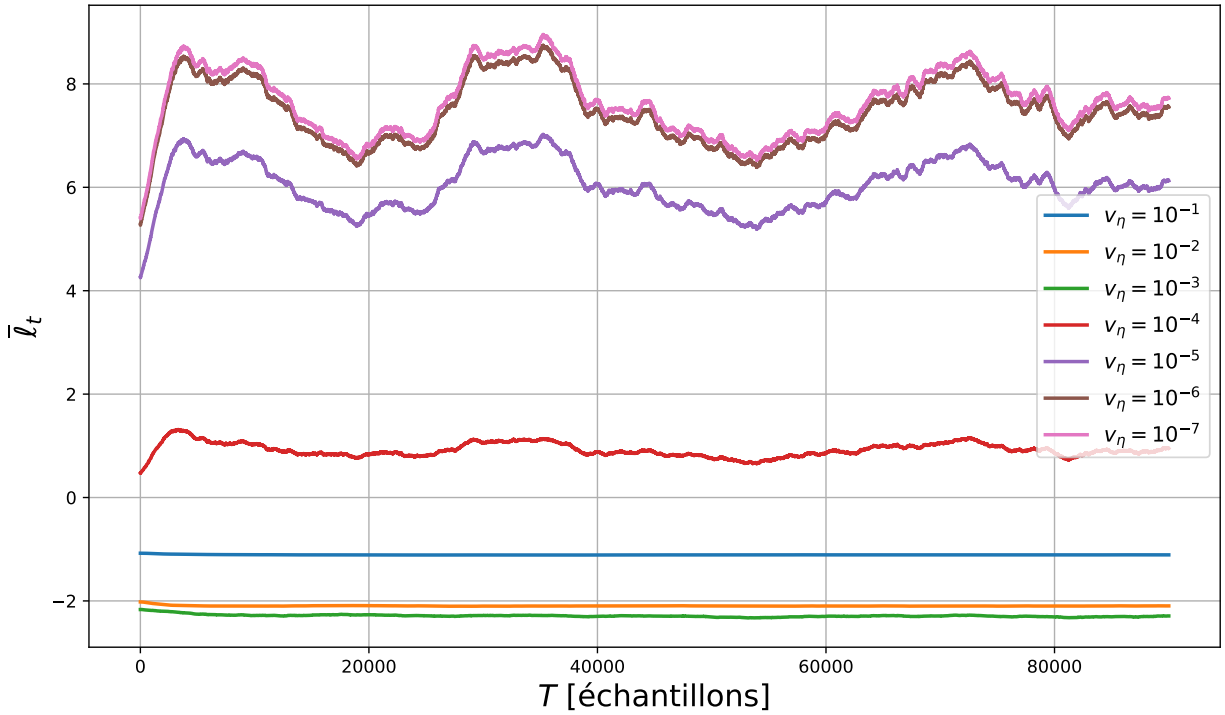


Figure 3.1: Courbe d'évolution de la moyenne glissante de la log-vraisemblance $\bar{\ell}_t$ obtenue pour différentes valeurs du bruit d'observation v_η dans le cadre du SKF, avec des paramètres initiaux du filtre fixés à $\varepsilon_0 = 10^{-10}$ et $v_0 = 8 \cdot 10^{-6}$. L'évaluation temporelle de h_t est réalisée en considérant $\varepsilon^* = 10^{-10}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$.

La figure 3.1 montre que la perte atteint son minimum lorsque v_η est fixé à 10^{-3} , valeur proche de $v_\eta^* = 2.1 \times 10^{-3}$. Un écart important entre la valeur initiale et la valeur réelle dégrade les performances, confirmant l'intérêt de choisir v_η initial au plus près de v_η^* .

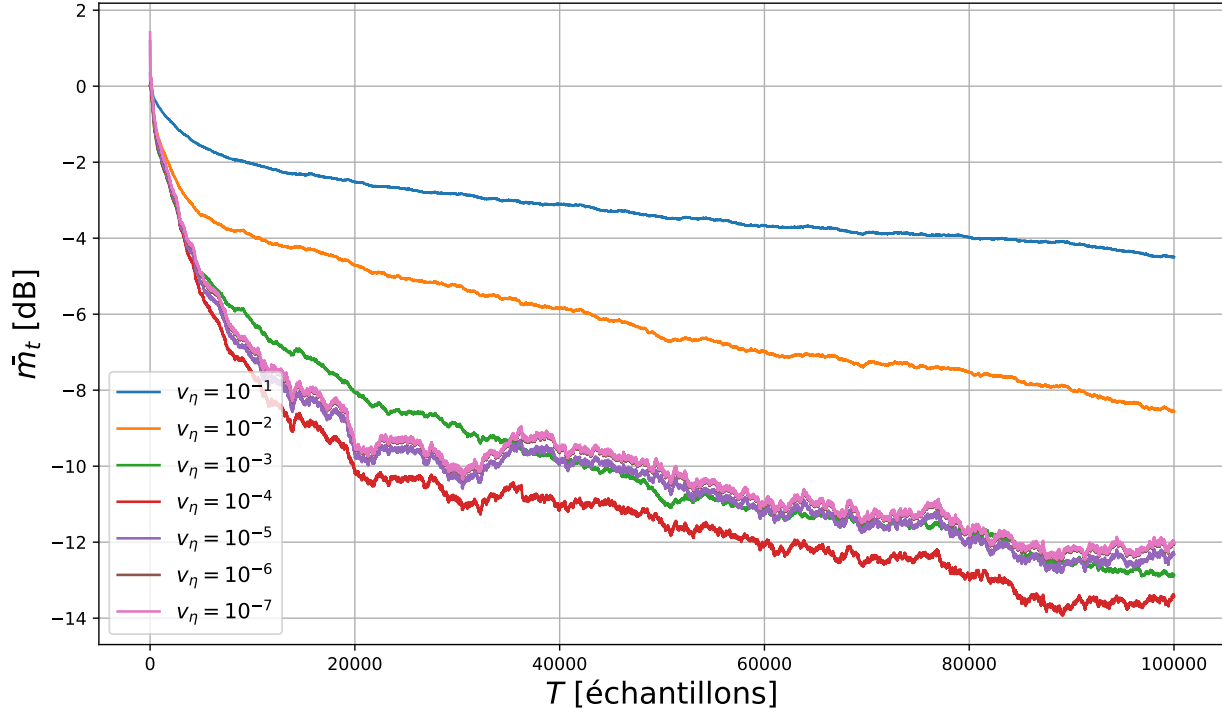


Figure 3.2: Évolution temporelle de l'erreur quadratique moyenne \bar{m}_t pour différentes valeurs du bruit d'observation v_η dans le cadre du SKF, avec des paramètres initiaux du filtre fixés à $\varepsilon_0 = 10^{-10}$ et $v_0 = 8 \cdot 10^{-6}$. L'évaluation temporelle de h_t est réalisée en considérant $\varepsilon^* = 10^{-10}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$.

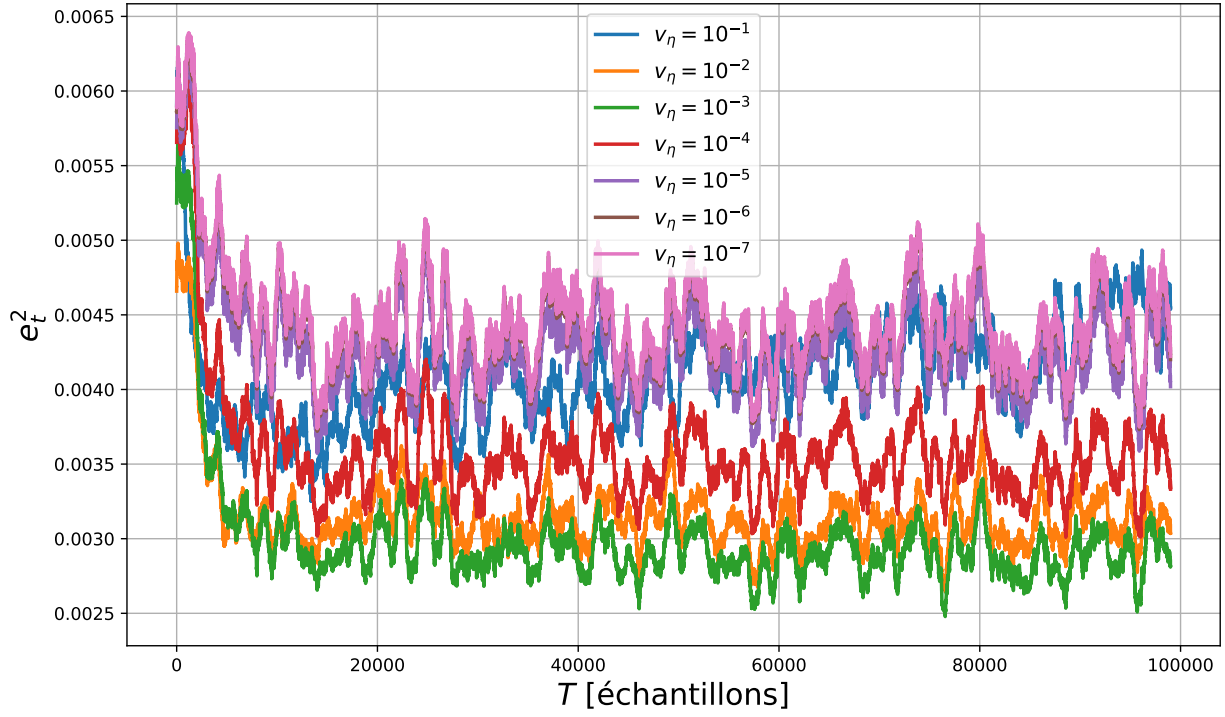


Figure 3.3: Évolution de l'erreur de prédiction au carré e_t^2 au cours du temps pour différents v_η dans le cadre du SKF, avec des paramètres initiaux du filtre fixés à $\varepsilon_0 = 10^{-10}$ et $v_0 = 8 \cdot 10^{-6}$. L'évaluation temporelle de h_t est réalisée en considérant $\varepsilon^* = 10^{-10}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$.

2. Effet de la variation de ε (variance de transition) :

Les deux figures suivantes mettent en évidence l'effet de la variance de transition ε sur la dynamique d'apprentissage du filtre.

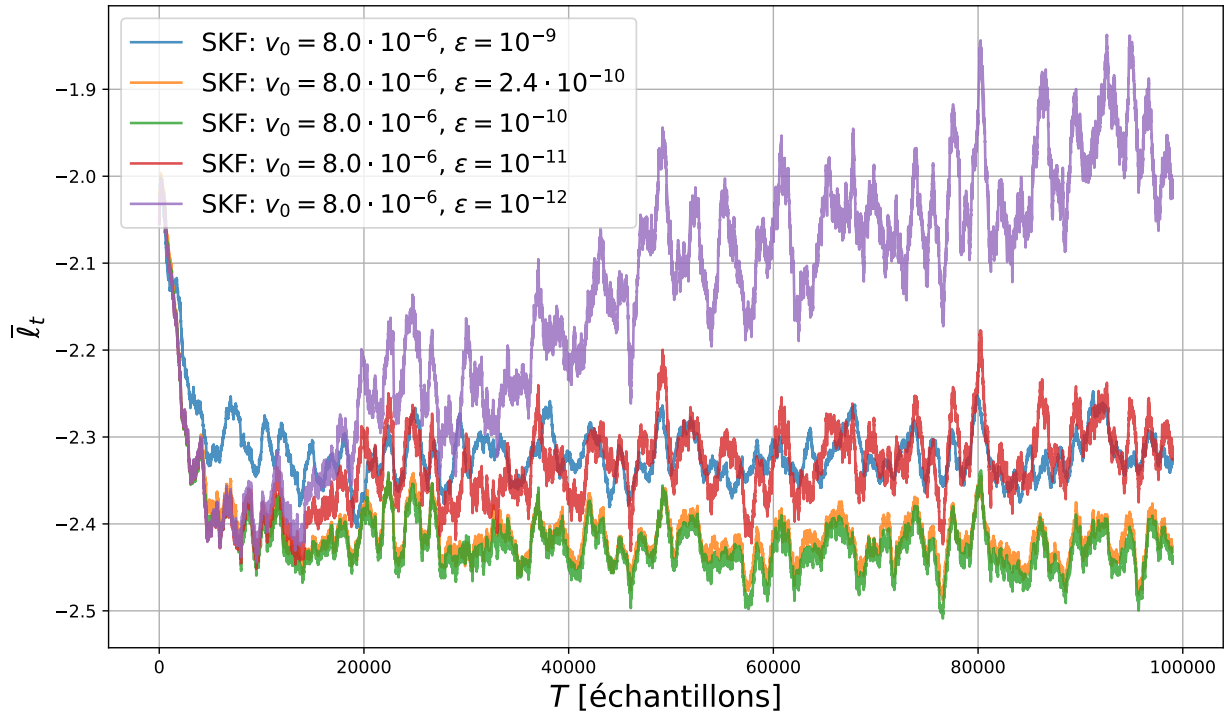


Figure 3.4: Évolution de la moyenne glissante de la log-vraisemblance $\bar{\ell}_t$ pour différentes valeurs de la variance de transition ε , lors de l'évaluation temporelle de h_t avec $\varepsilon^* = 10^{-10}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$, dans le cadre de l'application du filtre SKF.

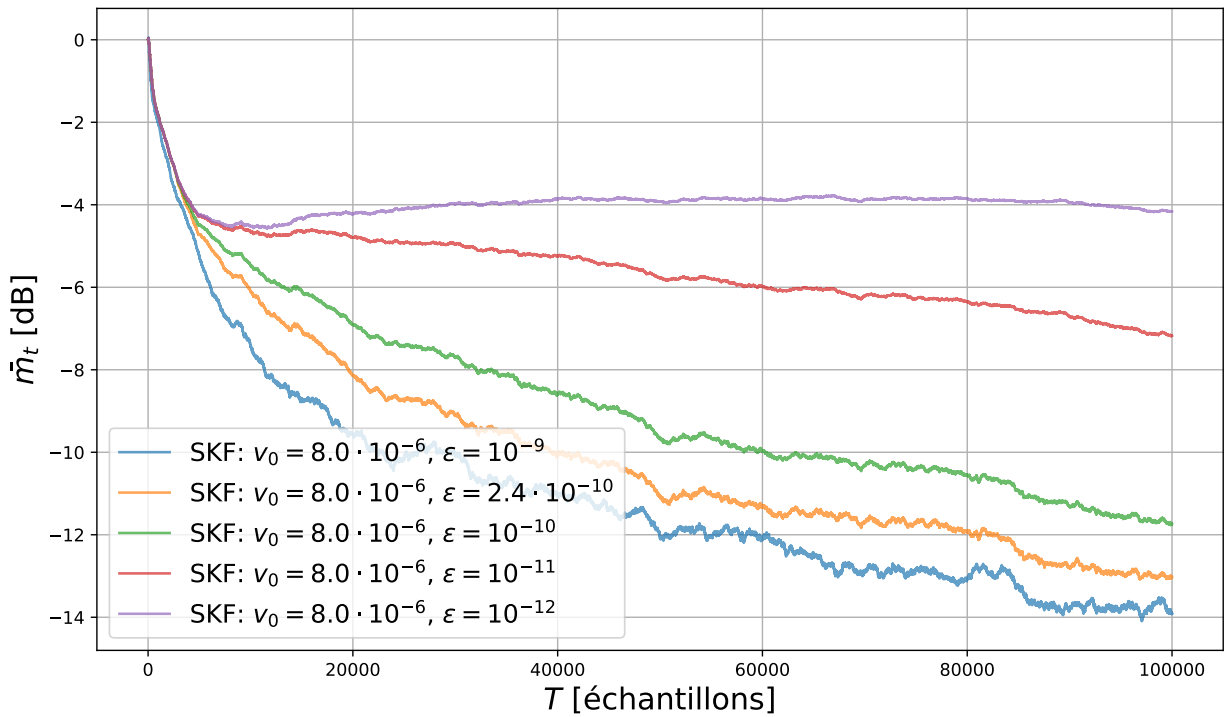


Figure 3.5: Évolution de l'erreur quadratique moyenne \bar{m}_t pour différentes valeurs de la variance de transition ε , illustrant l'impact de ce paramètre sur la performance du filtre SKF, lors de l'évaluation temporelle de h_t avec $\varepsilon^* = 10^{-10}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$.

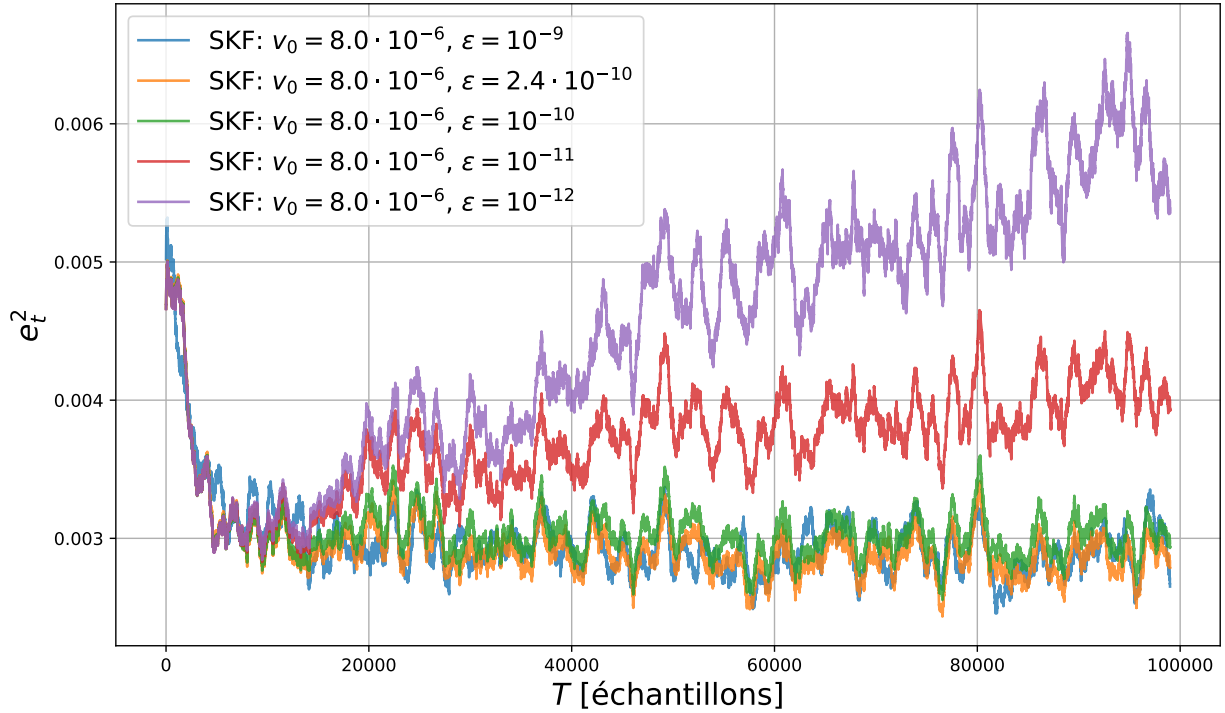


Figure 3.6: Évolution de l'erreur de prédiction au carré e_t^2 au cours du temps pour différentes valeurs de ε , lors de l'évaluation temporelle de h_t avec $\varepsilon^* = 10^{-10}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$, dans le cadre de l'application du filtre SKF.

L'analyse des résultats montre que lorsque la valeur estimée de ε est proche de la valeur réelle $\varepsilon^* = 10^{-10}$ (courbe verte), la log-vraisemblance moyenne $\bar{\ell}_t$ atteint un minimum plus prononcé que pour les autres valeurs testées. Ce résultat souligne la sensibilité de cette métrique à un réglage précis et réaliste de ε . En revanche, les deux autres métriques considérées, à savoir l'erreur quadratique moyenne m_t et l'erreur instantanée e_t^2 , ne révèlent pas de différences significatives permettant de discriminer clairement les valeurs de ε dans ce contexte.

3.1.5.2 Performance du filtre pour $\varepsilon^* = 1 \cdot 10^{-12}$ et $\beta = 1$

Afin de compléter l'étude, nous considérons un nouveau scénario où l'ensemble des simulations est exécuté en fixant la variance de transition à $\varepsilon^* = 1 \cdot 10^{-12}$ (au lieu de $\varepsilon^* = 1 \cdot 10^{-10}$ dans les analyses précédentes). Ce choix permet d'évaluer la robustesse et la sensibilité du filtre face à une variance de transition très faible. Les indicateurs de performance sont présentés et comparés à ceux obtenus précédemment en analysant les deux scénarios suivants :

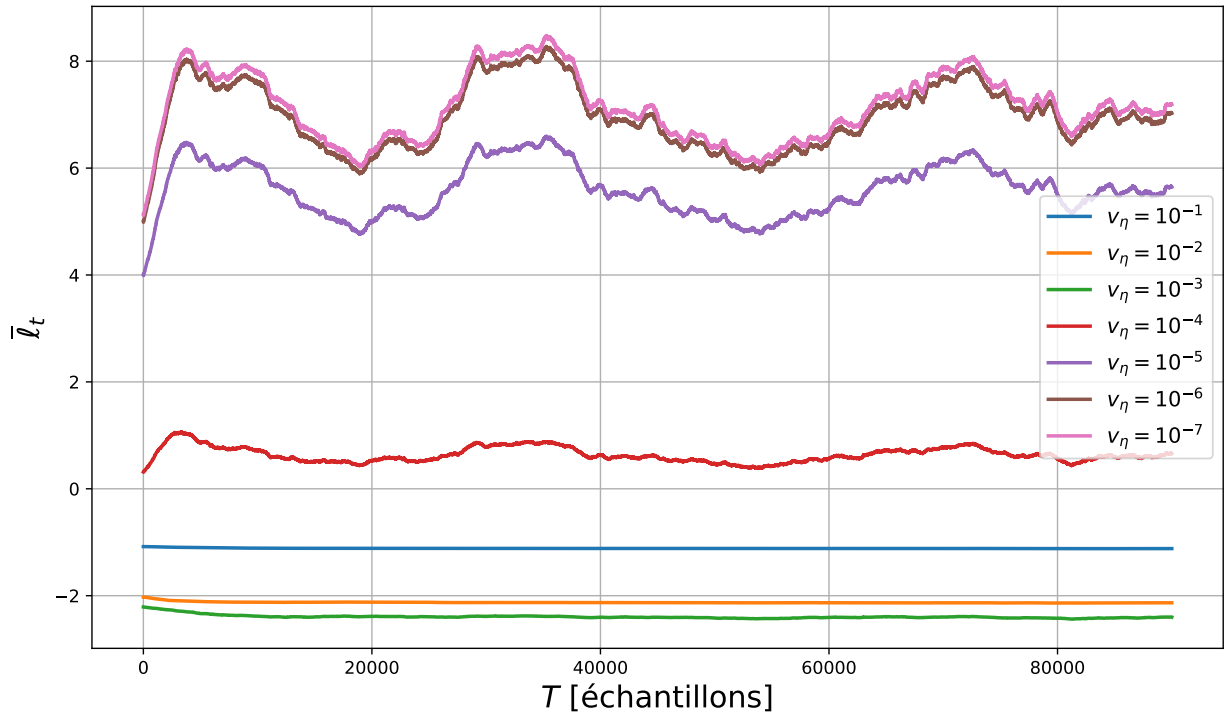
1. Effet de la variation de v_η (bruit d'observation)

Figure 3.7: Évolution de la moyenne glissante de la log-vraisemblance $\bar{\ell}_t$ pour différentes valeurs du bruit d'observation v_η dans le cadre du SKF, avec des paramètres initiaux du filtre fixés à $\varepsilon_0 = 10^{-12}$ et $v_0 = 8 \cdot 10^{-6}$. L'évaluation temporelle de h_t est réalisée en considérant $\varepsilon^* = 10^{-12}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$.

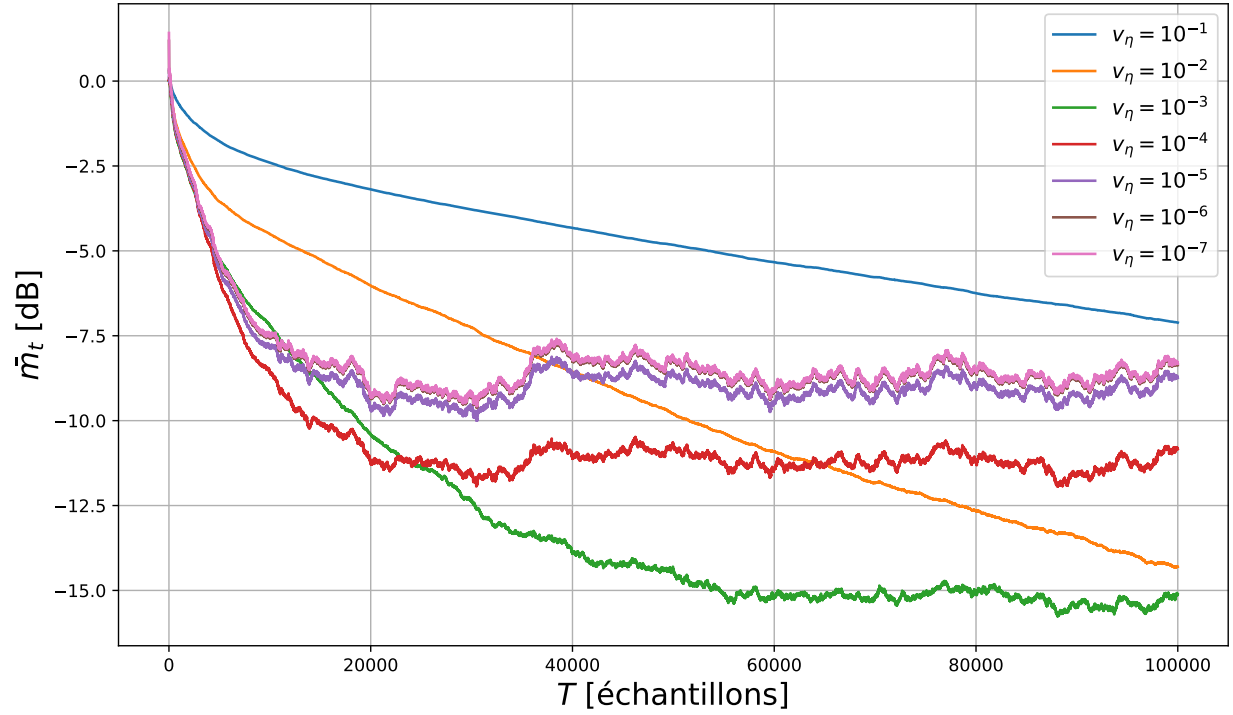


Figure 3.8: Évolution temporelle de l'erreur quadratique moyenne \bar{m}_t pour différentes valeurs du bruit d'observation v_η dans le cadre du SKF, avec des paramètres initiaux du filtre fixés à $\varepsilon_0 = 10^{-12}$ et $v_0 = 8 \cdot 10^{-6}$. L'évaluation temporelle de h_t est réalisée en considérant $\varepsilon^* = 10^{-12}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$.

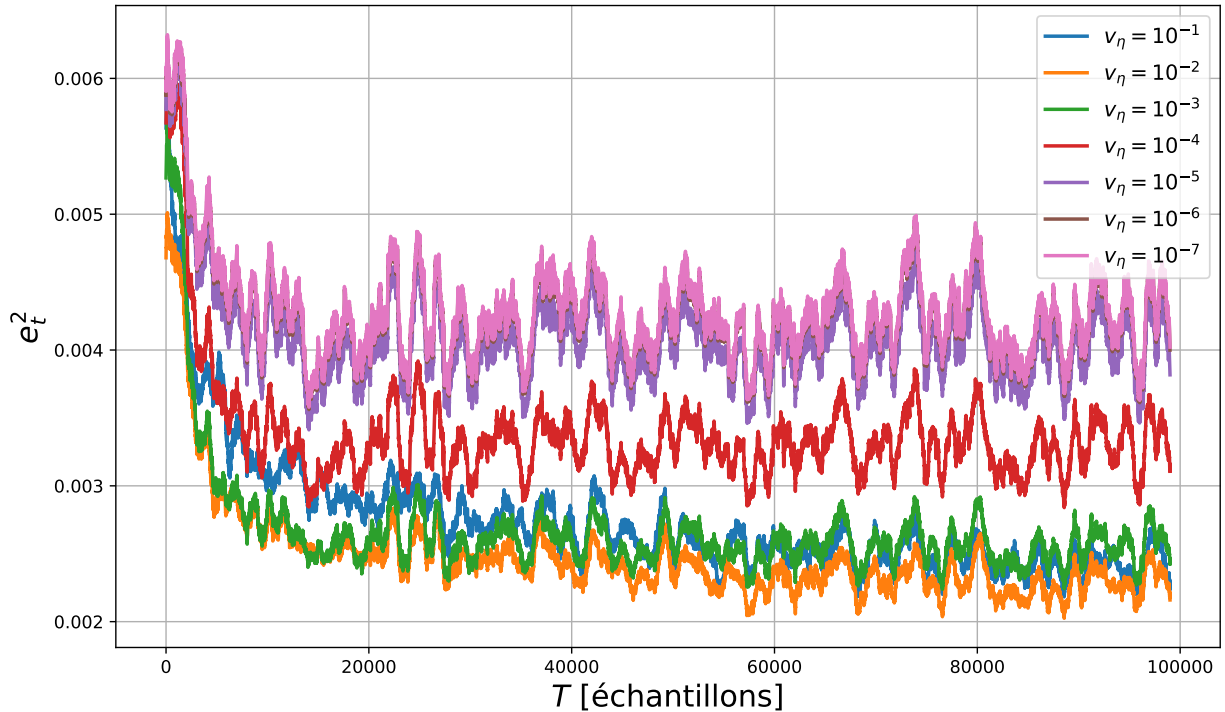


Figure 3.9: Évolution de l'erreur de prédiction au carré e_t^2 au cours du temps pour différents v_η dans le cadre du SKF, avec des paramètres initiaux du filtre fixés à $\varepsilon_0 = 10^{-12}$ et $v_0 = 8 \cdot 10^{-6}$. L'évaluation temporelle de h_t est réalisée en considérant $\varepsilon^* = 10^{-12}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$.

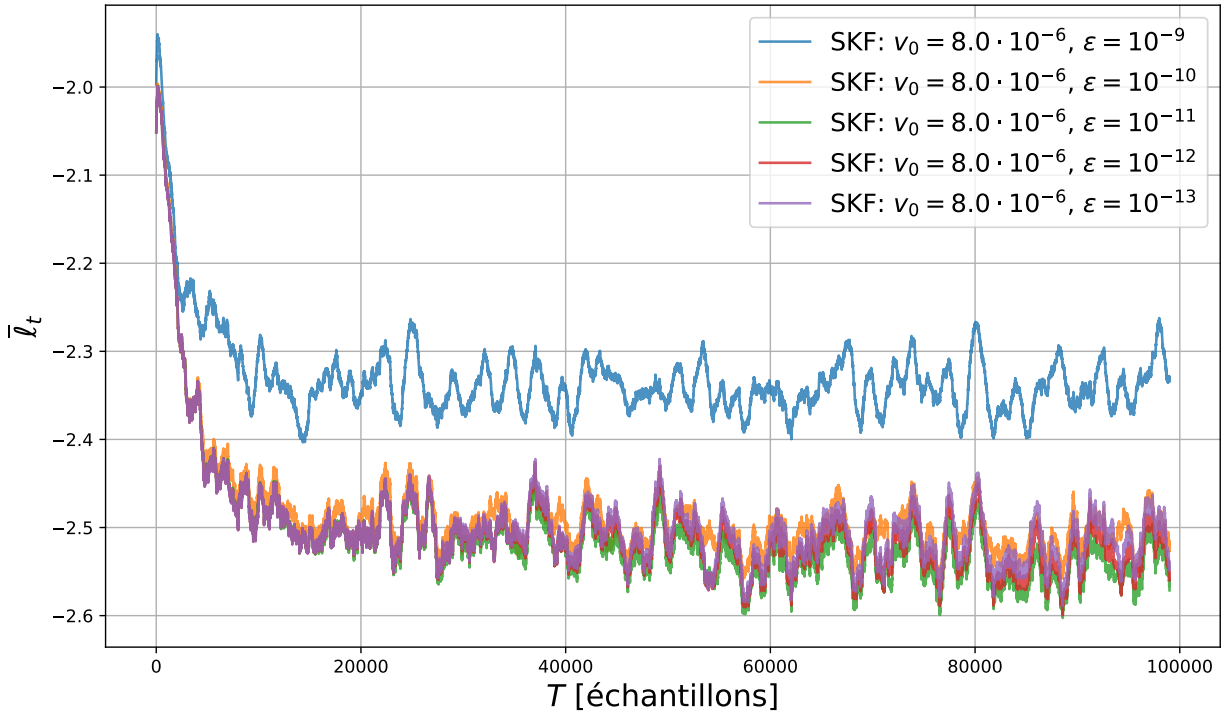
2. Effet de la variation de ε (variance de transition)

Figure 3.10: Évolution de la moyenne glissante de la log-vraisemblance $\bar{\ell}_t$ pour différentes valeurs de la variance de transition ε , lors de l'évaluation temporelle de \mathbf{h}_t avec $\varepsilon^* = 1 \cdot 10^{-12}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$, dans le cadre de l'application du filtre SKF.

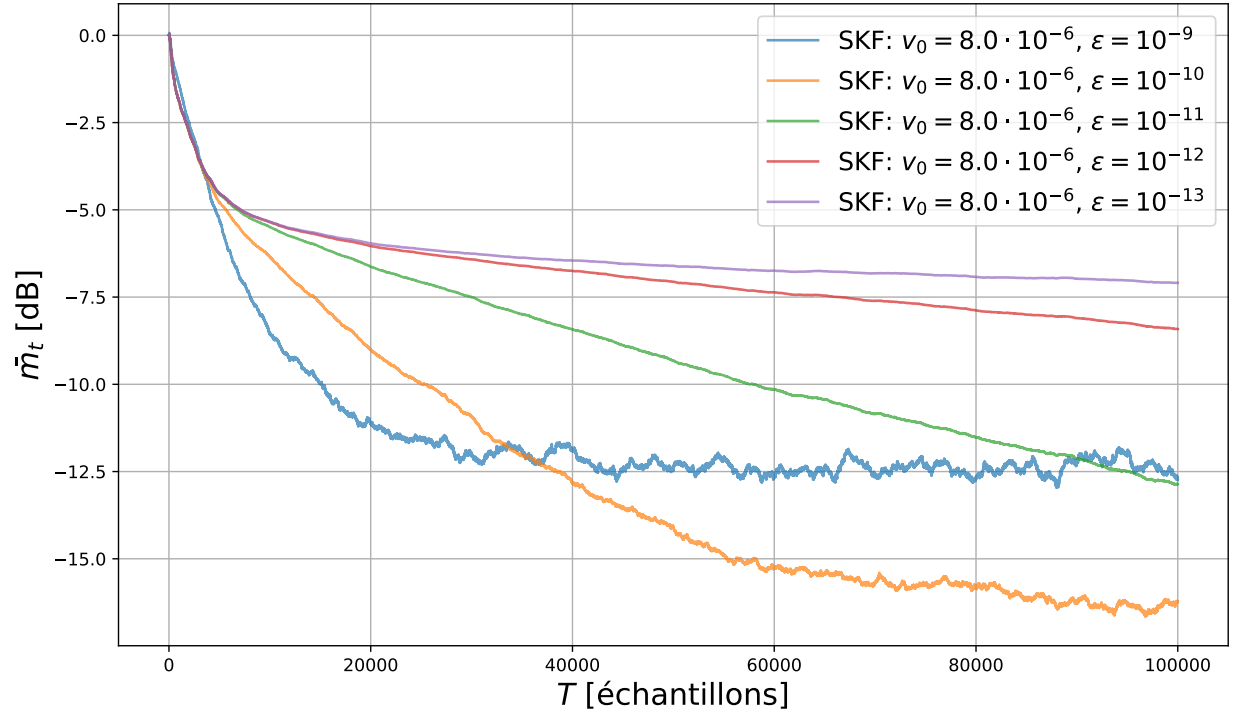


Figure 3.11: Effet de la variance de transition ε sur \tilde{m}_t lors de l'estimation temporelle de \mathbf{h}_t ($\varepsilon^* = 10^{-12}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$), dans le contexte d'utilisation du filtre SKF.

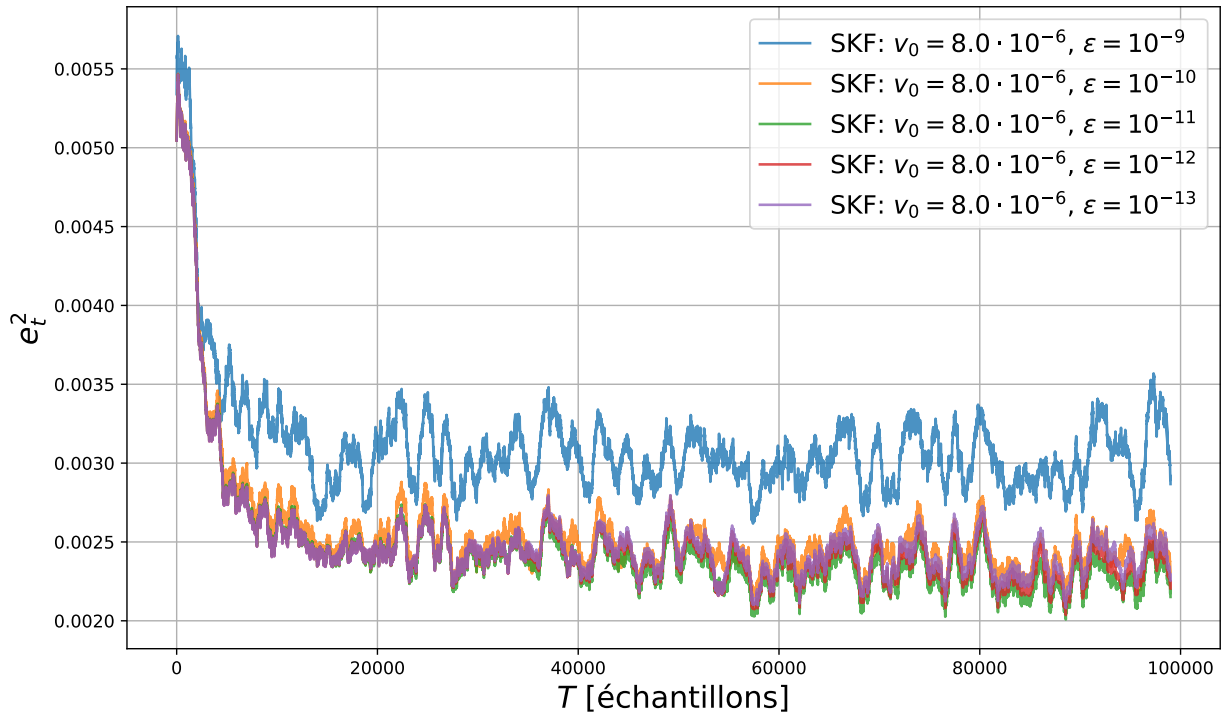


Figure 3.12: Évolution de l'erreur de prédiction au carré e_t^2 au cours du temps pour différents ε , lors de l'évaluation temporelle de \mathbf{h}_t avec $\varepsilon^* = 10^{-12}$, $v_\eta^* = 2.1 \cdot 10^{-3}$ et $\beta = 1$, dans le cadre de l'application du filtre SKF.

Ces résultats sont comparés au cas statique, présenté dans la section suivante, où $\varepsilon^* = 0$, correspondant à une réponse impulsionnelle constante.

3.1.5.3 Performance du filtre avec le modèle statique ($\varepsilon^* = 0$ et $\beta = 1$)

Dans cette section, nous examinons le comportement du filtre dans un contexte statique, où la variance de transition est fixée à zéro ($\varepsilon^* = 0$). En reprenant la même méthodologie de simulation et de comparaison présente précédemment.

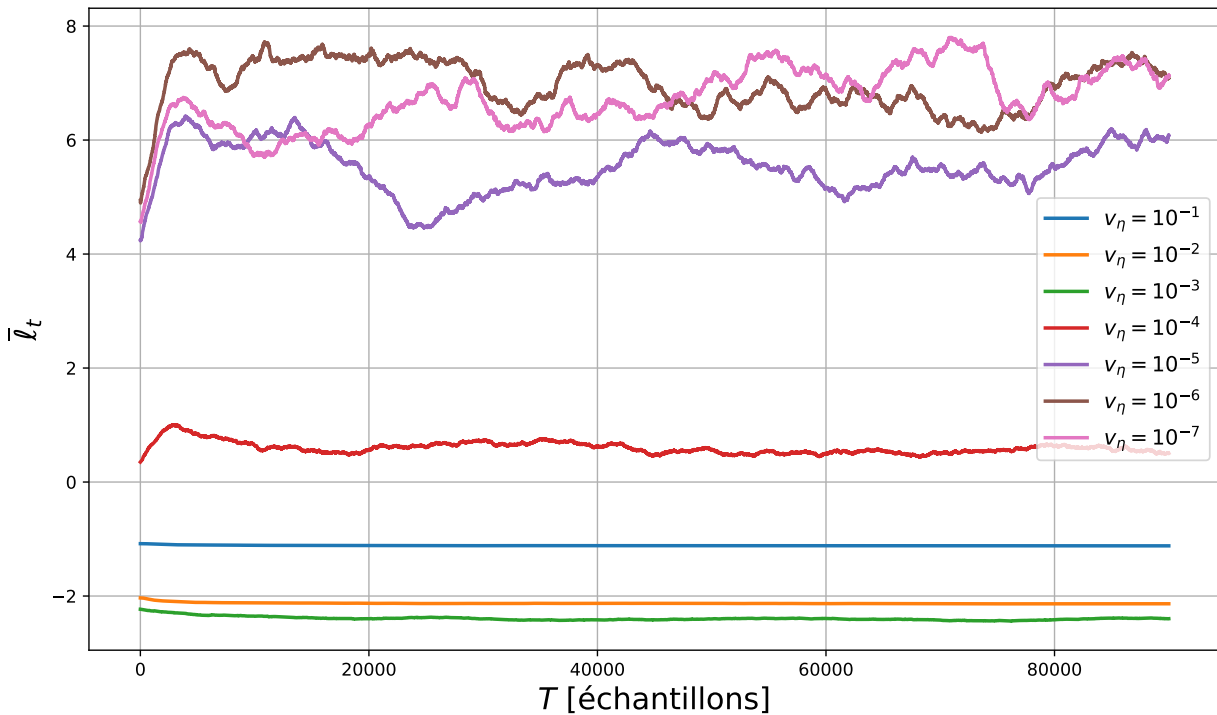


Figure 3.13: Évolution de la moyenne glissante de la log-vraisemblance $\bar{\ell}_t$ pour différents v_η dans le cadre du SKF, avec des paramètres initiaux du filtre fixés à $\varepsilon_0 = 10^{-10}$ et $v_0 = 8 \cdot 10^{-6}$, avec des données issues d'un modèle statique ($\varepsilon^* = 0$ et $\beta = 1$).

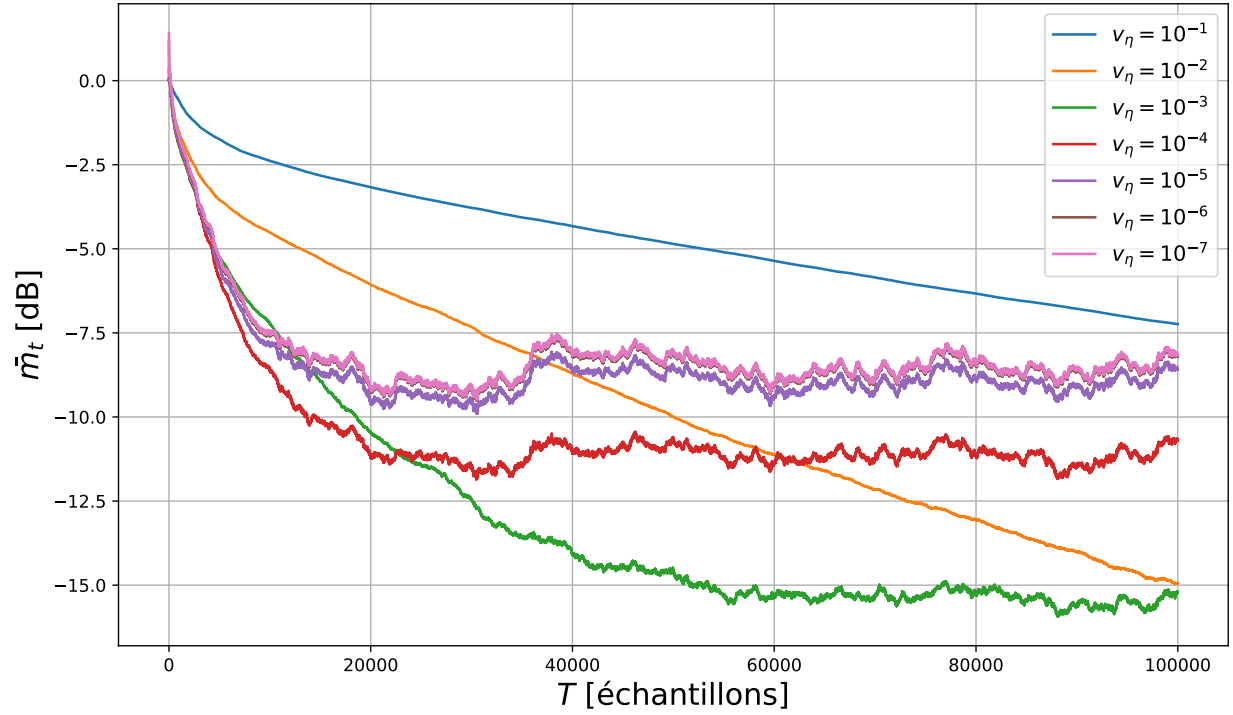


Figure 3.14: Évolution temporelle de l'erreur quadratique moyenne \bar{m}_t pour différentes valeurs du bruit d'observation v_η dans le cadre du SKF, avec des paramètres initiaux du filtre fixés à $\varepsilon_0 = 10^{-10}$ et $v_0 = 8 \cdot 10^{-6}$, avec des données issues d'un modèle statique ($\varepsilon^* = 0$ et $\beta = 1$).

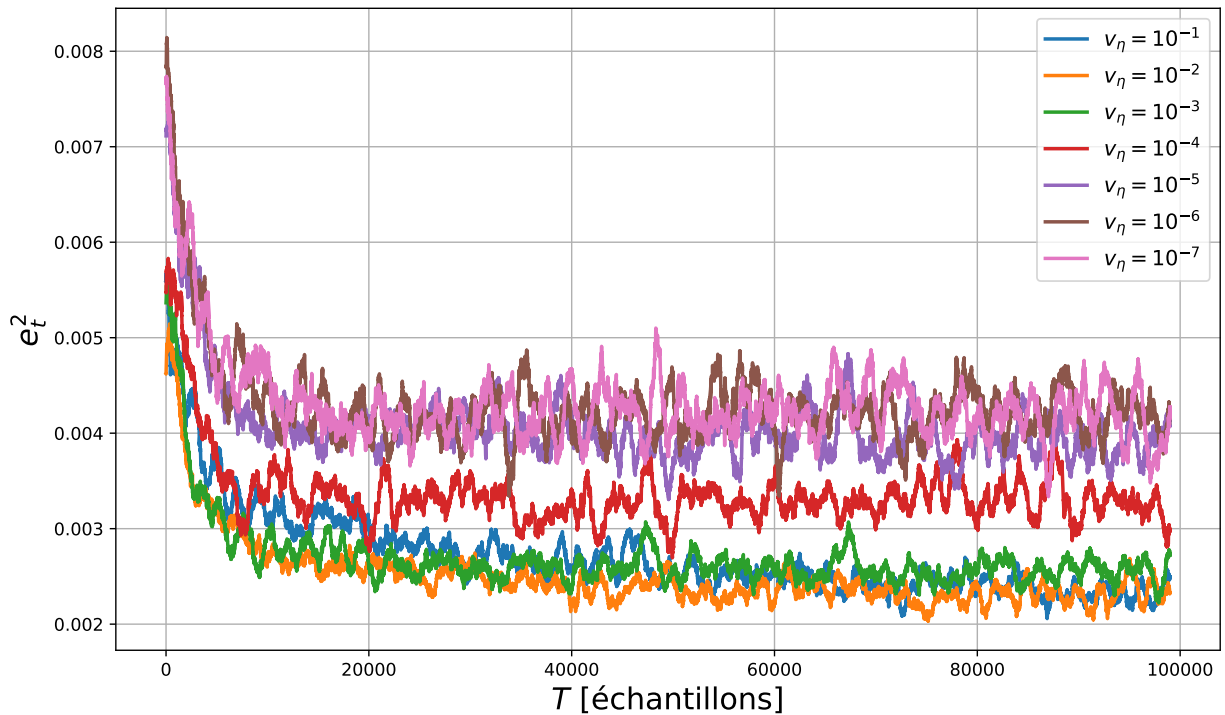


Figure 3.15: Évolution de l'erreur de prédiction au carré e_t^2 au cours du temps pour différents v_η dans le cadre du SKF, avec des paramètres initiaux du filtre fixés à $\varepsilon_0 = 10^{-10}$ et $v_0 = 8 \cdot 10^{-6}$, avec des données issues d'un modèle statique ($\varepsilon^* = 0$ et $\beta = 1$).

Cette comparaison montre que les paramètres optimaux selon cette approche ne correspondent pas nécessairement à la solution la plus rapide en termes de convergence de l'erreur quadratique moyenne. Autrement dit, si cette analyse permet une adaptation optimale au sens de la vraisemblance, elle n'assure pas toujours la stabilité ou la rapidité de convergence du filtre.

Les résultats numériques illustrent clairement cette distinction: une optimisation fondée uniquement sur la vraisemblance n'est pas toujours suffisante pour garantir la performance du filtre. Cela souligne la nécessité d'évaluer le filtre selon plusieurs critères complémentaires, au-delà du seul cadre statistique.

L'ensemble des figures présentées dans cette section montre que la perte est minimale lorsque v_η est fixé à 10^{-3} , valeur proche de la référence $v_\eta^* = 2.1 \times 10^{-3}$. Dans tous les cas étudiés, un écart important entre la valeur initiale et la valeur réelle de v_η entraîne une diminution notable de la perte et de l'erreur quadratique moyenne, confirmant l'intérêt de choisir une valeur initiale au plus près de v_η^* . On constate également que l'erreur quadratique moyenne reste systématiquement plus faible pour la configuration où v_η est proche de v_η^* , illustrée par la courbe verte dans l'ensemble des figures.

3.1.6 Optimisation de la variance de transition par descente de gradient

Dans cette section, nous présentons une méthode d'optimisation de la variance de transition ε par descente de gradient, appliquée à la minimisation de la log-vraisemblance négative. Cette approche est fondée sur le calcul explicite du gradient de la fonction de coût par rapport à ε . La démarche se structure comme suit :

1. Expression du gradient de la log-vraisemblance négative.

La fonction de coût (log-vraisemblance négative) au temps t est notée ℓ_t . Le gradient de ℓ_t par rapport à ε s'écrit :

$$\frac{\partial \ell_t}{\partial \varepsilon} = -\frac{e_t}{\omega_t + v_\eta} \mathbf{x}_t^\top \frac{\partial \boldsymbol{\mu}_{t-1}}{\partial \varepsilon} - \frac{e_t^2}{2(\omega_t + v_\eta)^2} \frac{\partial \omega_t}{\partial \varepsilon} + \frac{1}{2(\omega_t + v_\eta)} \frac{\partial \omega_t}{\partial \varepsilon}, \quad (3.12)$$

$$= \frac{1}{\omega_t + v_\eta} \left[e_t \mathbf{x}_t^\top \frac{\partial \boldsymbol{\mu}_{t-1}}{\partial \varepsilon} + \frac{\|\mathbf{x}_t\|^2 (\omega_t + v_\eta - e_t^2)}{(\omega_t + v_\eta)^2} \left(\frac{\partial v_{t-1}}{\partial \varepsilon} + 1 \right) \right], \quad (3.13)$$

où $e_t = y_t - \mathbf{x}_t^\top \boldsymbol{\mu}_{t-1}$ désigne l'erreur de prédiction.

2. Calcul des dérivées récursives.

Les dérivées nécessaires à la mise à jour du gradient s'obtiennent par récurrence:

$$\frac{\partial \boldsymbol{\mu}_{t-1}}{\partial \varepsilon} = \frac{\partial \boldsymbol{\mu}_{t-2}}{\partial \varepsilon} + \mathbf{x}_{t-1} \left[\frac{\frac{\partial v_{t-1,t-2}}{\partial \varepsilon} e_{t-1} + v_{t-1,t-2} \frac{\partial e_{t-1}}{\partial \varepsilon}}{\omega_{t-1} + v_\eta} - \frac{v_{t-1,t-2} e_{t-1}}{(\omega_{t-1} + v_\eta)^2} \frac{\partial \omega_{t-1}}{\partial \varepsilon} \right], \quad (3.14)$$

$$= \frac{\partial \boldsymbol{\mu}_{t-2}}{\partial \varepsilon} + \mathbf{x}_{t-1} \left[\frac{e_{t-1} \left(\frac{\partial v_{t-2}}{\partial \varepsilon} + 1 \right) - v_{t-1,t-2} \mathbf{x}_{t-1}^\top \frac{\partial \boldsymbol{\mu}_{t-2}}{\partial \varepsilon}}{\omega_{t-1} + v_\eta} - \frac{v_{t-1,t-2} e_{t-1}}{(\omega_{t-1} + v_\eta)^2} \|\mathbf{x}_{t-1}\|^2 \left(\frac{\partial v_{t-2}}{\partial \varepsilon} + 1 \right) \right], \quad (3.15)$$

$$= \frac{\partial \boldsymbol{\mu}_{t-2}}{\partial \varepsilon} + \mathbf{x}_{t-1} \left[\frac{-v_{t-1,t-2} \mathbf{x}_{t-1}^\top \frac{\partial \boldsymbol{\mu}_{t-2}}{\partial \varepsilon}}{\omega_{t-1} + v_\eta} + \frac{e_{t-1} v_\eta}{(\omega_{t-1} + v_\eta)^2} \left(\frac{\partial v_{t-2}}{\partial \varepsilon} + 1 \right) \right], \quad (3.16)$$

et

$$\frac{\partial v_t}{\partial \varepsilon} = \left(\frac{\partial v_{t-1}}{\partial \varepsilon} + 1 \right) \left(1 - \frac{1}{M} \frac{\omega_t}{\omega_t + v_\eta} \right) - \frac{v_{t,t-1}}{M} \frac{v_\eta}{\omega_t + v_\eta} \frac{\partial \omega_t}{\partial \varepsilon}, \quad (3.17)$$

$$= \left(\frac{\partial v_{t-1}}{\partial \varepsilon} + 1 \right) \left(1 - \frac{1}{M} \frac{\omega_t}{\omega_t + v_\eta} \right) - \frac{v_{t,t-1}}{M} \frac{v_\eta}{\omega_t + v_\eta} \|\mathbf{x}_t\|^2 \left(\frac{\partial v_{t-1}}{\partial \varepsilon} + 1 \right), \quad (3.18)$$

$$= \left(\frac{\partial v_{t-1}}{\partial \varepsilon} + 1 \right) \left(1 - \frac{1}{M} \frac{\omega_t(1 + v_\eta)}{\omega_t + v_\eta} \right), \quad (3.19)$$

avec les dérivées intermédiaires :

$$\frac{\partial v_{t,t-1}}{\partial \varepsilon} = \frac{\partial v_{t-1}}{\partial \varepsilon} + 1. \quad (3.20)$$

$$\frac{\partial \omega_t}{\partial \varepsilon} = \|\mathbf{x}_t\|^2 \left(\frac{\partial v_{t-1}}{\partial \varepsilon} + 1 \right). \quad (3.21)$$

3. Algorithme d'estimation MLE

- Initialiser $\varepsilon_0 > 0$ et les paramètres du filtre.
- **Pour** chaque $t = 1, \dots, T$:
 - (a) Calculer $\boldsymbol{\mu}_t$ et v_t via le filtre de Kalman.
 - (b) Calculer le gradient :

$$\frac{\partial \ell_t}{\partial \varepsilon} \Big|_{\varepsilon=\varepsilon_{t-1}} = \frac{1}{\omega_t + v_\eta} \left(e_t \mathbf{x}_t^\top \frac{\partial \boldsymbol{\mu}_{t-1}}{\partial \varepsilon} + \frac{\|\mathbf{x}_t\|^2 (\omega_t + v_\eta - e_t^2)}{(\omega_t + v_\eta)^2} \left(\frac{\partial v_{t-1}}{\partial \varepsilon} + 1 \right) \right). \quad (3.22)$$

(c) Mettre à jour :

$$\varepsilon_t = \varepsilon_{t-1} - \mu_\varepsilon \frac{\partial \ell_t}{\partial \varepsilon} \Big|_{\varepsilon=\varepsilon_{t-1}}. \quad (3.23)$$

4. **Reparamétrisation exponentielle** Afin de garantir que $\varepsilon_t > 0$ à chaque instant, nous introduisons une reparamétrisation exponentielle. On pose :

$$\varepsilon_t = \exp(\xi_t) \quad \Rightarrow \quad \frac{\partial \varepsilon_t}{\partial \xi_t} = \exp(\xi_t) = \varepsilon_t, \quad (3.24)$$

où $\xi_t \in \mathbb{R}$, cette transformation permet d'effectuer la mise à jour dans un espace non contraint (\mathbb{R}), tout en s'assurant que ε_t reste strictement positif

Ainsi, la dérivée de la log-vraisemblance par rapport à ξ_t est donnée par la règle de chaîne :

$$\frac{\partial \ell}{\partial \xi_t} = \frac{\partial \ell}{\partial \varepsilon_t} \cdot \frac{\partial \varepsilon_t}{\partial \xi_t} = \varepsilon_t \cdot \frac{\partial \ell}{\partial \varepsilon_t}. \quad (3.25)$$

5. **Mise à jour de ξ_t .** L'estimation adaptative suit une descente de gradient classique:

$$\xi_{t+1} = \xi_t - \mu_\xi \cdot \varepsilon_t \cdot \frac{\partial \ell_t}{\partial \varepsilon} \Big|_{\varepsilon=\varepsilon_{t-1}}. \quad (3.26)$$

6. **Résumé final.**

$$\boxed{\xi_{t+1} = \xi_t - \mu_\xi \cdot \varepsilon_t \cdot \left[\frac{1}{\omega_t + v_\eta} \left(e_t \mathbf{x}_t^\top \frac{\partial \boldsymbol{\mu}_{t-1}}{\partial \varepsilon} + \frac{\|\mathbf{x}_t\|^2 (\omega_t + v_\eta - e_t^2)}{(\omega_t + v_\eta)^2} \left(\frac{\partial v_{t-1}}{\partial \varepsilon} + 1 \right) \right) \right]}. \quad (3.27)$$

où $\mu_\xi > 0$ est le pas d'adaptation dans l'espace log-transformé. Cette approche permet une estimation positive de la variance de transition au fil du temps. Cette stratégie est particulièrement utile pour assurer la stabilité numérique lors de la mise à jour adaptative [12, 21].

Nous procédons ainsi à une série d'expérimentations numériques pour analyser, d'une part, le comportement de l'erreur quadratique moyenne \bar{m}_t , log-vraisemblance ℓ_t et d'autre part, l'évolution temporelle de ε_t selon différentes valeurs de μ_ξ . Cette analyse permet d'évaluer à la fois l'impact du pas d'adaptation sur la précision du filtre et sur la stabilité dynamique du paramètre ε_t , en s'appuyant sur l'équation (3.27).

3.1.6.1 Expérimentation numérique

Cette section présente les résultats obtenus à partir de simulations numériques réalisées avec diverses valeurs du paramètre d'adaptation μ_ξ . L'objectif est d'évaluer la capacité du filtre SKF à estimer dynamiquement la variance de transition ε_t , et de mettre en évidence l'effet de l'initialisation de l'état sur la stabilité et la convergence du filtre.

Ces simulations ont été effectuées à partir d'un modèle dynamique en fixant la variance de transition à $\varepsilon^* = 1 \cdot 10^{-10}$. Dans le but d'examiner l'influence de l'initialisation sur la trajectoire de convergence et la robustesse du filtre SKF, trois valeurs initiales pour ε_0 ont été considérées, avec v_0 constant à $8 \cdot 10^{-6}$, où chaque style de ligne représente un scénario d'initialisation distinct, ce qui permet de comparer directement la sensibilité du filtre au choix des paramètres initiaux et au pas d'adaptation μ_ξ .

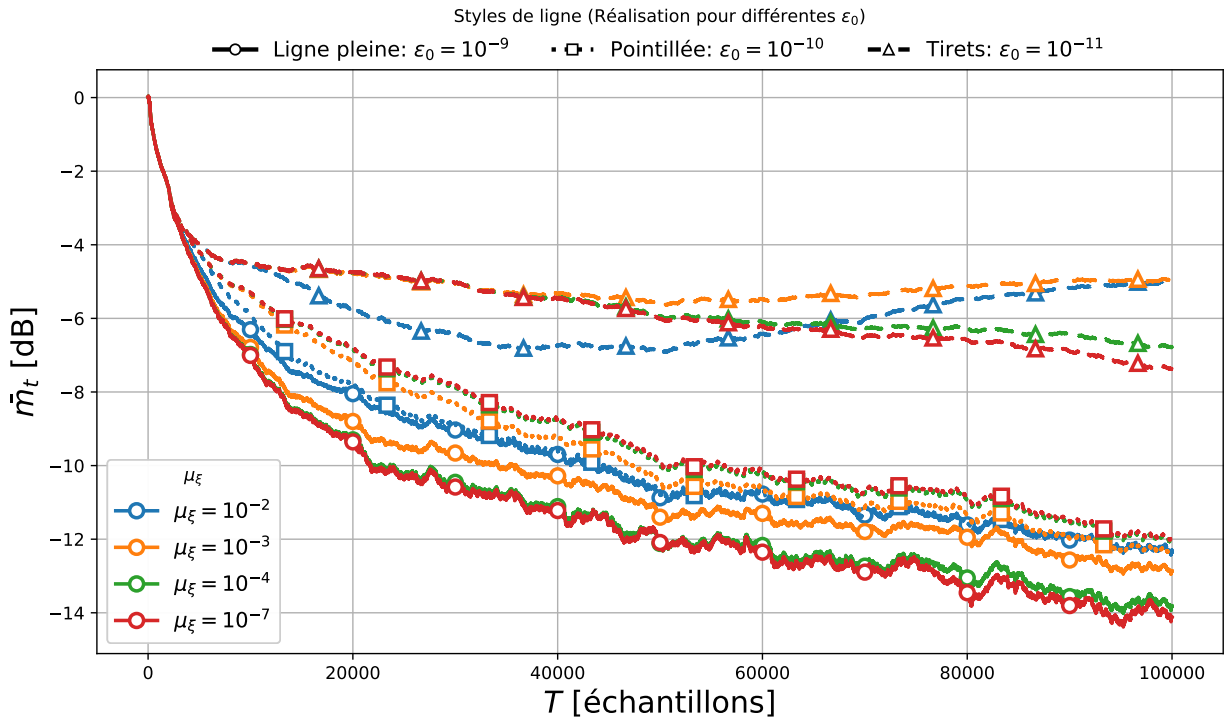


Figure 3.16: Erreur quadratique moyenne \tilde{m}_t pour différentes valeurs de μ_ξ lors de l'évaluation temporelle de h_t avec $\varepsilon^* = 10^{-10}$ et $\beta = 1$, selon trois scénarios initiaux $\varepsilon_0 \in \{10^{-9}, 10^{-10}, 10^{-11}\}$ et $v_0 = 8 \cdot 10^{-6}$, dans le cadre de l'application du filtre SKF.

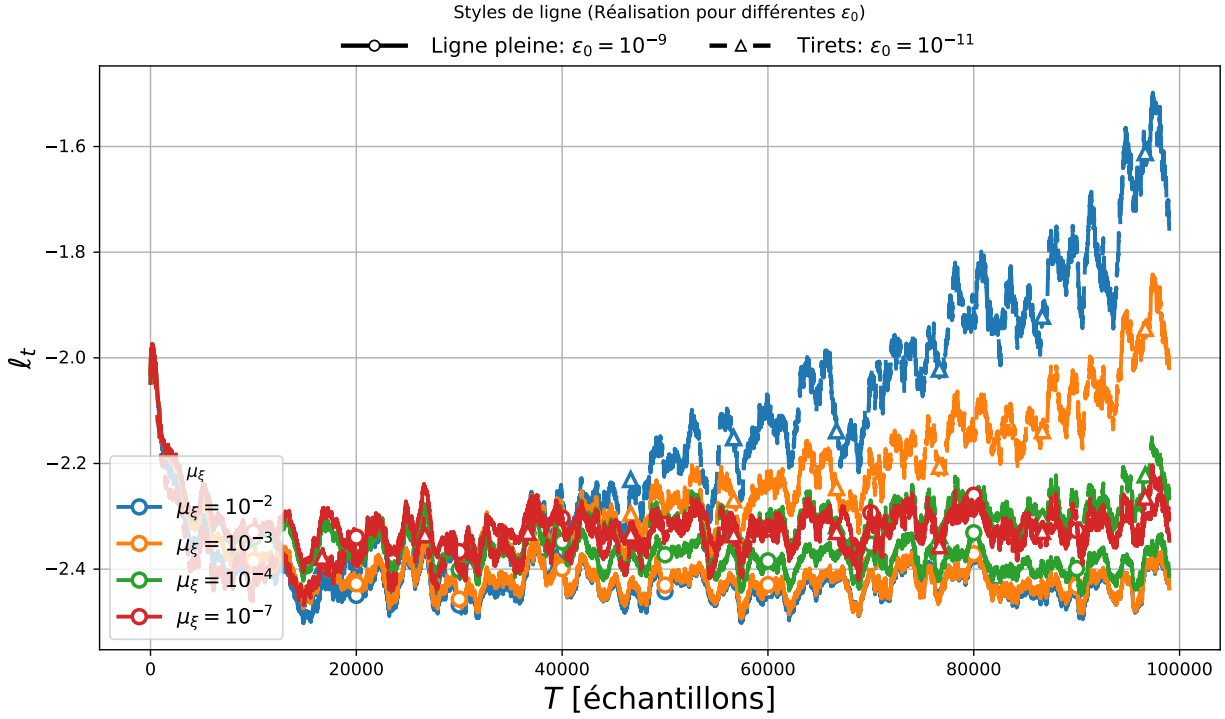


Figure 3.17: Évolution de la moyenne glissante de la log-vraisemblance \bar{l}_t pour différentes valeurs de μ_ξ lors de l'évaluation temporelle de h_t avec $\varepsilon^* = 10^{-10}$, selon deux scénarios initiaux $\varepsilon_0 \in \{10^{-9}, 10^{-11}\}$ et $v_0 = 8 \cdot 10^{-6}$.

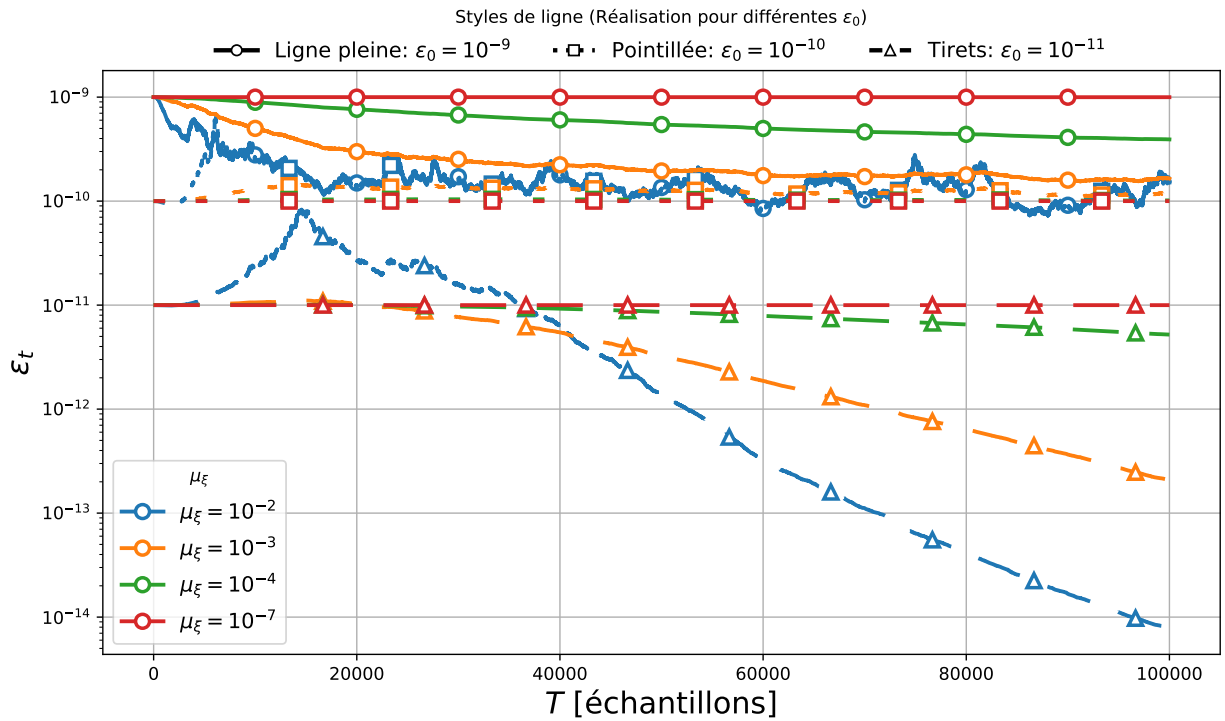


Figure 3.18: Évolution temporelle de ε_t pour différentes valeurs de μ_ξ lors de l'évaluation temporelle de h_t avec $\varepsilon^* = 10^{-10}$, selon trois scénarios initiaux $\varepsilon_0 \in \{10^{-9}, 10^{-10}, 10^{-11}\}$ et $v_0 = 8 \cdot 10^{-6}$.

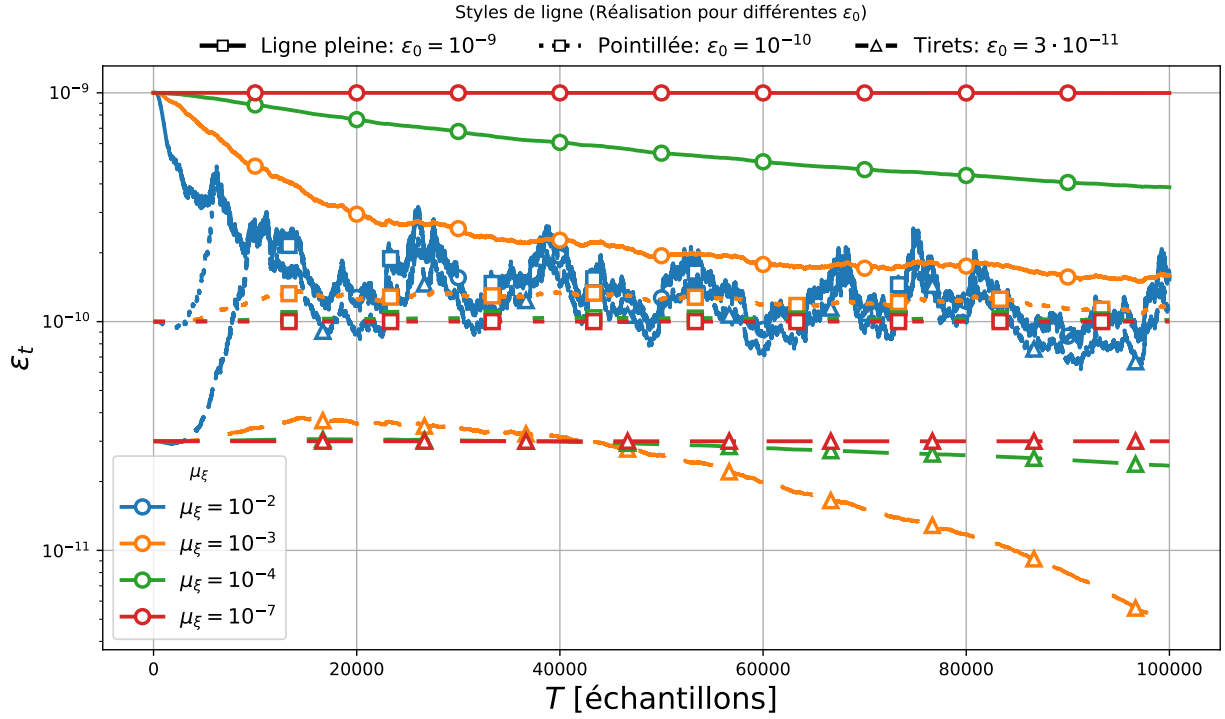


Figure 3.19: Évolution temporelle de ε_t pour différentes valeurs de μ_ξ lors de l'évaluation temporelle de h_t avec $\varepsilon^* = 10^{-10}$, selon trois scénarios initiaux $\varepsilon_0 \in \{10^{-9}, 10^{-10}, 3 \cdot 10^{-11}\}$ et $v_0 = 8 \cdot 10^{-6}$.

L'analyse des Figures présentes dans cette section met en évidence que l'algorithme d'estimation adaptative de la variance de transition (Eq. 3.27) guide systématiquement ε_t vers la valeur cible $\varepsilon^* = 10^{-10}$, et ce pour des jeux d'initialisation (ε_0) avec v_0 fixé. On observe en parallèle une baisse de la log-vraisemblance négative et une stabilisation de \bar{m}_t . Le pas d'adaptation μ_ξ module le compromis vitesse/stabilité : des valeurs plus élevées accélèrent la convergence mais peuvent induire des oscillations transitoires, tandis que des valeurs plus faibles accroissent la stabilité au prix d'une convergence plus lente. Globalement, ces résultats attestent de la capacité du schéma adaptatif à identifier la variance de transition et à améliorer la performance du filtre.

3.2 Estimation adaptative des hyper-paramètres par lissage temporel.

Dans cette étude, nous avons commencé par estimer le paramètre ε à chaque itération en utilisant l'équation (3.28) qui permet d'adapter dynamiquement ε en fonction des variations instantanées entre les états successifs.

La variance du bruit de processus ε_t , qui reflète les changements dynamiques des états, est calculée comme suit :

$$\hat{\varepsilon}_t = \frac{1}{M} \|\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-1}\|^2, \quad (3.28)$$

où M est la taille de la réponse impulsionnelle et la dimension du vecteur d'état, et $\|\cdot\|^2$ désigne la norme euclidienne au carré. Cette formulation permet de quantifier les fluctuations des états estimés entre deux instants consécutifs, comme détaillé dans [18, Sec. VIII].

Le hyper-paramètre $\hat{\varepsilon}_t$ doit être ajusté pour équilibrer une précision de suivi optimale et une minimisation de l'erreur quadratique moyenne. Les variations du système dynamique rendent l'utilisation de valeurs constantes pour ces paramètres ne permettent pas de garantir un compromis satisfaisant. Sur la base de cette motivation, l'estimation donnée par l'équation (3.28) est conçue pour atteindre cet objectif. Lorsque l'algorithme commence à converger ou lorsqu'il y a un changement brusque dans le système, la différence entre $\boldsymbol{\mu}_t$ et $\boldsymbol{\mu}_{t-1}$ devient significative. Dans ce cas, le paramètre $\hat{\varepsilon}_t$ prend des valeurs élevées, permettant ainsi une convergence rapide et un suivi efficace. En revanche, lorsque l'algorithme commence à atteindre son régime permanent, la différence entre $\boldsymbol{\mu}_t$ et $\boldsymbol{\mu}_{t-1}$ diminue, ce qui entraîne de faibles valeurs de $\hat{\varepsilon}_t$ et, par conséquent, une faible erreur quadratique moyenne.

Ces hyper-paramètres doivent être ajustés finement pour garantir des estimations cohérentes et minimiser l'erreur globale.

Cette relation permet de mettre à jour dynamiquement l'hyper-paramètre ε_t améliorant ainsi l'efficacité et la robustesse des prédictions du filtre de Kalman simplifié (SKF). Cette approche a été comparée à des algorithmes où ε est maintenu constant pour différentes valeurs fixes.

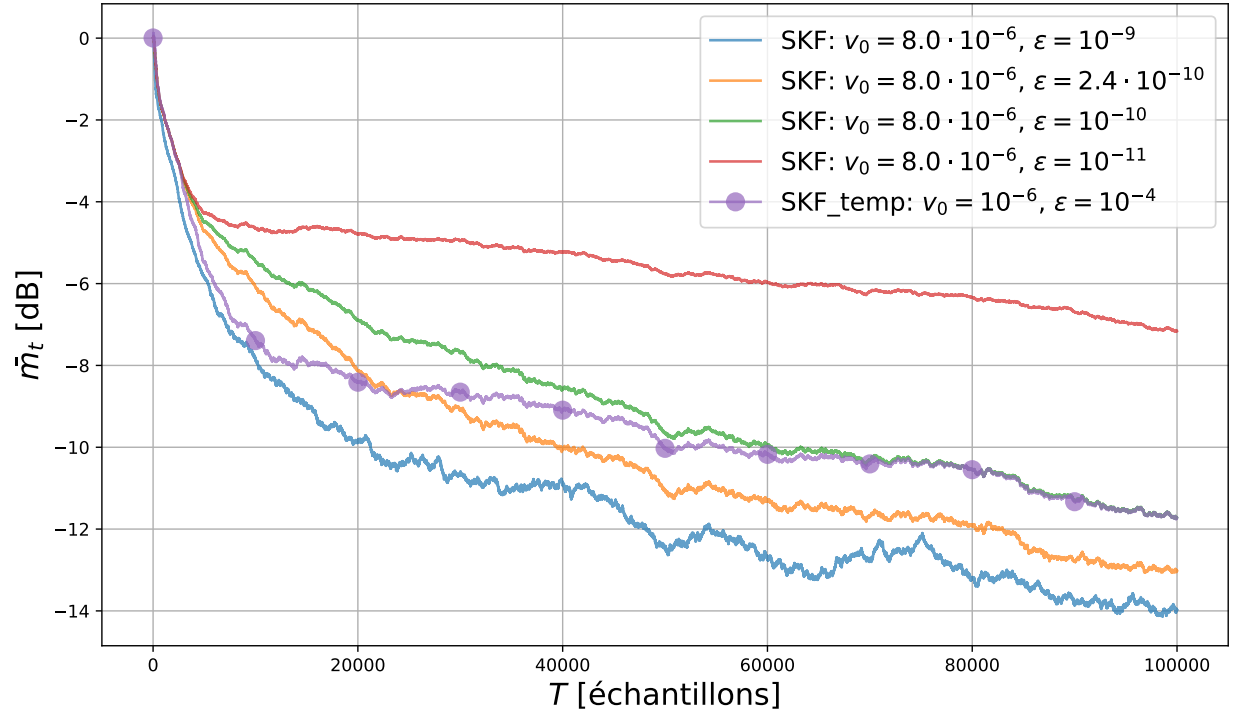


Figure 3.20: Erreur quadratique moyenne \tilde{m}_t , en comparant SKF avec différentes valeurs ε fixe et SKF avec ε estimée temporellement (SKF_temp) suivant l'équation (3.28), lors de l'évaluation temporelle de h_t avec $\varepsilon^* = 10^{-10}$ et $\beta = 1$

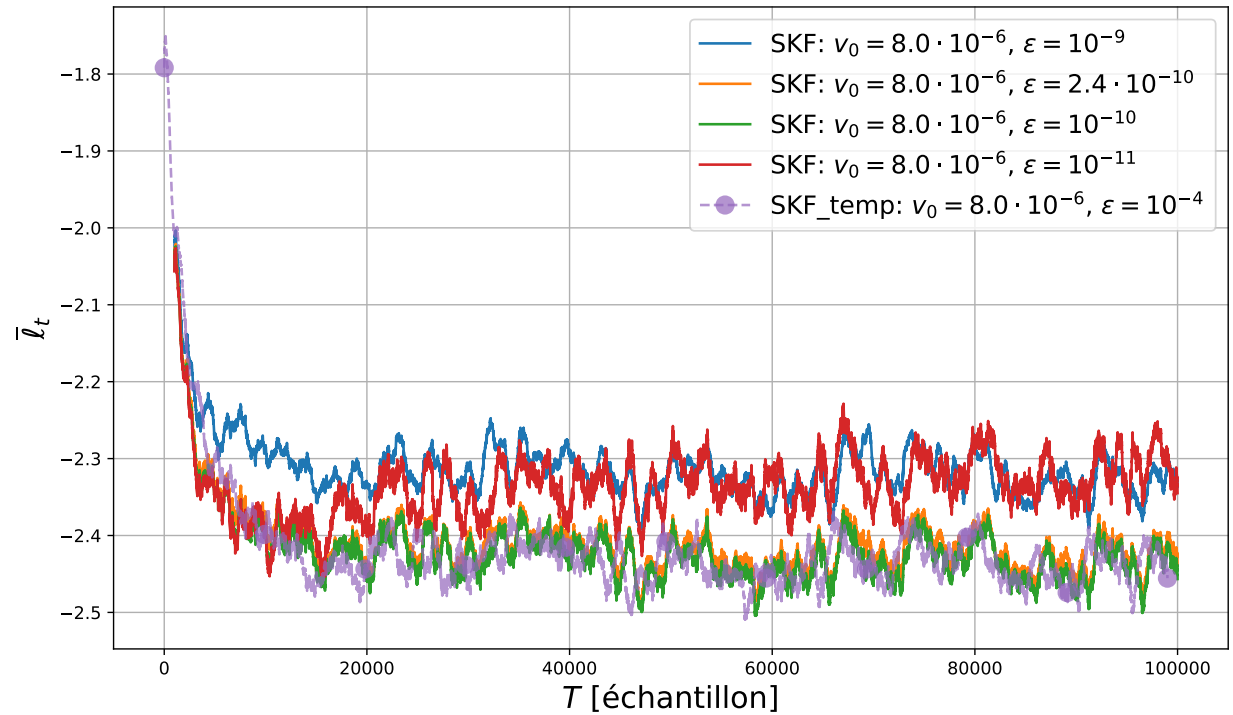


Figure 3.21: Évolution de la moyenne glissante de la log-vraisemblance $\bar{\ell}_t$ pour différentes valeurs de ε fixe et SKF avec ε estimée temporellement (SKF_temp) suivant l'équation (3.28), lors de l'évaluation temporelle de h_t avec $\varepsilon^* = 10^{-10}$ et $\beta = 1$

Afin d'améliorer la robustesse de l'estimation de ε , nous proposons une autre stratégie basée sur le lissage temporel. En suivant [18], nous reprenons l'estimateur instantané de la variance de transition, Eq.(3.28). Par extension, nous proposons une version lissée sur une fenêtre temporelle, donnée par

$$\hat{\varepsilon}_t = \frac{1}{MJ} \sum_{j=1}^{J-1} \|\mu_{t-j} - \mu_{t-1-j}\|^2, \quad (3.29)$$

où J correspond à la taille de la fenêtre de moyennage. Cette méthode vise à réduire l'impact des variations locales tout en conservant les tendances globales.

Les résultats obtenus en appliquant cette méthode de lissage temporel sont présentés dans les figures ci-dessous, où une comparaison entre l'approche classique avec ε constant et celle avec ε estimé temporellement est effectuée. Ces figures mettent en évidence les différences de performances en termes de rapidité de convergence de l'erreur quadratique moyenne \bar{m}_t ,

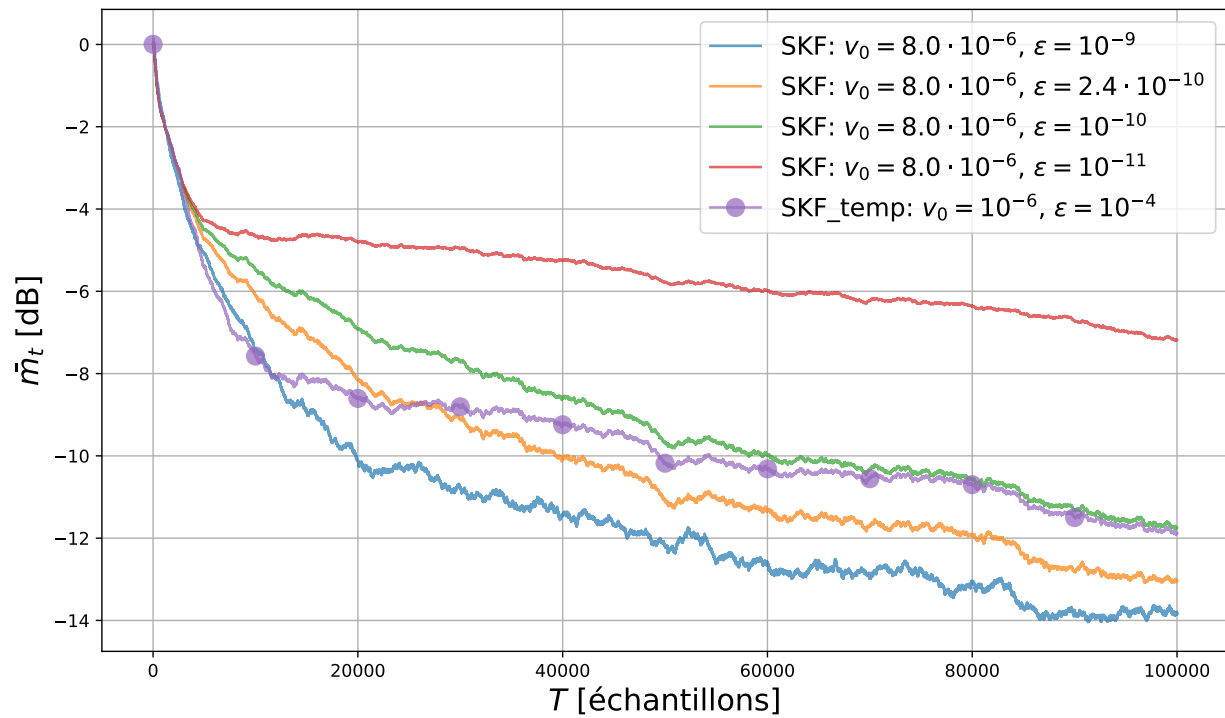


Figure 3.22: Erreur quadratique moyenne \bar{m}_t obtenue en comparant le SKF avec ε constant à une version où ε est estimé par lissage temporel (SKF_temp : équation (3.29), fenêtre $J = 10$), lors de l'évaluation temporelle de h_t avec $\varepsilon^* = 10^{-10}$ et $\beta = 1$

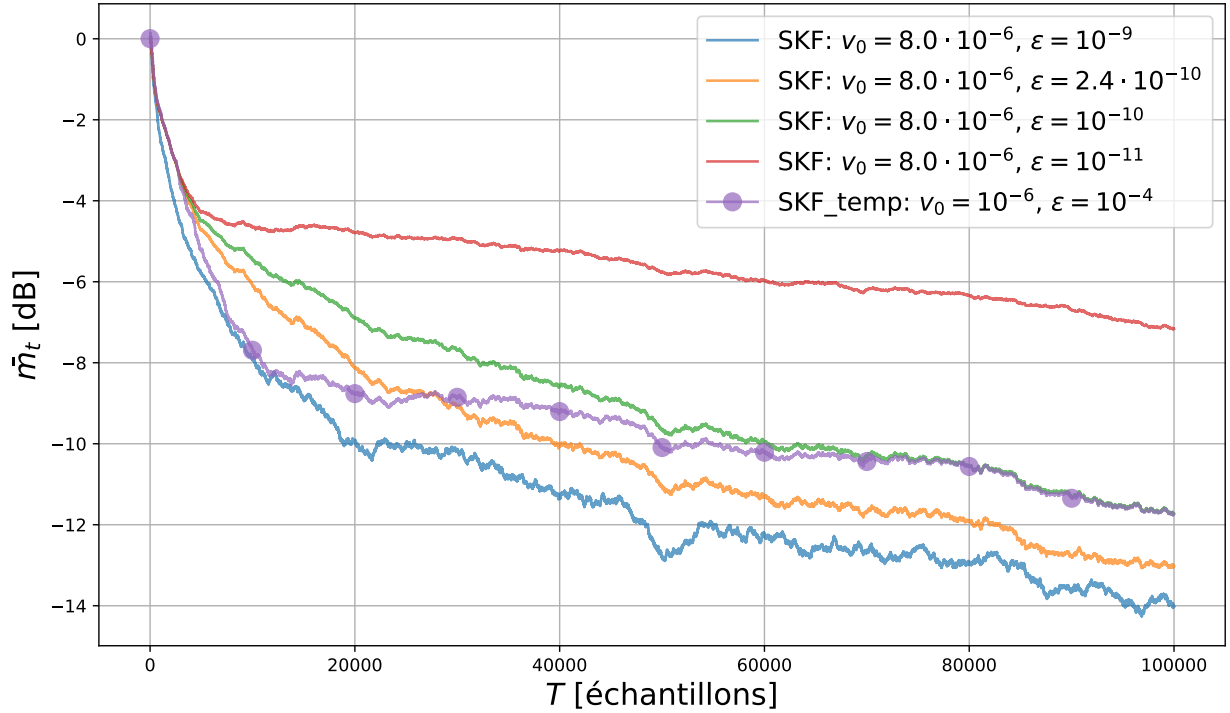


Figure 3.23: Erreur quadratique moyenne \tilde{m}_t , en comparant le SKF avec ε constant à une version de SKF avec ε estimé par lissage temporel (SKF_temp) selon l'équation (3.29), avec une fenêtre de lissage de taille $J = 100$, dans le cadre du modèle dynamique ($\varepsilon^* = 10^{-10}$ et $\beta = 1$).

Ces comparaisons permettent d'évaluer l'impact de la stratégie de lissage temporel sur les performances du filtre dans différentes réalisations, en variant la taille de la fenêtre de moyennage J . Cette analyse met en lumière comment cette approche influence la capacité de la méthode à réduire les fluctuations locales tout en préservant les tendances globales. Les mêmes simulations seront reprises dans la section suivante pour le modèle statique où $\varepsilon^* = 0$, afin de comparer les performances obtenues dans ce cadre particulier.

3.2.1 Analyse comparative avec le modèle statique ($\varepsilon^* = 0$ et $\beta = 1$)

Dans cette section, nous examinons le comportement du filtre dans un contexte statique, où la variance de transition est fixée à zéro ($\varepsilon^* = 0$). En reprenant la même méthodologie de simulation et de comparaison présentée précédemment. Cette analyse vise à déterminer si les avantages observés avec le lissage temporel se maintiennent également lorsque les paramètres du système restent constants.

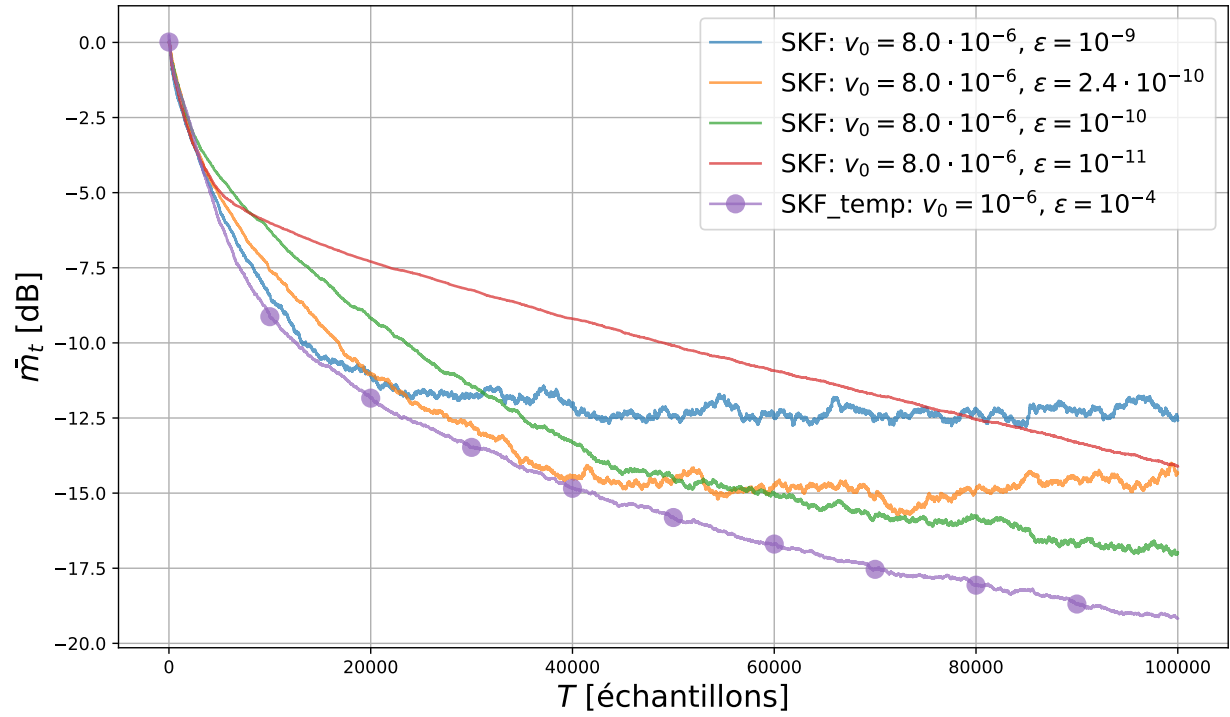


Figure 3.24: Erreur quadratique moyenne \bar{m}_t , en comparant SKF avec différentes valeurs ε fixe et SKF avec ε estimée temporellement (SKF_temp) suivant l'équation (3.28). avec des données issues d'un modèle statique ($\varepsilon^* = 0$ et $\beta = 1$)

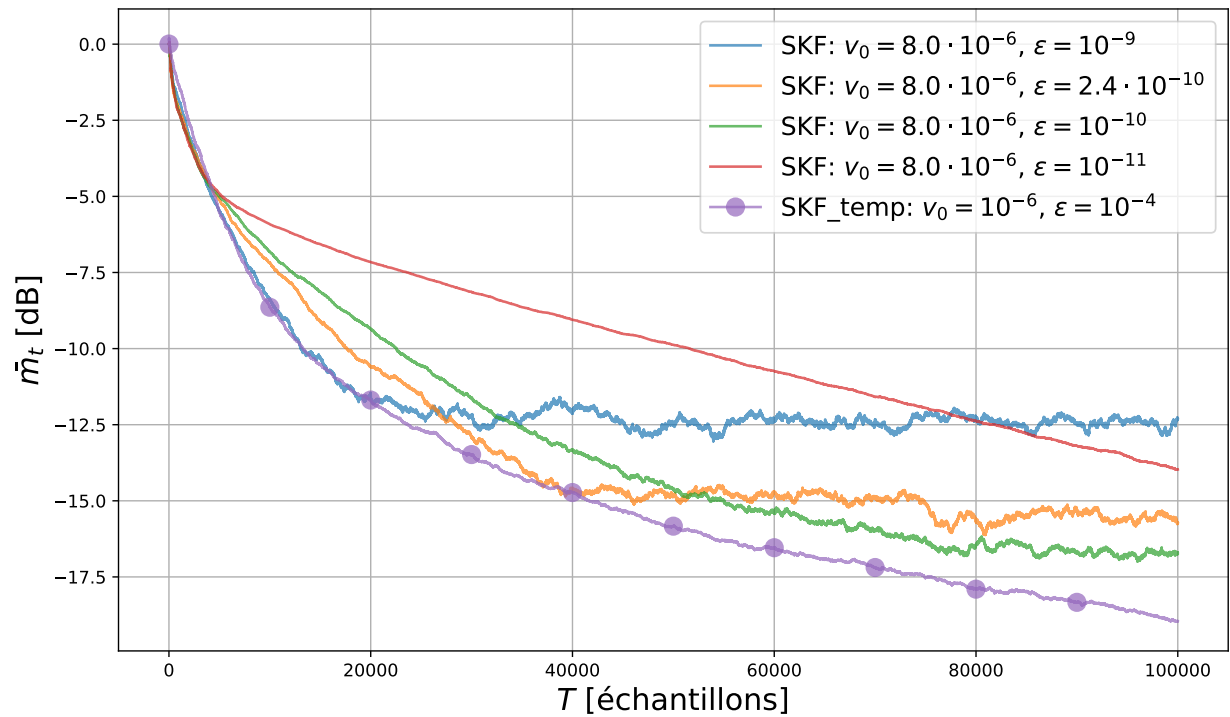


Figure 3.26: Erreur quadratique moyenne \bar{m}_t obtenue en comparant SKF avec ε constant à une version où ε est estimé par lissage temporel (SKF_temp : équation (3.29), fenêtre $J = 10$), avec des données issues d'un modèle statique ($\varepsilon^* = 0$ et $\beta = 1$).

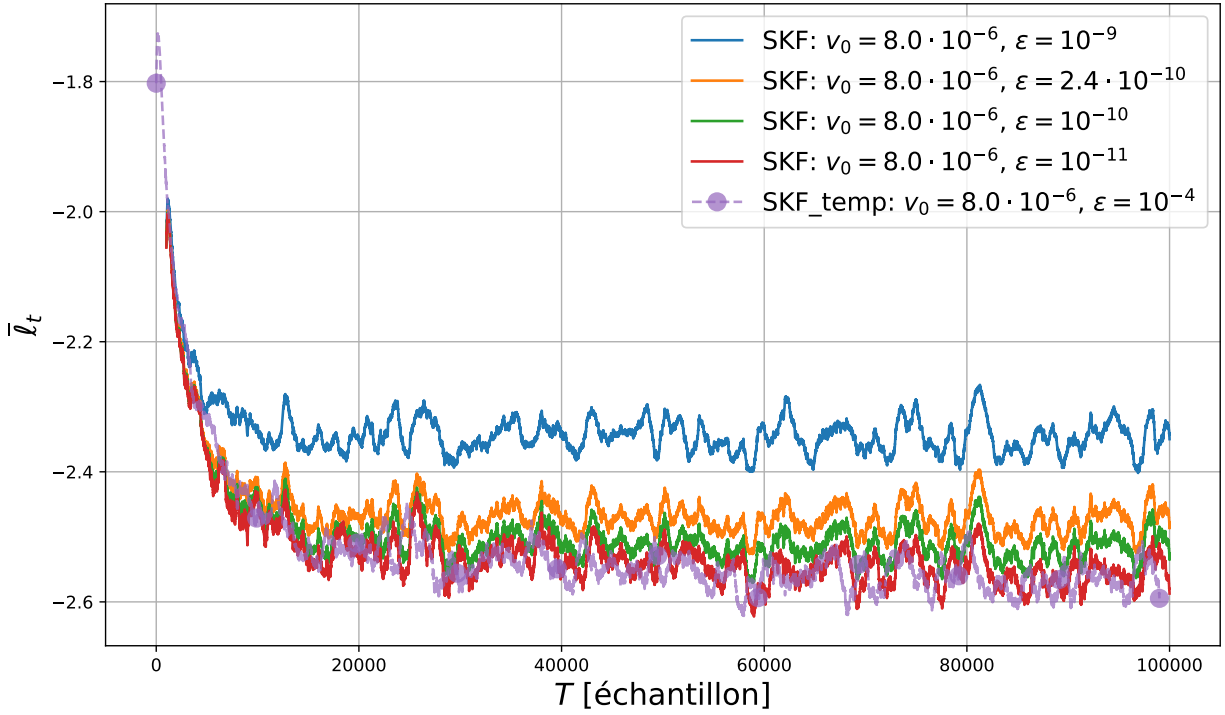


Figure 3.25: Évolution de la moyenne glissante de la log-vraisemblance $\bar{\ell}_t$ pour différentes valeurs de ε fixe et SKF avec ε estimée temporellement (SKF_temp) suivant l'équation (3.28), avec des données issues d'un modèle statique ($\varepsilon^* = 0$ et $\beta = 1$)

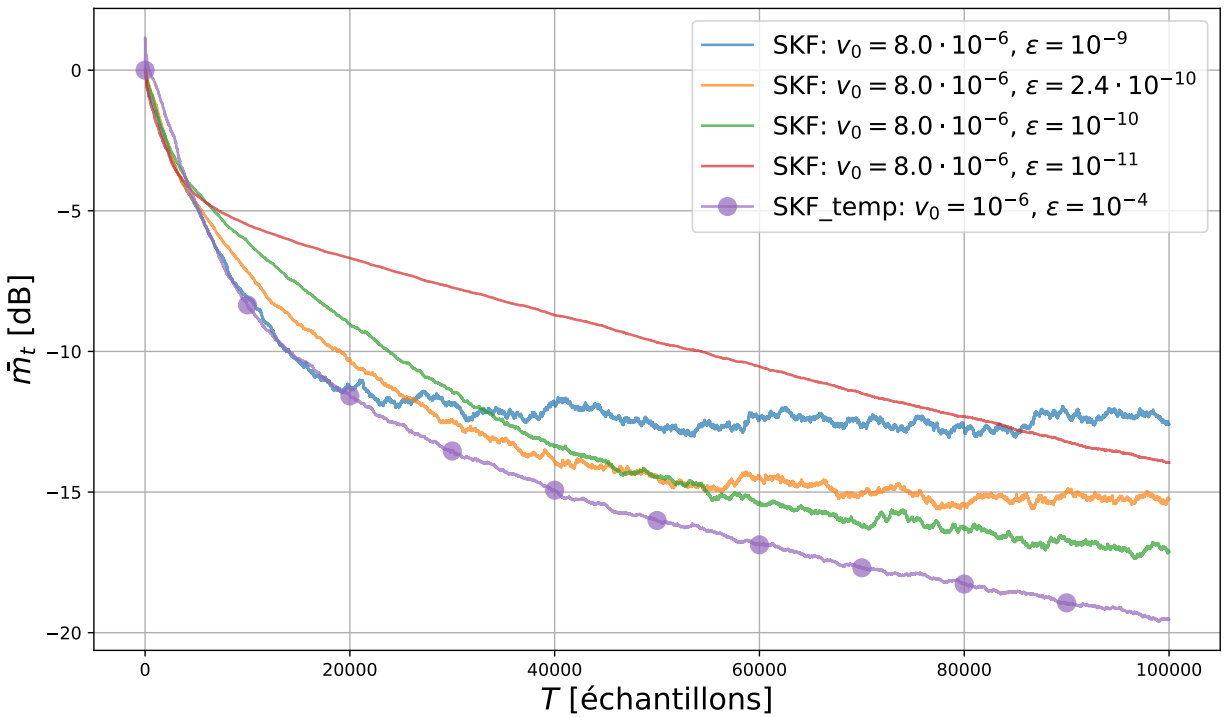


Figure 3.27: Erreur quadratique moyenne \tilde{m}_t comparant le SKF à ε constant et une estimation par lissage temporel SKF_temp ($J = 100$) selon (3.29), pour un modèle statique ($\varepsilon^* = 0$ et $\beta = 1$).

Les résultats illustrés par les figures montrent que, contrairement à l'hypothèse initiale, l'estimation adaptative de ε_t selon les équations (3.28) et (3.29) n'assure pas systématiquement une convergence plus rapide que l'estimation fixe. On observe que l'estimation temporelle (représentée par la courbe violette avec marqueurs) atteint un niveau de perte minimal plus bas et une erreur quadratique moyenne plus faible en moyenne.

3.3 Observations

Dans ce chapitre, nous avons examiné deux approches adaptatives d'estimation en ligne des hyper-paramètres du filtre de Kalman. Ces approches jouent un rôle crucial pour améliorer la précision et l'efficacité des estimations fournies par le filtre.

À partir des résultats présentés, plusieurs observations peuvent être faites :

1. La log-vraisemblance a posteriori constitue un critère robuste pour déterminer les hyper-paramètres optimaux du modèle. Comme montré explicitement dans la Section 3.1.5, cette métrique fournit une indication fiable et facilite l'identification des hyper-paramètres (ε et v_η).
2. Cependant, il est important de noter que la log-vraisemblance n'est pas directement aux deux autres métriques telles que l'erreur de prédiction e_t ou l'erreur quadratique moyenne m_t . En conséquence, la maximisation directe de cette log-vraisemblance ne permet pas nécessairement d'accélérer la convergence du filtre. Cette limitation est mise en évidence dans la Section 3.1.6, où les méthodes heuristiques d'adaptation dynamique l'hyper-paramètre (ε), fondées sur des critères empiriques, surpassent les approches bayésiennes en termes de vitesse et de performance de convergence.
3. L'analyse révèle également une sensibilité importante des méthodes d'adaptation en ligne à l'initialisation des hyper-paramètres. Ces méthodes sont fortement affectées par le choix initial des valeurs de ces paramètres. Cette étude, bien qu'instructive, n'est pas entièrement conclusive à ce stade et nécessiterait des recherches supplémentaires afin d'approfondir la compréhension des résultats obtenus.

En résumé, ces approches adaptatives d'estimation en ligne étudiées ici démontrent leur capacité à s'ajuster efficacement aux variations dynamiques du système analysé. Elles permettent une

réduction significative de l'erreur quadratique moyenne \bar{m}_t et de la log-vraisemblance moyenne $\bar{\ell}_t$ ainsi que de l'erreur de prédiction e_t . En comparaison avec les approches statiques traditionnelles, ces méthodes offrent une alternative robuste et prometteuse, à condition d'un réglage attentif et adapté des hyper-paramètres.

3.4 Conclusion

Ce chapitre a présenté deux approches pour l'estimation des hyper-paramètres du Kalman via (i) une MLE avec descente de gradient (positivité assurée) et (ii) une estimation en ligne lissée de ε_t fondée sur les écarts d'états. Les expériences, en régimes statique et dynamique, montrent que l'adaptation de ε_t améliore le suivi mais reste sensible au pas μ_ξ et à l'initialisation.

Chapitre 4

Conclusion

Dans ce mémoire, nous avons proposé et analysé une approche unifiée pour l'estimation dynamique dans les modèles linéaires, en nous appuyant sur le **filtre de Kalman** et ses variantes. Nous avons d'abord rappelé les fondements théoriques du filtrage de Kalman dans un cadre bayésien. Cette section a détaillé la formulation probabiliste du filtre, son implémentation récursive, ainsi que différentes simplifications possibles : covariance matricielle, vectorielle, scalaire et à variance fixe [13]. Nous avons également mis en évidence l'équivalence, sous certaines hypothèses, entre la mise à jour bayésienne du filtre et des méthodes classiques comme la descente de gradient stochastique (SGD) [19].

Les simulations menées pour l'identification de la réponse impulsionnelle dans un contexte acoustique ont validé l'efficacité des méthodes proposées. Les différents scénarios ont mis en évidence les compromis entre vitesse de convergence, précision asymptotique et robustesse au bruit.

Et puis, nous avons proposé et analysé différentes stratégies pour l'estimation des hyper-paramètres du **filtre de Kalman** appliqué à l'identification dynamique de la réponse impulsionnelle acoustique.

Nous avons porté une attention particulière à l'analyse comparative des principales stratégies d'estimation des hyper-paramètres, à savoir:

- l'optimisation par maximisation de la vraisemblance (MLE) [14];
- l'estimation adaptative basée sur l'évolution temporelle des états;

- l'introduction d'un lissage temporel afin d'atténuer les fluctuations locales des paramètres estimés [18].

Ces résultats mettent en évidence plusieurs axes d'amélioration et de recherche, tant sur l'optimisation des hyper-paramètres que sur l'extension du cadre méthodologique. Ainsi, plusieurs pistes d'amélioration et d'exploration peuvent être envisagées pour les travaux futurs :

1. Optimisation avancée des hyper-paramètres

- Étudier des méthodes d'apprentissage automatique (ex. Bayesian Optimization) pour sélectionner dynamiquement les valeurs optimales des hyper-parameters du modèle. [30].
- Explorer des techniques de régularisation adaptative pour stabiliser encore plus l'estimation des poids [16, 31].

2. Généralisation aux modèles non linéaires

- Étendre la méthodologie à des modèles de Kalman non linéaires, comme le filtre de Kalman étendu (EKF) ou le filtre de Kalman unscented (UKF) [4, 32].
- Intégrer des méthodes à base de filtre particulaire pour les cas fortement non linéaires ou non gaussiens [23].

3. Comparaison avec des méthodes d'apprentissage profond

- Comparer le filtre de Kalman adaptatif aux approches récentes, telles que les RNN et transformers pour séries temporelles [24, 33].
- Explorer des approches hybrides combinant filtre de Kalman et réseaux neuronaux, notamment par la fusion bayésienne des prédictions afin de tirer parti des forces respectives de chaque méthode.

En conclusion, cette étude met en évidence le potentiel du filtrage de Kalman adaptatif pour l'estimation dynamique des paramètres. Cependant, plusieurs défis restent à relever, notamment l'optimisation des hyper-paramètres et l'adaptation du modèle à des contextes plus complexes. Les perspectives identifiées permettront d'améliorer davantage ces techniques et d'en élargir les applications à des domaines variés.

Références

- [1] S. Haykin, *Kalman Filtering and Neural Networks*, Wiley-Interscience, 2001.
- [2] P. S. Maybeck, *Stochastic Models, Estimation, and Control*, Academic Press, New York, 1979.
- [3] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
- [4] M. H. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice Using MATLAB*, 4th ed. Wiley, 2014.
- [5] A. C. Harvey, “Forecasting, structural time series models and the kalman filter,” *Cambridge University Press*, 1990.
- [6] S. Thrun, W. Burgard, and D. Fox, “Probabilistic robotics,” *MIT Press*, 2005.
- [7] F. Gustafsson, “Adaptive filtering and change detection,” *Wiley*, 2001.
- [8] A. H. Sayed, *Adaptive Filters*, John Wiley & Sons, Hoboken, New Jersey, 2008.
- [9] S. Ruder, “An overview of gradient descent optimization algorithms,” *arXiv preprint arXiv:1609.04747*, 2016.
- [10] M. Dogariu, J. Benesty, C. Paleologu, and S. Ciochină, “An insightful overview of the wiener filter for system identification,” *Applied Sciences*, vol. 11, no. 17, 2021. [Online]. Available: <https://www.mdpi.com/2076-3417/11/17/7774>
- [11] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*, Wiley, 2006.
- [12] S. Särkkä, *Bayesian Filtering and Smoothing*, Cambridge University Press, 2013.
- [13] L. Szczecinski and R. Tihon, “Simplified kalman filter for on-line rating: one-fits-all approach,” *Journal of Quantitative Analysis in Sports*, vol. 19, no. 4, pp. 1–20, June 2023, institut National de la Recherche Scientifique et Université de Montréal.
- [14] R.-J. Lange, “Bellman filtering and smoothing for state-space models,” *Journal of Econometrics*, vol. 238, no. 2, p. 105632, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0304407623003482>
- [15] L. Szczecinski, J. Benesty, and E. Kuhn, “A unified bayesian perspective for conventional and robust adaptive filters,” 02 2025.

- [16] S. Haykin, *Adaptive Filter Theory*, 4th ed. Prentice Hall, 2002.
- [17] L. Bottou, “Large-scale machine learning with stochastic gradient descent,” in *Proceedings of COMPSTAT’2010*. Springer, 2010, pp. 177–186.
- [18] C. Paleologu, J. Benesty, and S. Ciochină, “Study of the general kalman filter for echo cancellation,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 8, pp. 1539–1549, 2013.
- [19] S. Ruder, “An overview of gradient descent optimization algorithms,” *arXiv preprint arXiv:1609.04747*, 2016.
- [20] T. Moon and W. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Upper Saddle River, NJ: Prentice Hall, 2000.
- [21] D. Barber, *Bayesian reasoning and Machine Learning*. New York: Cambridge University Press, 2012.
- [22] Piotrg commonswiki, “Multivariate_gaussian.png,” Wikimedia Commons, 2006, creative Commons Attribution–Share Alike 3.0 Unported (CC BY-SA 3.0). [Online]. Available: https://commons.wikimedia.org/wiki/File:Multivariate_Gaussian.png
- [23] A. Doucet, N. de Freitas, and N. Gordon, “An introduction to sequential monte carlo methods,” in *Sequential Monte Carlo Methods in Practice*. New York, NY: Springer, 2001, pp. 3–14, [Online]. [Online]. Available: https://doi.org/10.1007/978-1-4757-3437-9_1
- [24] C. M. Bishop, *Pattern Recognition and Machine Learning*, ser. Information Science and Statistics. New York: Springer, 2006.
- [25] S. Theodoridis, “Chapter 2 - probability and stochastic processes,” in *Machine Learning*, S. Theodoridis, Ed. Oxford: Academic Press, 2015, pp. 9–51. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780128015223000021>
- [26] G. Enzner, “Bayesian inference model for applications of time-varying acoustic system identification,” in *2010 18th European Signal Processing Conference (EUSIPCO)*, 2010, pp. 2126–2130.
- [27] N. Werner, “audiolabs/rir-generator: Version 0.2.0,” <https://github.com/audiolabs/rir-generator>, 2023.
- [28] D. Zanco, L. Szczecinski, and J. Benesty, “Automatic regularization for linear mmse filters,” *Signal Processing*, vol. 230, pp. 1–10, Dec. 2024.
- [29] S. J. Koopman, A. Lucas, and M. Scharth, “Predicting time-varying parameters with parameter-driven and observation-driven models,” *The Review of Economics and Statistics*, vol. 98, no. 1, pp. 97–110, Mar. 2016, [Online]. [Online]. Available: https://doi.org/10.1162/REST_a_00533
- [30] E. Brochu, V. M. Cora, and N. de Freitas, “A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning,” 2010, arXiv preprint arXiv:1012.2599.
- [31] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice Hall, 1985.

- [32] S. J. Julier and J. K. Uhlmann, “Unscented filtering and nonlinear estimation,” *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.
- [33] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.