

# Identification d'un systeme non-lineaire partiellement observe par la methode de la distance minimale

Yurii Kutoyants, François Le Gland, Rivo Rakotozafy

► **To cite this version:**

Yurii Kutoyants, François Le Gland, Rivo Rakotozafy. Identification d'un systeme non-lineaire partiellement observe par la methode de la distance minimale. [Rapport de recherche] RR-2125, INRIA. 1993. <inria-00074547>

**HAL Id: inria-00074547**

**<https://hal.inria.fr/inria-00074547>**

Submitted on 24 May 2006

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

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

***Identification d'un Système Non-Linéaire  
Partiellement Observé par la  
Méthode de la Distance Minimale***

Yurii Kutoyants

François Le Gland

Rivo Rakotozafy

**N° 2125**

Décembre 1993

PROGRAMME 5

Traitement du signal,  
automatique  
et productique



***R*** ***apport  
de recherche***

1993





# Identification d'un Système Non-Linéaire Partiellement Observé par la Méthode de la Distance Minimale

Yurii Kutoyants \* \*\*  
François Le Gland \*\*\*  
Rivo Rakotozafy \*\*\*\*

Programme 5 — Traitement du signal, automatique et productique  
Projet Mefisto

Rapport de recherche n° 2125 — Décembre 1993 — 32 pages

**Résumé :** Nous considérons le problème de l'estimation de paramètres par la méthode de la distance minimale, pour les processus de type diffusion partiellement observés. Nous construisons l'estimateur de la distance minimale (EDM) à partir de  $n$  observations indépendantes, et nous montrons la consistance et la normalité asymptotique de l'EDM quand  $n$  tend vers l'infini. Nous étudions ensuite le cas particulier des systèmes linéaires, et nous présentons quelques résultats numériques.

**Mots-clé :** processus de type diffusion, observation partielle, estimation paramétrique, estimateur de la distance minimale.

*(Abstract: pto)*

\*. Département de Mathématiques, Université du Maine, BP 535, 72017 Le Mans Cedex, France.

\*\* Université de Paris VI, URA CNRS 1321, 4 place Jussieu, 75252 Paris Cedex 05, France.

\*\*\* IRISA, Campus Universitaire de Beaulieu, 35042 Rennes Cedex, France.

\*\*\*\* INRIA Sophia-Antipolis, BP 93, 06902 Sophia-Antipolis Cedex, France.

Unité de recherche INRIA Sophia-Antipolis  
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)  
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

# Identification of a Partially Observed Non-Linear System by the Minimum Distance Method

**Abstract:** We consider the problem of parameter estimation by the minimum distance method, for partially observed diffusion type processes. We construct the minimum distance estimate (MDE) based on  $n$  independent observations, and we prove consistency and asymptotic normality for the MDE when  $n$  goes to infinity. Then, we study the particular case of linear systems, and we present some numerical results.

**Key-words:** diffusion type process, partially observation, parameter estimation, minimum distance estimate.

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>EDM pour les processus de type diffusion</b>	<b>5</b>
2.1	Consistance de l'EDM . . . . .	7
2.2	Normalité asymptotique de la moyenne empirique . . . . .	11
2.3	Normalité asymptotique de l'EDM . . . . .	14
<b>3</b>	<b>EDM pour les diffusions linéaires</b>	<b>16</b>
3.1	Applications des résultats précédents . . . . .	18
<b>4</b>	<b>Applications numériques</b>	<b>21</b>
4.1	Modèle I . . . . .	22
4.2	Modèle II . . . . .	25
<b>5</b>	<b>Conclusion</b>	<b>28</b>



## 1 Introduction

Nous nous proposons d'étudier le problème de l'identification d'un processus partiellement observé de type diffusion. Pour toute valeur  $\theta \in \Theta$  du paramètre inconnu, nous définissons la probabilité  $\mathbf{P}_\theta$  sur l'espace canonique  $C([0, T]; \mathbb{R}^m \times \mathbb{R}^d)$  telle que sous  $\mathbf{P}_\theta$

$$\begin{cases} dX_t = b_t(\theta, X) dt + \sigma_t(\theta, X) dW_t^\theta, & X_0 = x_0 \\ dY_t = h_t(\theta, X) dt + dV_t^\theta, & Y_0 = 0, \end{cases} \quad (1)$$

où  $X = \{X_t, 0 \leq t \leq T\}$  et  $Y = \{Y_t, 0 \leq t \leq T\}$  sont deux processus à valeurs dans  $\mathbb{R}^m$  et  $\mathbb{R}^d$  respectivement,  $\Theta \subset \mathbb{R}^p$  est l'ensemble des paramètres,  $\{W_t^\theta, 0 \leq t \leq T\}$  et  $\{V_t^\theta, 0 \leq t \leq T\}$  sont deux processus de Wiener indépendants de  $\mathbb{R}^m$  et  $\mathbb{R}^d$  respectivement.

Nous faisons les hypothèses habituelles sur les coefficients, de façon à assurer l'existence et l'unicité de la solution du système (1), voir Liptser–Shiryayev [20].

Il s'agit d'estimer le paramètre inconnu  $\theta \in \Theta$  à partir de l'observation du seul processus  $Y$

Un certain nombre de résultats sont disponibles dans la littérature. La plupart de ces résultats concernent le comportement en temps long, quand  $T$  tend vers l'infini, de l'estimateur du maximum de vraisemblance  $\hat{\theta}_T$ , dans le cas particulier où le système est linéaire, c'est-à-dire que, sous  $\mathbf{P}_\theta$

$$\begin{cases} dX_t = B_\theta X_t dt + G_\theta dW_t^\theta, & X_0 = x_0 \\ dY_t = H_\theta X_t dt + dV_t^\theta, & Y_0 = 0, \end{cases} \quad (2)$$

Le premier résultat est dû à Balakrishnan [4, Chapitre VIII] qui a montré la consistance faible (en probabilité) de  $\hat{\theta}_T$ , sous des hypothèses de régularité et d'identifiabilité, et en supposant que, pour tout  $\theta \in \Theta$ ,  $F_\theta$  est stable et la paire  $(F_\theta, H_\theta)$  est complètement observable.

Bagchi [2], [3] a montré sous les mêmes hypothèses la consistance forte (p.s.) de  $\hat{\theta}_T$ , puis la normalité asymptotique et l'efficacité asymptotique de  $\hat{\theta}_T$ , c'est-à-dire la convergence en loi de  $\sqrt{T}(\hat{\theta}_T - \theta_0)$  vers un vecteur aléatoire gaussien centré de matrice de covariance  $Q^{-1}(\theta_0)$  où  $Q(\theta_0)$  est l'information de Fisher du modèle (2). Voir aussi la thèse de Le Breton [19, Chapitre III, §2]



Dans le cas simple où  $m = d = p = 1$ , Kutoyants [17, §3.4.4] a utilisé l'approche asymptotique locale d'Ibragimov–Khasminskii [12] pour montrer la normalité asymptotique locale du modèle (2) sous des hypothèses de régularité et de différentiabilité, et en déduire la consistance forte, la normalité asymptotique et l'efficacité asymptotique de  $\hat{\theta}_T$ , et la convergence des moments de  $\sqrt{T}(\hat{\theta}_T - \theta_0)$ .

Ce résultat a été ensuite généralisé par Kallianpur–Sedukar [15]. D'après ces auteurs, il est possible de prouver le même résultat dans le cas multidimensionnel, sous des hypothèses de régularité et de différentiabilité, et en supposant que pour tout  $\theta \in \Theta$ ,  $F_\theta$  est stable, la paire  $(F_\theta, H_\theta)$  est complètement observable, et la paire  $(F_\theta, G_\theta)$  est complètement contrôlable.

Parmi les autres références, on peut citer le travail récent de Bertrand [6].

Dans le cas général où le système (1) est non-linéaire, les résultats sont beaucoup moins nombreux. En ce qui concerne le comportement en temps long, quand  $T$  tend vers l'infini, de l'estimateur du maximum de vraisemblance  $\hat{\theta}_T$ , Borkar–Bagchi [8] utilisent la représentation

$$dY_t = m_t(\theta, Y) dt + dI_t^\theta,$$

où  $m_t(\theta, Y) = \mathbf{E}_\theta[h_\theta(X_t) | \mathcal{Y}_t]$  et  $I^\theta = \{I_t^\theta, t \geq 0\}$  est le processus d'innovation, pour montrer le résultat de consistance forte suivant

$$d(\hat{\theta}_T, D) \rightarrow 0 \quad \text{p.s.}$$

Dans cet énoncé, l'ensemble aléatoire  $D \subset \Theta$  est défini par :

$$D = \left\{ \theta \in \Theta : \liminf_{T \rightarrow \infty} \frac{1}{T} \int_0^T |m_t(\theta, Y) - m_t(\theta_0, Y)|^2 dt = 0 \right\}$$

Malheureusement, les conditions sous lesquelles ce résultat est démontré sont difficiles à vérifier en pratique, car elles portent sur le processus  $\{m_t(\theta, Y), t \geq 0\}$  et pas directement sur les coefficients du système (1).

Plus récemment, James–Le Gland [13], [14] ont considéré une asymptotique de type “petit bruit” :

$$\begin{cases} dX_t = b_\theta(X_t) dt + \varepsilon dW_t^{\theta, \varepsilon}, & X_0 = \bar{x}_0^\theta \\ dY_t = h_\theta(X_t) dt + \varepsilon dV_t^{\theta, \varepsilon}, & Y_0 = 0 \end{cases}$$

et ont montré le résultat de convergence faible suivant pour l'estimateur du maximum de vraisemblance  $\hat{\theta}_\varepsilon$

$$d(\hat{\theta}_\varepsilon, M(\theta_0)) \rightarrow 0$$

en probabilité, quand  $\varepsilon$  tend vers zéro. Dans cet énoncé :

$$M(\theta_0) = \left\{ \theta \in \Theta : \dot{y}_t^{\theta_0} = \dot{y}_t^\theta, \quad \text{pour tout } 0 \leq t \leq T \right\}$$

et pour tout  $\theta \in \Theta$ ,  $\{y_t^\theta, 0 \leq t \leq T\}$  est la sortie du système déterministe limite

$$\begin{cases} \dot{x}_t^\theta = b_\theta(x_t^\theta), & x_0^\theta = \bar{x}_0^\theta \\ \dot{y}_t^\theta = h_\theta(x_t^\theta), & y_0^\theta = 0. \end{cases}$$

Leur méthode repose sur un résultat de grandes déviations, qui permet de montrer la convergence uniforme sur  $\Theta$  de la log-vraisemblance vers une fonction de contraste appropriée, en probabilité. Voir aussi Dembo–Zeitouni [11] et Campillo–LeGland [9], où l'algorithme EM est étudié pour le calcul de l'estimateur du maximum de vraisemblance.

Dans ce travail, nous considérons le problème de l'estimation du paramètre inconnu  $\theta \in \Theta$ , dans le modèle

$$\begin{cases} dX_t = b_t(\theta, X) dt + \sigma_t(\theta, X) dW_t^\theta, & X_0 = x_0 \\ dY_t = h_t(\theta, X) dt + dV_t^\theta, & Y_0 = 0. \end{cases}$$

Nous utilisons une asymptotique du type “grand nombre d'observations indépendantes”. Nous définissons le processus *moyenne empirique*  $\hat{Y}^n = \{\hat{Y}_t^n, 0 \leq t \leq T\}$  par

$$\hat{Y}_t^n = \frac{1}{n} \sum_{j=1}^n Y_t^j,$$

où  $\{Y^j, 1 \leq j \leq n\}$  sont  $n$  observations indépendantes. Sous la probabilité  $\mathbf{P}_\theta$ , ce processus vérifie l'équation d'Itô suivante :

$$d\hat{Y}_t^n = \left\{ \frac{1}{n} \sum_{j=1}^n h_t(\theta, X^j) \right\} dt + \frac{1}{\sqrt{n}} d\hat{V}_t^{n,\theta}, \quad \hat{Y}_0^n = 0,$$

où pour tout  $n \geq 1$ ,  $\widehat{V}^{n,\theta} = \{\widehat{V}_t^{n,\theta}, 0 \leq t \leq T\}$  est un processus de Wiener standard. Cette approche permet de se ramener au cas d'une observation faiblement bruitée : l'asymptotique  $n \rightarrow \infty$  est remplacée par l'asymptotique  $\varepsilon = \frac{1}{\sqrt{n}} \rightarrow 0$ . Si la vraie valeur du paramètre était  $\theta$ , alors pour  $n$  assez grand on aurait

$$\widehat{Y}_t^n \simeq y_t(\theta) = \int_0^t \bar{h}_s(\theta) ds, \quad 0 \leq t \leq T,$$

avec :  $\bar{h}_t(\theta) = \mathbf{E}_\theta[h_t(\theta, X)]$  pour tout  $0 \leq t \leq T$ . L'idée consiste à définir un estimateur  $\theta_n^*$  qui minimise par rapport à  $\theta \in \Theta$  une certaine distance entre le processus moyenne empirique  $\widehat{Y}^n$  et le processus déterministe limite  $y(\theta)$ . Dans la suite de ce travail, nous considérons la distance définie par la norme dans  $L_\mu^2([0, T])$  où  $\mu$  est une mesure positive bornée sur  $[0, T]$ .

Cet estimateur est appelé *estimateur de la distance minimale (EDM)* : il s'agit d'un cas particulier d'estimateur du minimum de contraste, voir Dacunha Castelle–Dufflo [10, Section 3.2]. L'estimateur de la distance minimale a été étudié dans un cadre général par Millar [21] et [22]. Récemment, Kutoyants [18] a étudié l'estimateur de la distance minimale pour les processus de diffusion complètement observés, dans une asymptotique de type “petit bruit”.

Dans la Section 2, nous étudions l'EDM pour l'identification d'un processus partiellement observé de type diffusion. Sous l'hypothèse d'*identifiabilité* du système déterministe limite, nous montrons la consistance de l'EDM  $\theta_n^*$  avec convergence polynomiale. Sous les hypothèses de *différentiabilité* et *non-dégénérescence* nous montrons la normalité asymptotique de l'EDM  $\theta_n^*$ .

Dans la Section 3, nous étudions l'EDM pour l'identification d'un système linéaire partiellement observé. La consistance et la normalité asymptotique de l'EDM  $\theta_n^*$  découlent des résultats précédents. Dans ce cas particulier, nous montrons que la convergence est exponentielle.

Dans la Section 4, nous considérons deux exemples de système linéaire partiellement observé. La mesure positive  $\mu$  est une mesure discrète, ce qui à pour effet de ramener le problème à celui de l'identification d'un système linéaire échantillonné. A partir de simulations de l'observation échantillonnée, nous avons calculé l'estimateur  $\theta_n^*$  pour différentes valeurs de  $n$ . Nous avons comparé les deux modèles du point de vue de l'*identifiabilité* en calculant explicitement les fonctions de contraste. Nous avons aussi comparé les fonctions de contraste pour différentes valeurs de la durée d'observation  $T$ .

Dans la Conclusion, nous présentons quelques extensions possibles de ce travail : covariance du bruit d'observation dépendant de la trajectoire passée du processus d'état  $X$ , asymptotique du type "petits bruits" sur l'état et sur l'observation. La conclusion est que l'EDM permet de traiter des cas plus généraux que d'autres estimateurs, comme l'estimateur du maximum de vraisemblance, et avec des techniques beaucoup plus faciles à mettre en œuvre. Un autre avantage de l'EDM est qu'il est sans doute possible de choisir la mesure positive  $\mu$  sur  $[0, T]$  de façon à optimiser les propriétés de l'estimateur.

## 2 EDM pour les processus de type diffusion

Nous considérons  $X = \{X_t, 0 \leq t \leq T\}$  et  $Y = \{Y_t, 0 \leq t \leq T\}$  deux processus à valeurs dans  $\mathbb{R}^m$  et  $\mathbb{R}^d$  respectivement. Pour chaque valeur  $\theta \in \Theta \subset \mathbb{R}^p$ , nous considérons la probabilité  $\mathbf{P}_\theta$  sur l'espace canonique  $C([0, T]; \mathbb{R}^{m+d})$  telle que, sous  $\mathbf{P}_\theta$

$$\begin{cases} dX_t = b_t(\theta, X) dt + \sigma_t(\theta, X) dW_t^\theta, & X_0 = x_0 \\ dY_t = h_t(\theta, X) dt + dV_t^\theta, & Y_0 = 0, \end{cases} \quad (3)$$

où  $\theta \in \Theta$  est le paramètre inconnu. Nous supposons aussi que sous  $\mathbf{P}_\theta$ ,  $\{W_t^\theta, 0 \leq t \leq T\}$  et  $\{V_t^\theta, 0 \leq t \leq T\}$  sont des processus de Wiener indépendants de  $\mathbb{R}^m$  et  $\mathbb{R}^d$  respectivement.

Nous faisons l'hypothèse suivante sur les coefficients du système (3) :

**(H I)** *Pour tout  $\theta \in \Theta$ , les fonctions  $(t, x) \mapsto b_t(\theta, x)$ ,  $(t, x) \mapsto \sigma_t(\theta, x)$  et  $(t, x) \mapsto h_t(\theta, x)$  sont nonanticipatives, à valeurs dans  $\mathbb{R}^m$  et  $\mathbb{R}^d$  respectivement.*

*En outre, ces fonctions sont lipschitziennes et à croissance au plus linéaire : pour tout  $0 \leq t \leq T$  et  $x, x' \in C([0, T]; \mathbb{R}^m)$*

$$|b_t(\theta, x) - b_t(\theta, x')| + |\sigma_t(\theta, x) - \sigma_t(\theta, x')| + |h_t(\theta, x) - h_t(\theta, x')| \leq L \sup_{0 \leq s \leq t} |x_s - x'_s|,$$

$$|b_t(\theta, x)| + |\sigma_t(\theta, x)| + |h_t(\theta, x)| \leq L \left( 1 + \sup_{0 \leq s \leq t} |x_s| \right),$$

*et la constante  $L > 0$  est indépendante de  $\theta \in \Theta$ .*

Sous l'hypothèse **(H I)**, le système (3) admet une solution unique, voir Liptser–Shiryayev [20], et on peut vérifier que cette solution possède des moments de tout ordre : pour tout  $r \geq 1$ , il existe une constante  $C_r > 0$  telle que

$$\mathbf{E}_\theta \left( \sup_{0 \leq t \leq T} |X_t|^r \right) \leq C_r (1 + \mathbf{E}_\theta |X_0|^r) = C_r (1 + |x_0|^r) , \quad (4)$$

où la constante  $C_r$  ne dépend pas du paramètre inconnu  $\theta \in \Theta$ .

Dans ce chapitre, nous nous intéressons à l'estimateur du paramètre  $\theta$  dans le modèle décrit ci-dessus, par la méthode de la distance minimale. Pour l'étude asymptotique d'autres estimateurs (estimateur du maximum de vraisemblance, estimateur bayésien) dans les processus stochastiques, on consultera par exemple Basawa–Prakasa Rao [5], Ibragimov–Khasminskii [12] et Kutoyants [17].

Nous considérons ici une asymptotique de “type grand nombre d'échantillons indépendants”. A partir de  $n$  observations indépendantes  $\{Y^j, 1 \leq j \leq n\}$ , nous définissons le processus moyenne empirique  $\hat{Y}^n = \{\hat{Y}_t^n, 0 \leq t \leq T\}$  par :

$$\hat{Y}_t^n = \frac{1}{n} \sum_{j=1}^n Y_t^j .$$

Nous définissons la fonction  $\hat{h}_t^n(\theta)$  par :

$$\hat{h}_t^n(\theta) = \frac{1}{n} \sum_{j=1}^n h_t(\theta, X^j) , \quad (5)$$

où  $\{X^j, 1 \leq j \leq n\}$  sont  $n$  copies indépendantes de la solution de l'équation :

$$dX_t = b_t(\theta, X) dt + \sigma_t(\theta, X) dW_t^\theta , \quad X_0 = x_0 .$$

Sous la probabilité  $\mathbf{P}_\theta$ , ce processus vérifie l'équation d'Itô suivante :

$$d\hat{Y}_t^n = \hat{h}_t^n(\theta) dt + \frac{1}{\sqrt{n}} d\hat{V}_t^{n,\theta} , \quad \hat{Y}_0^n = 0 , \quad (6)$$

où pour tout  $n \geq 1$ ,  $\hat{V}^{n,\theta} = \{\hat{V}_t^{n,\theta}, 0 \leq t \leq T\}$  est un processus de Wiener standard, et la fonction  $\hat{h}_t^n(\theta)$  est définie par (5).

D'après la loi des grands nombres, nous avons la convergence suivante :

$$\hat{h}_t^n(\theta) \xrightarrow{n \uparrow \infty} \bar{h}_t(\theta) = \mathbf{E}_\theta[h_t(\theta, X)] , \quad \mathbf{P}_\theta\text{-p.s.}$$

Notons par  $y(\theta)$  la solution de l'équation déterministe limite de l'équation (6) quand  $n$  tend vers l'infini. Nous avons :

$$y_t(\theta) = \int_0^t \bar{h}_s(\theta) ds . \quad (7)$$

L'estimateur de la distance minimale pour le paramètre inconnu  $\theta$  dans le modèle statistique ci-dessus, est fondé sur la remarque suivante : quand la taille  $n$  de l'échantillon est grand, le processus moyenne empirique  $\hat{Y}^n$  est proche de  $y(\theta_0)$ , où  $\theta_0$  désigne la *vraie* valeur du paramètre. Nous définissons donc l'estimateur de la distance minimale comme l'ensemble des points  $\theta \in \Theta$  tels que  $\hat{Y}^n$  soit proche de  $y(\theta)$ . Pour mesurer cette distance, nous considérons une mesure positive bornée  $\mu$  sur  $[0, T]$ , et nous notons  $\|\cdot\|$  et  $\langle \cdot, \cdot \rangle$  la norme et le produit scalaire dans  $L^2_\mu([0, T])$ , c'est à dire

$$\|\phi\| = \left\{ \int_0^T |\phi_t|^2 d\mu(t) \right\}^{1/2} .$$

Si en outre  $\phi$  est continue sur  $[0, T]$ , nous avons l'estimation suivante

$$\|\phi\| \leq \sqrt{\mu_T} \sup_{0 \leq t \leq T} |\phi_t|$$

où  $\mu_T = \mu([0, T])$ .

**Définition 2.1** *L'estimateur de distance minimale (EDM) est défini par :*

$$\theta_n^* \in \operatorname{argmin}_{\theta \in \Theta} \|\hat{Y}^n - y(\theta)\| , \quad (8)$$

où  $\hat{Y}^n$  et  $y(\theta)$  sont les solutions respectives des équations (6) et (7).

## 2.1 Consistance de l'EDM

Nous faisons l'hypothèse d'*identifiabilité* suivante :

**(H II)** *Pour tout  $\nu > 0$*

$$g(\nu) \triangleq \inf_{|\theta - \theta_0| > \nu} \|y(\theta) - y(\theta_0)\| > 0 ,$$

où  $\theta_0$  désigne la *vraie* valeur du paramètre inconnu.

Nous obtenons une vitesse de convergence polynomiale de tout ordre, pour l'estimateur de distance minimale.

**Théorème 2.2** *Sous les hypothèses (H I) et (H II), pour tout  $r \geq 1$  il existe une constante  $C_r > 0$  telle que*

$$\mathbf{P}_{\theta_0}(|\theta_n^* - \theta_0| > \nu) \leq \frac{C_r}{n^r g^{2r}(\nu)} .$$

Nous commençons par établir le résultat suivant.

**Lemme 2.3**

$$\{|\theta_n^* - \theta_0| > \nu\} \subset \left\{ \|\widehat{Y}^n - y(\theta_0)\| > \frac{1}{2} g(\nu) \right\} .$$

PREUVE. Nous avons d'abord :

$$\{|\theta_n^* - \theta_0| \leq \nu\} \supset \left\{ \inf_{|\theta - \theta_0| \leq \nu} \|\widehat{Y}^n - y(\theta)\| < \inf_{|\theta - \theta_0| > \nu} \|\widehat{Y}^n - y(\theta)\| \right\} .$$

L'inégalité triangulaire nous donne :

$$\|\widehat{Y}^n - y(\theta)\| \leq \|\widehat{Y}^n - y(\theta_0)\| + \|y(\theta) - y(\theta_0)\| ,$$

d'où

$$\begin{aligned} \inf_{|\theta - \theta_0| \leq \nu} \|\widehat{Y}^n - y(\theta)\| &\leq \|\widehat{Y}^n - y(\theta_0)\| + \inf_{|\theta - \theta_0| \leq \nu} \|y(\theta) - y(\theta_0)\| \\ &= \|\widehat{Y}^n - y(\theta_0)\| , \end{aligned}$$

car la borne inférieure 0 est atteinte pour  $\theta = \theta_0$ . Nous avons aussi l'inégalité suivante :

$$\|\widehat{Y}^n - y(\theta)\| \geq \|y(\theta) - y(\theta_0)\| - \|\widehat{Y}^n - y(\theta_0)\| ,$$

d'où

$$\begin{aligned} \inf_{|\theta - \theta_0| > \nu} \|\widehat{Y}^n - y(\theta)\| &\geq \inf_{|\theta - \theta_0| > \nu} \|y(\theta) - y(\theta_0)\| - \|\widehat{Y}^n - y(\theta_0)\| \\ &= g(\nu) - \|\widehat{Y}^n - y(\theta_0)\| . \end{aligned}$$

Nous déduisons de ces inégalités l'inclusion suivante :

$$\begin{aligned} \{|\theta_n^* - \theta_0| > \nu\} &\subset \left\{ \inf_{|\theta - \theta_0| \leq \nu} \|\widehat{Y}^n - y(\theta)\| \geq \inf_{|\theta - \theta_0| > \nu} \|\widehat{Y}^n - y(\theta)\| \right\} \\ &\subset \left\{ \|\widehat{Y}^n - y(\theta_0)\| \geq \frac{1}{2}g(\nu) \right\} . \quad \square \end{aligned}$$

Considérons le processus  $\xi^n(\theta) = \{\xi_t^n(\theta), 0 \leq t \leq T\}$ , où  $\xi_t^n(\theta) = \sqrt{n} [\widehat{Y}_t^n - y_t(\theta)]$ , et  $\widehat{Y}^n$  et  $y(\theta)$  sont les solutions respectives des équations (6) et (7). Ce processus vérifie l'équation d'Itô suivante :

$$d\xi_t^n(\theta) = \sqrt{n} [\widehat{h}_t^n(\theta) - \bar{h}_t(\theta)] dt + d\widehat{V}_t^n . \quad (9)$$

où  $\{\widehat{V}_t^n, 0 \leq t \leq T\}$  est un processus de Wiener standard.

**Lemme 2.4** *Sous l'hypothèse (H I), le processus  $\xi^n(\theta_0) = \{\xi_t^n(\theta_0), 0 \leq t \leq T\}$ , solution de l'équation (9) possède sous  $\mathbf{P}_{\theta_0}$  des moments de tout ordre, bornés indépendamment de  $n$ .*

PREUVE. La solution de l'équation (9) s'écrit

$$\xi_t^n(\theta_0) = \int_0^t \sqrt{n} [\widehat{h}_s^n(\theta_0) - \bar{h}_s(\theta_0)] ds + \widehat{V}_t^n ,$$

d'où

$$\begin{aligned} \mathbf{E}_{\theta_0} |\xi_t^n(\theta_0)|^r &= \mathbf{E}_{\theta_0} \left| \left( \int_0^t \sqrt{n} [\widehat{h}_s^n(\theta_0) - \bar{h}_s(\theta_0)] ds + \widehat{V}_t^n \right) \right|^r \\ &\leq N_r \left( \mathbf{E}_{\theta_0} \left| \int_0^t \sqrt{n} [\widehat{h}_s^n(\theta_0) - \bar{h}_s(\theta_0)] ds \right|^r + \mathbf{E}_{\theta_0} |\widehat{V}_t^n|^r \right) \\ &\leq N_r \left( T^{r-1} \int_0^T \mathbf{E}_{\theta_0} \left| \sqrt{n} [\widehat{h}_s^n(\theta_0) - \bar{h}_s(\theta_0)] \right|^r ds + \mathbf{E}_{\theta_0} |\widehat{V}_t^n|^r \right) \end{aligned}$$

d'après l'inégalité de Jensen et  $N_r = 2^{r-1} > 0$  est une constante qui ne dépend que de  $r$ . Nous remarquons d'abord que :

$$\mathbf{E}_{\theta_0} \left| \sqrt{n} [\widehat{h}_s^n(\theta_0) - \bar{h}_s(\theta_0)] \right|^r = \mathbf{E}_{\theta_0} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^n \zeta_s^j(\theta_0) \right|^r ,$$



où

$$\zeta_s^j(\theta_0) = h_s(\theta_0, X^j) - \bar{h}_s(\theta_0), \quad 1 \leq j \leq n. \quad (10)$$

On vérifie immédiatement que les vecteurs aléatoires  $\{\zeta_s^j(\theta_0), 1 \leq j \leq n\}$  sont i.i.d. et d'espérance nulle sous  $\mathbf{P}_{\theta_0}$ . D'après le Lemme 3.1 du Chapitre III de [12], nous avons :

$$\mathbf{E}_{\theta_0} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^n \zeta_s^j(\theta_0) \right|^r \leq C_r' \mathbf{E}_{\theta_0} |\zeta_s^1(\theta_0)|^r.$$

Et en utilisant l'hypothèse **(H I)**, nous avons :

$$\begin{aligned} \mathbf{E}_{\theta_0} |\zeta_s^1(\theta_0)|^r &= \mathbf{E}_{\theta_0} |h_s(\theta_0, X^1) - \bar{h}_s(\theta_0)|^r \\ &\leq C_r'' \mathbf{E}_{\theta_0} |h_s(\theta_0, X^1)|^r + C_r'' |\bar{h}_s(\theta_0)|^r \\ &\leq C_r'' L (1 + \mathbf{E}_{\theta_0} \sup_{0 \leq t \leq T} |X_t^1|^r) + C_r'' |\bar{h}_s(\theta_0)|^r. \end{aligned} \quad (11)$$

Nous obtenons l'inégalité suivante :

$$\begin{aligned} \int_0^T \mathbf{E}_{\theta_0} \left| \sqrt{n} [\widehat{h}_s^n(\theta_0) - \bar{h}_s(\theta_0)] \right|^r ds &\leq C_r'' L (1 + \mathbf{E}_{\theta_0} \sup_{0 \leq t \leq T} |X_t^1|^r) \\ &\quad + C_r'' \int_0^T |\bar{h}_s(\theta_0)|^r ds < \infty, \end{aligned}$$

ce qui fournit une majoration indépendante de  $n$ .

D'autre part  $\widehat{V}^n = \{\widehat{V}_t^n, 0 \leq t \leq T\}$  est un processus de Wiener standard sous  $\mathbf{P}_{\theta_0}$ , donc il a ses moments d'ordre  $r$  finis. Nous en déduisons que  $\mathbf{E}_{\theta_0} |\xi_t^n(\theta_0)|^r$  est majoré par une constante indépendante de  $n$ .  $\square$

**PREUVE DU THÉORÈME 2.2.** Le Lemme 2.3 ci-dessus nous permet d'établir l'inégalité :

$$\begin{aligned} \mathbf{P}_{\theta_0} (|\theta_n^* - \theta_0| > \nu) &\leq \mathbf{P}_{\theta_0} \left( \|\xi^n(\theta_0)\| \geq \frac{1}{2} \sqrt{n} g(\nu) \right) \\ &\leq \frac{4^r}{n^r g^{2r}(\nu)} \mathbf{E}_{\theta_0} \|\xi^n(\theta_0)\|^{2r}, \end{aligned} \quad (12)$$

d'après l'inégalité de Markov. Ici,  $\|\cdot\|$  désigne la norme dans  $L^2_\mu([0, T])$ , et  $\xi^n(\theta_0)$  est la solution de l'équation (9). Par définition nous avons :

$$\|\xi^n(\theta_0)\|^2 = \int_0^T |\xi_t^n(\theta_0)|^2 d\mu(t) ,$$

En utilisant l'inégalité de Jensen, nous avons

$$\begin{aligned} \|\xi^n(\theta_0)\|^{2r} &= \left\{ \int_0^T |\xi_t^n(\theta_0)|^2 d\mu(t) \right\}^r \\ &\leq \mu_T^{r-1} \int_0^T |\xi_t^n(\theta_0)|^{2r} d\mu(t) . \end{aligned}$$

ous en déduisons que

$$\begin{aligned} \mathbf{E}_{\theta_0} \|\xi^n(\theta_0)\|^{2r} &\leq \mu_T^{r-1} \int_0^T \mathbf{E}_{\theta_0} |\xi_t^n(\theta_0)|^{2r} d\mu(t) \\ &\leq \mu_T^r \sup_{0 \leq t \leq T} \mathbf{E}_{\theta_0} |\xi_t^n(\theta_0)|^{2r} . \end{aligned}$$

D'après le Lemme 2.4, nous pouvons trouver une constante  $C_r > 0$  qui ne dépend pas de  $n$ , telle que :

$$\mathbf{E}_{\theta_0} \|\xi^n(\theta_0)\|^{2r} \leq C_r . \quad \square$$

## 2.2 Normalité asymptotique de la moyenne empirique

Considérons le processus  $\xi^n(\theta) = \{\xi_t^n(\theta), 0 \leq t \leq T\}$  solution de l'équation (9). Nous avons le résultat de convergence faible suivant.

**Proposition 2.5** *Sous l'hypothèse (H I) :*

$$\xi^n(\theta_0) \Rightarrow \xi(\theta_0) ,$$

dans  $C([0, T]; \mathbb{R}^d)$ , quand  $n \rightarrow \infty$ , où  $\xi(\theta_0) = \{\xi_t(\theta_0), 0 \leq t \leq T\}$  est un processus gaussien centré.

Pour démontrer ce résultat, nous utilisons la démarche classique suivante : nous commençons par montrer que la suite  $\{\xi^n(\theta), n \geq 1\}$  est tendue. Pour

cela nous utilisons le Théorème 12.3 et l'inégalité (12.5) dans Billingsley [7]. Ensuite nous caractérisons de façon unique la limite de toute sous-suite de la suite  $\{\xi^n(\theta_0), n \geq 1\}$ . Pour cela, nous montrons que le processus limite est gaussien, et nous explicitons sa fonction de corrélation  $\Xi(\theta_0) = \{\Xi_{s,t}(\theta_0), 0 \leq s, t \leq T\}$ .

Nous commençons par établir la proposition suivante, qui montre que la suite  $\{\xi^n(\theta_0), n \geq 1\}$  est équitendue sur  $C([0, T]; \mathbb{R}^d)$ .

**Proposition 2.6** *Sous l'hypothèse (H I)*

$$\mathbf{E}_{\theta_0} |\xi_t^n(\theta_0) - \xi_s^n(\theta_0)|^4 < C |t - s|^2, \quad 0 \leq s, t \leq T, \quad C > 0,$$

PREUVE. Nous avons :

$$\begin{aligned} \mathbf{E}_{\theta_0} |\xi_t^n(\theta_0) - \xi_s^n(\theta_0)|^4 &= \mathbf{E}_{\theta_0} \left| \int_s^t \sqrt{n} [\hat{h}_r^n(\theta_0) - \bar{h}_r(\theta_0)] dr + \hat{V}_t^n - \hat{V}_s^n \right|^4 \\ &\leq 8 \mathbf{E}_{\theta_0} |\hat{V}_t^n - \hat{V}_s^n|^4 + 8 \mathbf{E}_{\theta_0} \left| \int_s^t \sqrt{n} [\hat{h}_r^n(\theta_0) - \bar{h}_r(\theta_0)] dr \right|^4 \\ &\leq C_1 |t - s|^2 + 8 \mathbf{E}_{\theta_0} \left| \int_s^t \frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta_r^i(\theta_0) dr \right|^4, \end{aligned}$$

où  $\zeta_r^i(\theta_0)$  est définie par (10). En utilisant l'inégalité de Jensen, nous avons :

$$\mathbf{E}_{\theta_0} |\xi_t^n(\theta_0) - \xi_s^n(\theta_0)|^4 \leq C_1 |t - s|^2 + 8 |t - s|^3 \int_s^t \mathbf{E}_{\theta_0} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta_r^i(\theta_0) \right|^4 dr.$$

Comme dans la preuve du Lemme 2.4 ci-dessus, il existe une constante  $C_4'' > 0$  indépendante de  $n$ , tels que

$$\mathbf{E}_{\theta_0} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta_r^i(\theta_0) \right|^4 \leq C_4'' L (1 + \mathbf{E}_{\theta_0} \sup_{0 \leq t \leq T} |X_t^1|^4) + C_4'' |\bar{h}_r(\theta_0)|^4 < \infty,$$

et nous avons :

$$\mathbf{E}_{\theta_0} |\xi_t^n(\theta_0) - \xi_s^n(\theta_0)|^4 \leq C |t - s|^2,$$

où

$$C = C_1 + 8TC_4'' \left( LT (1 + \mathbf{E}_{\theta_0} \sup_{0 \leq t \leq T} |X_t^1|^4) + \int_0^T |\bar{h}_s(\theta_0)|^4 ds \right) < \infty. \quad \square$$

D'après le théorème de Prohorov dans [7], toute suite extraite de  $\{\xi^n(\theta_0), n \geq 1\}$  contient une sous-suite qui converge en loi dans  $C([0, T]; \mathbb{R}^d)$ . Soit  $\xi(\theta_0)$  la limite de cette sous-suite. Le but de la proposition suivante est de caractériser cette limite de façon unique.

**Proposition 2.7** *Soit  $\xi(\theta_0)$  la limite d'une sous-suite de  $\{\xi^n(\theta_0), n \geq 1\}$ . Le processus  $\xi(\theta_0) = \{\xi_t(\theta_0), 0 \leq t \leq T\}$  est gaussien centré, et la fonction de corrélation  $\{\Xi_{s,t}(\theta_0), 0 \leq s, t \leq T\}$  est définie par*

$$\Xi_{s,t}(\theta_0) = \mathbf{E}_{\theta_0}[\xi_s(\theta_0)\xi_t^*(\theta_0)] = \int_0^t \int_0^s \mathbf{E}_{\theta_0}[\zeta_u(\theta_0)\zeta_v^*(\theta_0)] du dv + (t \wedge s), \quad (13)$$

avec  $\zeta_t(\theta_0) = h_t(\theta_0, X) - \bar{h}_t(\theta_0)$ .

PREUVE. Soit  $0 \leq t_0 < \dots < t_l \leq T$  une suite croissante quelconque dans  $[0, T]$ , et  $\lambda_1, \dots, \lambda_l$  des scalaires quelconques. Nous définissons le vecteur aléatoire

$$\Upsilon_n = \sum_{k=1}^l \lambda_k \xi_{t_k}^n(\theta_0).$$

Par définition de  $\xi^n(\theta_0)$ , nous avons

$$\Upsilon_n = \sum_{k=1}^l \lambda_k \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{t_k} \zeta_t^i(\theta_0) dt + \widehat{V}_{t_k}^n \right\} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_i + v_n = u_n + v_n,$$

avec

$$\eta_i = \sum_{k=1}^l \lambda_k \int_0^{t_k} \zeta_t^i(\theta_0) dt, \quad v_n = \sum_{k=1}^l \lambda_k \widehat{V}_{t_k}^n.$$

Les vecteurs aléatoires  $\{\eta_i, i \geq 1\}$  sont i.i.d de moyenne nulle et de covariance finie : en effet

$$\begin{aligned} \Gamma = \mathbf{E}_{\theta_0}[\eta_i \eta_i^*] &= \sum_{k,k'=1}^l \lambda_k \lambda_{k'} \int_0^{t_k} \int_0^{t_{k'}} \mathbf{E}_{\theta_0}[\zeta_u(\theta_0)\zeta_v^*(\theta_0)] du dv \\ &= \sum_{k,k'=1}^l \lambda_k \lambda_{k'} \Gamma_{t_k, t_{k'}}(\theta_0), \end{aligned}$$

d'où

$$\|\Gamma\| \leq C \sup_{0 \leq t \leq T} \mathbf{E}_{\theta_0} |\zeta_t(\theta_0)|^2 < \infty,$$

d'après (11). D'après le théorème de la limite centrale, la suite  $\{u_n, n \geq 1\}$  converge en loi sous  $\mathbf{P}_{\theta_0}$  vers un vecteur aléatoire gaussien centré et de covariance  $\Gamma$ .

D'autre part, les vecteurs aléatoires  $\{v_n, n \geq 1\}$  ont tous la même loi gaussienne centrée et de covariance

$$\Lambda = \mathbf{E}_{\theta_0}(v_n v_n^*) = \sum_{k,k'=1}^l \lambda_k \lambda_{k'} t_k \wedge t_{k'} .$$

Compte tenu que pour tout  $n \leq 1$ , les vecteurs aléatoires  $u_n$  et  $v_n$  sont indépendants, nous en déduisons que la suite  $\{\Upsilon_n, n \geq 1\}$  converge en loi sous  $\mathbf{P}_{\theta_0}$  vers un vecteur aléatoire gaussien  $\Upsilon$  centré et de covariance  $\Gamma + \Lambda$ .

Par unicité de la limite, nous en déduisons que

$$\Upsilon = \sum_{k=1}^l \lambda_k \xi_{t_k}(\theta_0)$$

ce qui montre le résultat.  $\square$

### 2.3 Normalité asymptotique de l'EDM

Nous faisons les hypothèses de *différentiabilité* et de *non-dégénérescence* suivantes :

**(H III)** *La fonction  $y(\theta)$  est différentiable par rapport à  $\theta$  au point  $\theta_0$ , c'est-à-dire qu'il existe une fonction  $\dot{y}(\theta_0) \in L^2_{\mu}([0, T]; \mathbb{R}^{d \times p})$  telle que :*

$$\|y(\theta) - y(\theta_0) - \dot{y}(\theta_0)(\theta - \theta_0)\| = |\theta - \theta_0| \varepsilon(|\theta - \theta_0|) , \quad (14)$$

où  $\varepsilon(r) \rightarrow 0$  quand  $r \rightarrow 0$ .

**(H IV)** *La matrice  $p \times p$  définie par :*

$$J(\theta_0) = \int_0^T \dot{y}_t^*(\theta_0) \dot{y}_t(\theta_0) d\mu(t) ,$$

*est non-singulière, c'est-à-dire qu'il existe une constante  $c_0 > 0$  telle que : pour tout  $u \in \mathbb{R}^p$*

$$\|\dot{y}(\theta_0) u\| = [u^* J(\theta_0) u]^{1/2} \geq c_0 |u| . \quad (15)$$

Nous définissons le vecteur aléatoire

$$\widehat{u}(\theta_0) \triangleq J(\theta_0)^{-1} m(\theta_0), \quad \text{où: } m(\theta_0) \triangleq \int_0^T \dot{y}_t^*(\theta_0) \xi_t(\theta_0) d\mu(t).$$

$\xi(\theta_0)$  est un processus gaussien, de fonction de corrélation  $\Xi(\theta_0) = \{\Xi_{s,t}(\theta_0), 0 \leq s, t \leq T\}$  définie par (13), donc  $\widehat{u}(\theta_0)$  est un vecteur aléatoire gaussien centré et de matrice de covariance

$$I(\theta_0) = J(\theta_0)^{-1} \left[ \int_0^T \int_0^T \dot{y}_t^*(\theta_0) \Xi_{s,t}(\theta_0) \dot{y}_s(\theta_0) d\mu(t) d\mu(s) \right] J(\theta_0)^{-1}. \quad (16)$$

**Théorème 2.8** *Sous les hypothèses (H II), (H III) et (H IV)*

$$\sqrt{n} (\theta_n^* - \theta_0) \Rightarrow \widehat{u}(\theta_0), \quad \text{quand } n \rightarrow \infty,$$

où  $\widehat{u}(\theta_0)$  est un vecteur aléatoire gaussien centré, de matrice de covariance  $I(\theta_0)$  définie par (16).

**PREUVE DU THÉORÈME 2.8.** Nous commençons par vérifier que les hypothèses (2.2), (2.4), (2.5) et (2.6) du Théorème 2.9 de [21, Section X] sont vérifiées. L'hypothèse d'identifiabilité (2.2) découle de notre hypothèse (H II). L'hypothèse de convergence (2.4) résulte de la Proposition 2.5 ci-dessus. Nos hypothèses (H III) et (H IV) entraînent respectivement l'hypothèse de différentiabilité (2.5) et l'hypothèse de non-singularité (2.6).

Nous introduisons alors les notations suivantes. L'hypothèse de différentiabilité assure l'existence du vecteur gradient  $\dot{y}(\theta_0) = (\dot{y}^1(\theta_0), \dots, \dot{y}^p(\theta_0))$  avec

$$\dot{y}^i(\theta_0) = \frac{\partial}{\partial \theta^i} y(\theta_0) \quad i = 1, \dots, p.$$

L'hypothèse de non-singularité signifie que les vecteurs  $\{\dot{y}^1(\theta_0), \dots, \dot{y}^p(\theta_0)\}$  sont linéairement indépendants dans  $L_\mu^2([0, T]; \mathbb{R}^d)$ . Nous considérons le sous-espace  $B \subset L_\mu^2([0, T]; \mathbb{R}^d)$  engendré par ces vecteurs, et nous notons  $\pi$  la projection de  $L_\mu^2([0, T]; \mathbb{R}^d)$  sur  $B$ .

D'après le Théorème 2.9 de [21, Section X], avec une probabilité qui approche 1 quand  $n \rightarrow \infty$ , l'EDM  $\theta_n^*$  existe et est unique, et nous avons

$$\xi^n(\theta_n^*) - \xi^n(\theta_0) \Rightarrow \pi \xi(\theta_0),$$

dans  $L^2_\mu([0, T]; \mathbb{R}^d)$  et

$$\sqrt{n}(\theta_n^* - \theta_0) \Rightarrow \Upsilon ,$$

dans  $\mathbb{R}^p$ , où le vecteur aléatoire  $\Upsilon$  est défini par :

$$\pi \xi(\theta_0) = \dot{y}(\theta_0) \Upsilon .$$

En prenant le produit scalaire dans  $L^2_\mu([0, T])$  avec  $\dot{y}(\theta_0)$ , nous obtenons :

$$\begin{aligned} m(\theta_0) &= \int_0^T \dot{y}_t^*(\theta_0) \xi_t(\theta_0) d\mu(t) = \int_0^T \dot{y}_t^*(\theta_0) \pi \xi_t(\theta_0) d\mu(t) \\ &= \int_0^T \dot{y}_t^*(\theta_0) \dot{y}_t(\theta_0) d\mu(t) \Upsilon = J(\theta_0) \Upsilon . \end{aligned}$$

Nous en déduisons :

$$\Upsilon = J^{-1}(\theta_0) m(\theta_0) = \hat{u}(\theta_0) . \quad \square$$

### 3 EDM pour les diffusions linéaires

Nous considérons les processus  $X = \{X_t, 0 \leq t \leq T\}$  et  $Y = \{Y_t, 0 \leq t \leq T\}$  à valeurs dans  $\mathbb{R}^m$  et  $\mathbb{R}^d$  respectivement. Nous considérons, pour chaque valeur de  $\theta \in \Theta \subset \mathbb{R}^p$ , la probabilité  $\mathbf{P}_\theta$  définie sur l'espace canonique  $C([0, T]; \mathbb{R}^{m+d})$  telle que, sous  $\mathbf{P}_\theta$

$$\begin{cases} dX_t = B_t(\theta) X_t dt + G_t dW_t^\theta, & X_0 = x_0 \\ dY_t = H_t(\theta) X_t dt + dV_t^\theta, & Y_0 = 0 , \end{cases} \quad (17)$$

où  $\theta \in \Theta$  est le paramètre inconnu. Nous supposons aussi que sous  $\mathbf{P}_\theta$ ,  $\{W_t^\theta, 0 \leq t \leq T\}$  et  $\{V_t^\theta, 0 \leq t \leq T\}$  sont des processus de Wiener indépendants à valeurs dans  $\mathbb{R}^m$  et  $\mathbb{R}^d$  respectivement.

Nous faisons l'hypothèse suivante sur les coefficients du système (17) :

**(H I)** *Pour tout  $\theta \in \Theta$ , les fonctions  $t \mapsto B_t(\theta)$ ,  $t \mapsto G_t$  et  $t \mapsto H_t(\theta)$  sont bornées : pour tout  $0 \leq t \leq T$*

$$\|B_t(\theta)\| + \|G_t\| + \|H_t(\theta)\| \leq L ,$$

*et la constante  $L > 0$  est indépendante de  $\theta \in \Theta$ .*

Sous cette hypothèse, le système (17) admet une solution unique qui possède des moments de tout ordre.

A partir de  $n$  observations indépendantes  $\{Y^j, 1 \leq j \leq n\}$ , nous définissons comme dans la section précédente le processus *moyenne empirique* par :

$$\widehat{Y}_t^n = \frac{1}{n} \sum_{j=1}^n Y_t^j .$$

Ce processus vérifie l'équation d'Itô suivante :

$$d\widehat{Y}_t^n = H_t(\theta) \widehat{X}_t^n dt + \frac{1}{\sqrt{n}} d\widehat{V}_t^n, \quad \widehat{Y}_0^n = 0, \quad (18)$$

où  $\widehat{V}^n = \{\widehat{V}_t^n, 0 \leq t \leq T\}$  est un processus de Wiener standard sous  $\mathbf{P}_\theta$ . Le processus  $\widehat{X}^n = \{\widehat{X}_t^n, 0 \leq t \leq T\}$  est solution de l'équation suivante :

$$d\widehat{X}_t^n = B_t(\theta) \widehat{X}_t^n dt + \frac{1}{\sqrt{n}} G_t d\widehat{W}_t^n, \quad \widehat{X}_0^n = x_0, \quad (19)$$

où  $\widehat{W}^n = \{\widehat{W}_t^n, 0 \leq t \leq T\}$  est un processus de Wiener standard sous  $\mathbf{P}_\theta$ , indépendant de  $\widehat{V}^n$ .

Quand la taille  $n$  de l'échantillon tend vers l'infini, nous obtenons les équations déterministes limites suivantes :

$$\left\{ \begin{array}{l} \frac{d}{dt} x_t(\theta) = B_t(\theta) x_t(\theta), \quad x_0(\theta) = x_0 \\ \frac{d}{dt} y_t(\theta) = H_t(\theta) x_t(\theta) \end{array} \right.$$

Nous notons par  $\Phi(\theta) = \{\Phi_{t,s}(\theta), 0 \leq s \leq t\}$  la solution fondamentale de l'équation (3), de sorte que :

$$\begin{aligned} x_t(\theta) &= \Phi_{t,s}(\theta) x_s(\theta) \\ y_t(\theta) &= \int_0^t H_s(\theta) x_s(\theta) ds = \int_0^t H_s(\theta) \Phi_{s,0}(\theta) ds x_0 \end{aligned} \quad (20)$$



Comme dans la section précédente, nous considérons une mesure positive bornée  $\mu$  sur  $[0, T]$  et nous notons  $\|\cdot\|$  et  $\langle \cdot, \cdot \rangle$  la norme et le produit scalaire dans  $L^2_\mu([0, T]; \mathbb{R}^d)$ .

**Définition 3.1** *L'estimateur de la distance minimale (EDM) est défini par:*

$$\theta_n^* \in \operatorname{argmin}_{\theta \in \Theta} \|\widehat{Y}^n - y(\theta)\| ,$$

où  $\widehat{Y}^n$  et  $y(\theta)$  sont les solutions respectives des équations (18) et (3).

**Remarque 3.2** La fonction de contraste  $K(\theta, \theta_0) = \|y(\theta) - y(\theta_0)\|$  ne dépend pas de la matrice  $G_t$ . L'EDM ne permet donc pas d'identifier la matrice  $G_t$ .

### 3.1 Applications des résultats précédents

Nous faisons l'hypothèse d'*identifiabilité* suivante :

**(H II')** *Pour tout  $\nu > 0$*

$$g(\nu) \triangleq \inf_{|\theta - \theta_0| > \nu} \|y(\theta) - y(\theta_0)\| > 0 ,$$

où  $\theta_0$  désigne la vraie valeur du paramètre inconnu.

Dans le cas linéaire, nous obtenons une vitesse de convergence exponentielle pour l'EDM.

**Théorème 3.3** *Sous les hypothèses (H I') et (H II'), il existe  $\alpha > 0$  tel que :*

$$\mathbf{P}_{\theta_0}(|\theta_n^* - \theta_0| > \nu) \leq C \exp \left\{ -\alpha \frac{n g^2(\nu)}{4} \right\} ,$$

où  $C > 0$  est une constante qui est indépendante de  $\theta \in \Theta$ .

Nous considérons d'abord le processus  $\xi^n(\theta) = \{\xi_t^n(\theta), 0 \leq t \leq T\}$ , où  $\xi_t^n(\theta) = \sqrt{n}[\widehat{Y}_t^n - y_t(\theta)]$  et  $\widehat{Y}^n, y(\theta)$  sont les solutions respectives des équations (18) et (3). Sous la probabilité  $\mathbf{P}_\theta$ , le processus  $\xi^n(\theta)$  vérifie l'équation d'Itô suivante :

$$d\xi_t^n(\theta) = \sqrt{n} H_t(\theta) [\widehat{X}_t^n - x_t(\theta)] dt + d\widehat{V}_t^n, \quad \xi_0^n(\theta) = 0 , \quad (21)$$

où  $\widehat{V}^n = \{\widehat{V}_t^n, 0 \leq t \leq T\}$  est un processus de Wiener standard. Nous pouvons aussi écrire ce processus sous forme intégrale :

$$\xi_t^n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t H_s(\theta) [X_s^i - x_s(\theta)] ds + \widehat{V}_t^n$$

et nous avons le résultat suivant :

**Proposition 3.4** *Sous l'hypothèse (H I'), le processus  $\xi^n(\theta_0)$  est sous  $P_{\theta_0}$  un processus gaussien centré et de fonction de corrélation  $\Xi(\theta_0) = \{\Xi_{s,t}(\theta_0), 0 \leq s, t \leq T\}$  indépendante de  $n$  :*

$$\Xi_{s,t}(\theta_0) = \int_0^t \int_0^s H_u(\theta_0) \left( \int_0^{u \wedge v} \Phi_{u,w}(\theta_0) \Phi_{v,w}^*(\theta_0) dw \right) H_v^*(\theta_0) du dv + s \wedge t . \quad (22)$$

PREUVE DU THÉORÈME 3.3. En utilisant le Lemme 2.3, nous avons l'inégalité :

$$\mathbf{P}_{\theta_0}(|\theta_n^* - \theta_0| > \nu) \leq \mathbf{P}_{\theta_0} \left( \exp \{ \alpha \|\xi^n(\theta_0)\|^2 \} > \exp \left\{ \alpha \frac{n g^2(\nu)}{4} \right\} \right) .$$

D'après l'inégalité de Bienaimé Chebyshev, nous avons :

$$\mathbf{P}_{\theta_0}(|\theta_n^* - \theta_0| > \nu) \leq \exp \left\{ -\alpha \frac{n g^2(\nu)}{4} \right\} \mathbf{E}_{\theta_0} \left[ \exp \{ \alpha \|\xi^n(\theta_0)\|^2 \} \right] .$$

D'après la Proposition 3.4, on peut toujours trouver un  $\alpha > 0$  assez petit, qui ne dépend pas de  $n$ , tel que  $\mathbf{E}_{\theta_0}[\exp \{ \alpha \|\xi(\theta_0)\|^2 \}]$  soit finie.  $\square$

Nous faisons l'hypothèse de régularité suivante :

(H III') *Les fonctions  $B(\theta)$  et  $H(\theta)$  sont différentiables par rapport à  $\theta$  au point  $\theta_0$ , c'est-à-dire qu'il existe des fonctions  $\dot{B}(\theta_0) \in L_\mu^2([0, T]; \mathbb{R}^{m \times m \times p})$  et  $\dot{H}(\theta_0) \in L_\mu^2([0, T]; \mathbb{R}^{d \times m \times p})$  telles que :*

$$\|B(\theta) - B(\theta_0) - \dot{B}(\theta_0)(\theta - \theta_0)\| = |\theta - \theta_0| \varepsilon(|\theta - \theta_0|) ,$$

$$\|H(\theta) - H(\theta_0) - \dot{H}(\theta_0)(\theta - \theta_0)\| = |\theta - \theta_0| \varepsilon(|\theta - \theta_0|) .$$

où  $\varepsilon(r) \rightarrow 0$  quand  $r \rightarrow 0$ .

Sous cette hypothèse, la fonction  $y(\theta)$  est différentiable par rapport à  $\theta$  au point  $\theta_0$ , et le vecteur gradient  $\dot{y}(\theta_0) \in L^2_\mu([0, T]; \mathbb{R}^{d \times m \times p})$  est défini par :

$$\begin{aligned} \dot{y}_t(\theta) &= \int_0^t [\dot{H}_s(\theta) x_s(\theta) + H_s(\theta) \dot{x}_s(\theta)] ds \\ \frac{d}{dt} \dot{x}_t(\theta) &= B_t(\theta) \dot{x}_t(\theta) + \dot{B}_t(\theta) x_t(\theta), \quad \dot{x}_0(\theta) = 0, \end{aligned}$$

d'où

$$\dot{x}_t(\theta) = \int_0^t \Phi_{t,s}(\theta) \dot{B}_s(\theta) x_s(\theta) ds = \int_0^t \Phi_{t,s}(\theta) \dot{B}_s(\theta) \Phi_{s,0}(\theta) ds x_0,$$

et

$$\dot{y}_t(\theta) = \int_0^t [\dot{H}_s(\theta) \Phi_{s,0}(\theta) + H_s(\theta) \int_0^s \Phi_{s,u}(\theta) \dot{B}_u(\theta) \Phi_{u,0}(\theta) du] ds x_0. \quad (23)$$

Nous faisons aussi l'hypothèse de non-dégénérescence suivante :

**(H IV')** La matrice  $p \times p$  définie par :

$$J(\theta_0) = \int_0^T \dot{y}_t^*(\theta_0) \dot{y}_t(\theta_0) d\mu(t),$$

est non singulière, c'est-à-dire qu'il existe une constante  $c_0 > 0$  telle que : pour tout  $u \in \mathbb{R}^p$

$$\|\dot{y}(\theta_0) u\| = [u^* J(\theta_0) u]^{1/2} \geq c_0 |u|,$$

où la fonction  $\dot{y}(\theta_0)$  est définie par (23).

Nous définissons le vecteur aléatoire

$$\hat{u}(\theta_0) \triangleq J(\theta_0)^{-1} m(\theta_0), \quad \text{où : } m(\theta_0) \triangleq \int_0^T \dot{y}_t^*(\theta_0) \xi_t(\theta_0) d\mu(t).$$

$\xi(\theta_0)$  est un processus gaussien centré, de fonction de corrélation  $\Xi(\theta_0) = \{\Xi_{s,t}(\theta_0), 0 \leq s, t \leq T\}$  définie par (22), donc  $\hat{u}(\theta_0)$  est un vecteur aléatoire gaussien centré, de matrice de covariance :

$$I(\theta_0) = J(\theta_0)^{-1} \left[ \int_0^T \int_0^T \dot{y}_t^*(\theta_0) \Xi_{s,t}(\theta_0) \dot{y}_s(\theta_0) d\mu(t) d\mu(s) \right] J(\theta_0)^{-1}. \quad (24)$$

**Théorème 3.5** *Sous les hypothèses (H II), (H III') et (H IV')*

$$\sqrt{n}(\theta_n^* - \theta_0) \Rightarrow \hat{u}(\theta_0), \quad \text{quand } n \rightarrow \infty ,$$

où  $\hat{u}(\theta_0)$  est un vecteur aléatoire gaussien centré, de matrice de covariance  $I(\theta_0)$  définie par (24).

La preuve est identique à celle du Théorème 2.8.

## 4 Applications numériques

Dans cette section, nous étudions deux exemples de systèmes linéaires partiellement observés. Nous faisons le choix suivant pour la mesure positive  $\mu$  sur  $[0, T]$  : nous introduisons une partition  $0 \leq t_0 < \dots < t_i < \dots < t_I = T$  de l'intervalle  $[0, T]$  et nous définissons

$$\mu = \sum_{i=0}^I \delta_{t_i}$$

ce qui a pour effet de ramener le problème à celui de l'identification du système linéaire échantillonné avec un pas de constant  $\delta = t_{i+1} - t_i$ . En effet, le processus de contraste s'écrit

$$\|\hat{Y}^n - y(\theta)\|^2 = \sum_{i=0}^I |\hat{Y}_{t_i}^n - y_{t_i}(\theta)|^2$$

et la fonction de contraste s'écrit

$$K(\theta, \theta_0) = \sum_{i=0}^I |y_{t_i}(\theta) - y_{t_i}(\theta_0)|^2 .$$

Pour l'application numérique, il suffit de simuler les observations et de calculer la sortie  $y(\theta)$  du système déterministe limite aux instants discrets  $0 \leq t_0 < \dots < t_i < \dots < t_I = T$ . Dans la pratique cela revient à simuler la sortie du système linéaire échantillonné, voir Åström [1].

## 4.1 Modèle I

Nous considérons le problème de diffusion linéaire suivant :

$$\begin{cases} dX_t = -\theta X_t dt + dW_t^\theta, & X_0 = x_0 \\ dY_t = \theta X_t dt + dV_t^\theta, & Y_0 = 0, \end{cases} \quad (25)$$

où  $\theta \in \Theta \subset \mathbb{R}$  est inconnu. Pour ce modèle, la fonction  $\bar{h}_t(\theta)$  est donnée par :

$$\bar{h}_t(\theta) = \theta \mathbf{E}_\theta[X_t] = x_0 \theta e^{-\theta t},$$

et le processus déterministe limite par :

$$y_t(\theta) = \int_0^t \bar{h}_s(\theta) ds = x_0 (1 - e^{-\theta t}).$$

Nous remarquons que pour tout  $\theta > 0$ , la limite

$$y^*(\theta) = \lim_{t \rightarrow \infty} y_t(\theta) = x_0,$$

est indépendante de  $\theta$ .

### Identifiabilité du système

Nous représentons sur la Figure 1 le comportement de la fonction de contraste  $\theta \mapsto K_T(\theta, \theta_0) = \|y(\theta) - y(\theta_0)\|$  au voisinage de  $\theta_0$ , la vraie valeur du paramètre, et pour différentes valeurs de  $T$ . La norme  $\|\cdot\|$  désigne la norme dans  $L_\mu^2([0, T]; \mathbb{R}^d)$ . Nous remarquons que pour  $T$  suffisamment grand, la fonction de contraste ne dépend pas de  $T$ . Une observation plus longue du système n'améliore donc pas l'identification. Pour les simulations numériques, nous avons utilisé les valeurs suivantes :  $\theta_0 = 2$ ,  $x_0 = 1$ ,  $T = 0.5s, 1s, 2s, 4s, 20s$ , et  $\theta \in \Theta = [1, 4]$ .

La Figure 2 représente les trajectoires du processus déterministe limite  $y(\theta)$  pour deux valeurs différentes de  $\theta$  avec  $x_0 = 1$ ,  $T = 10s$ , et  $\theta_1 = 1$ ,  $\theta_2 = 2$ .

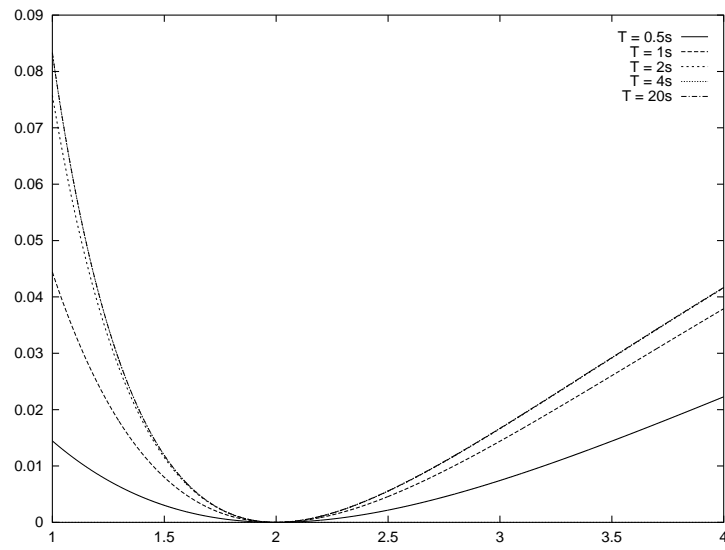


FIG. 1 - Fonction de contraste du Modèle I.

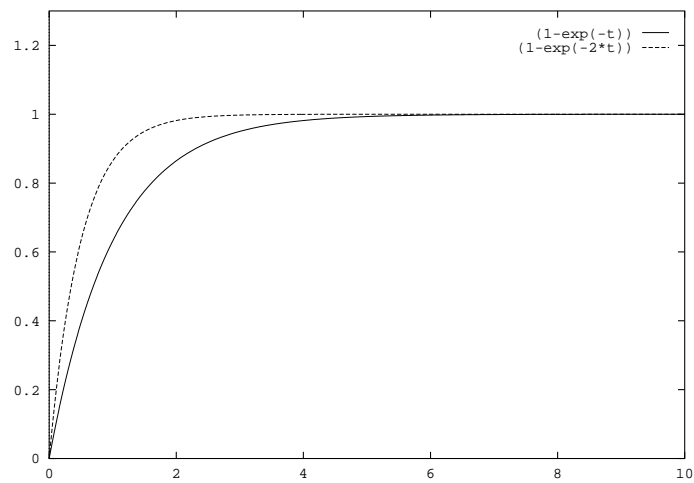


FIG. 2 - Processus déterministe limite pour  $\theta_1 = 1$ ,  $\theta_2 = 2$ .

Nous représentons sur la Figure 3 l'EDM en fonction de la taille  $n$  de l'échantillon, ainsi que les fonctions  $\theta_0 + 3\sqrt{\frac{I(\theta_0)}{n}}$  et  $\theta_0 - 3\sqrt{\frac{I(\theta_0)}{n}}$ , où  $\theta_0$  est la vraie valeur du paramètre,  $I(\theta_0)$  est la covariance définie par (24).

Pour les simulations numériques, nous avons utilisé les valeurs suivantes :

- Taille de l'échantillon :  $10^5, 5 \times 10^5, 10^6, \dots, 10^7$ .
- Vraie valeur du paramètre  $\theta_0 = 2$ .
- Ensemble des paramètres :  $\Theta = [1, 10]$ .
- Temps d'observation :  $T = 6s$ .
- Pas de discrétisation en temps :  $5 \times 10^{-3}$ .

Pour ces valeurs des paramètres, nous avons :  $I(\theta_0) = 56.99960208$ .

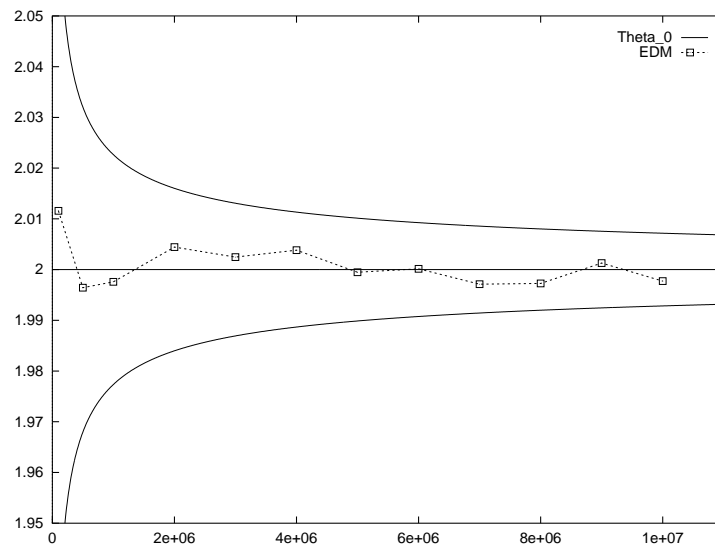


FIG. 3 - *Estimateur de la distance minimale (EDM).*

## 4.2 Modèle II

Dans cet exemple nous considérons la même équation d'état que dans le modèle précédent, en revanche l'équation d'observation ne dépend pas du paramètre inconnu. Ce modèle est décrit par l'équation suivante :

$$\begin{cases} dX_t = -\theta X_t dt + dW_t^\theta, & X_0 = x_0 \\ dY_t = X_t dt + dV_t^\theta, & Y_0 = 0, \end{cases} \quad (26)$$

où  $\theta \in \Theta \subset \mathbb{R}$  est inconnu. Pour ce deuxième modèle, l'expression de la fonction  $\bar{h}_t(\theta)$  est

$$\bar{h}_t(\theta) = \mathbf{E}_\theta[X_t] = x_0 e^{-\theta t}.$$

Le processus déterministe limite est donné par :

$$y_t(\theta) = \int_0^t \bar{h}_s(\theta) ds = x_0 \frac{1 - e^{-\theta t}}{\theta}.$$

Nous remarquons que pour tout  $\theta > 0$ , la limite

$$y^*(\theta) = \lim_{t \rightarrow \infty} y_t(\theta) = \frac{x_0}{\theta},$$

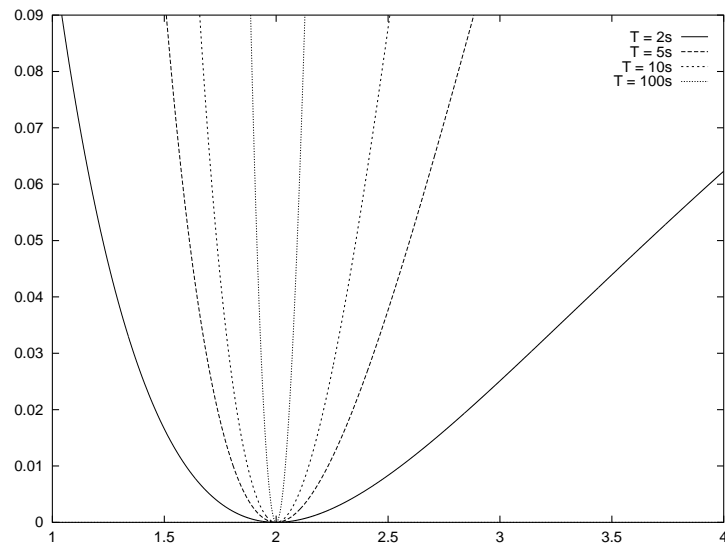
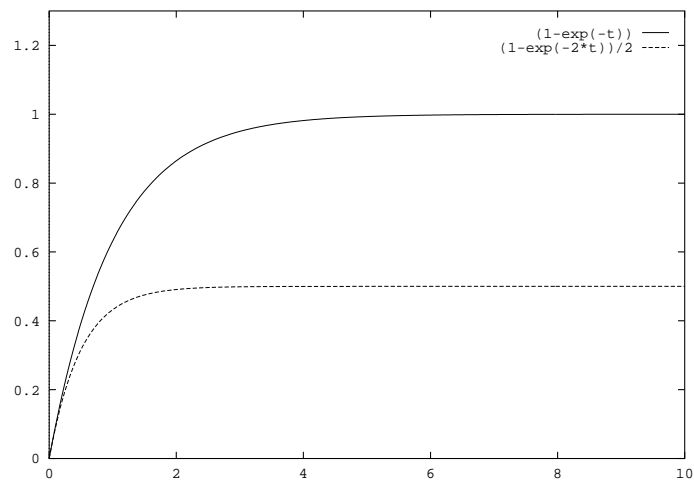
dépend du paramètre  $\theta$ .

### Identifiabilité du système

Nous représentons sur le Figure 4 le comportement de la fonction de contraste  $\theta \mapsto K_T(\theta, \theta_0) = \|y(\theta) - y(\theta_0)\|$  au voisinage de  $\theta_0$ , la vraie valeur du paramètre, et pour différentes valeurs de  $T$ . Nous remarquons que plus  $T$  est grand, plus la fonction de contraste est "contrastée". Une observation plus longue du système améliore donc l'identification. Pour les simulations numérique, nous avons utilisé les valeurs numériques suivantes :  $\theta_0 = 2$ ,  $x_0 = 1$ ,  $T = 2s, 5s, 10s, 100$ , et  $\theta \in \Theta = [1, 4]$ .

La Figure 5 représente les trajectoires du processus déterministe limite pour deux valeurs différentes de  $\theta$  avec  $x_0 = 1$ ,  $T = 10s$ , et  $\theta_1 = 1$ ,  $\theta_2 = 2$ .



FIG. 4 - *Fonction de contraste du Modèle II.*FIG. 5 - *Processus déterministe limite pour  $\theta_1 = 1$ ,  $\theta_2 = 2$*

Nous représentons sur la Figure 6 l'EDM en fonction de la taille  $n$  de l'échantillon, ainsi que les fonctions  $\theta_0 + 3\sqrt{\frac{I(\theta_0)}{n}}$  et  $\theta_0 - 3\sqrt{\frac{I(\theta_0)}{n}}$ , où  $\theta_0$  est la vraie valeur du paramètre,  $I(\theta_0)$  est la covariance définie par (24).

Pour les simulations numériques, nous avons utilisé les valeurs suivantes :

- Taille de l'échantillon :  $10^5, 5 \times 10^5, 10^6, \dots, 10^7$ .
- Vraie valeur du paramètre  $\theta_0 = 2$ .
- Ensemble des paramètres :  $\Theta = [1, 10]$ .
- Temps d'observation :  $T = 20s$ .
- Pas de discrétisation en temps :  $5 \times 10^{-3}$ .

Pour ces valeurs des paramètres, nous avons :  $I(\theta_0) = 149.9915470$ .

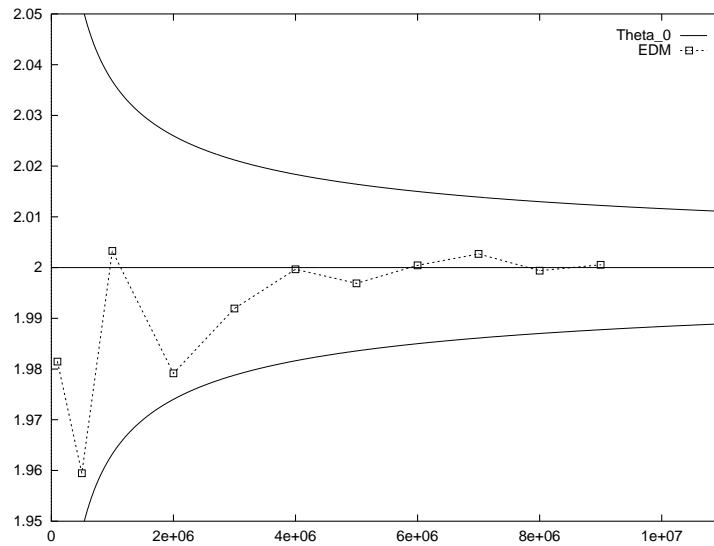


FIG. 6 - Estimateur de la distance minimale (EDM).

## 5 Conclusion

Dans ce travail, nous avons étudié l'identification des systèmes non-linéaires partiellement observés, par l'estimateur de la distance minimale, et obtenu pour une asymptotique de type "grand nombre d'observations indépendantes", les résultats de consistance et de normalité asymptotique.

Parmi les extensions possibles, nous pouvons considérer les problèmes suivants.

### Bruit d'observation dépendant de l'état

Considérons le modèle plus général suivant

$$\begin{cases} dX_t = b_t(\theta, X) dt + \sigma_t(\theta, X) dW_t^\theta, & X_0 = x_0 \\ dY_t = h_t(\theta, X) dt + r_t(X) dV_t^\theta, & Y_0 = 0, \end{cases}$$

où  $\{W_t^\theta, 0 \leq t \leq T\}$  et  $\{V_t^\theta, 0 \leq t \leq T\}$  sont deux processus de Wiener standard indépendants sous la probabilité  $\mathbf{P}_\theta$ . Dans ce modèle, la covariance du bruit d'observation dépend de la trajectoire passée du processus  $X$ .

Nous définissons encore le processus moyenne empirique

$$\hat{Y}_t^n = \frac{1}{n} \sum_{j=1}^n Y_t^j$$

et sous la probabilité  $\mathbf{P}_\theta$ , nous avons :

$$d\hat{Y}_t^n = \left\{ \frac{1}{n} \sum_{j=1}^n h_t(\theta, X^j) \right\} dt + \frac{1}{\sqrt{n}} \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^n r_t(X^j) dV_t^{\theta,j} \right\}, \quad \hat{Y}_0^n = 0.$$

Le processus  $\hat{V}^{n,\theta} = \{\hat{V}_t^{n,\theta}, 0 \leq t \leq T\}$  défini par

$$\hat{V}_t^{n,\theta} = \frac{1}{\sqrt{n}} \sum_{j=1}^n \int_0^t r_s(X^j) dV_s^{\theta,j}$$

est une martingale continue, nulle en zéro, de processus croissant

$$\langle \hat{V}^{n,\theta} \rangle_t = \int_0^t \left\{ \frac{1}{n} \sum_{j=1}^n r_s r_s^*(X^j) \right\} ds.$$

Si la vraie valeur du paramètre était  $\theta$ , alors pour  $n$  assez grand, on aurait encore

$$\widehat{Y}_t^n \simeq y_t(\theta) = \int_0^t \bar{h}_s(\theta) ds \quad 0 \leq t \leq T$$

avec  $\bar{h}_t(\theta) = \mathbf{E}_\theta[h_t(\theta, X)]$  pour tout  $0 \leq t \leq T$ .

L'estimateur de la distance minimale est alors défini comme précédemment par

$$\theta_n^* \in \underset{\theta \in \Theta}{\operatorname{argmin}} \|\widehat{Y}^n - y(\theta)\|.$$

Il semble possible de montrer la consistance et la normalité asymptotique de l'EDM, quand la taille  $n$  de l'échantillon tend vers l'infini.

**Remarque 5.1** La fonction de contraste  $K(\theta, \theta_0) = \|y(\theta) - y(\theta_0)\|$  ne dépend pas du coefficient  $r_t(\cdot)$ . L'EDM ne permet donc pas d'identifier ce coefficient.

**Remarque 5.2** Même dans le cas plus simple où les coefficients du système dépendent seulement de l'état courant  $X_t$ , c'est-à-dire dans le modèle :

$$\begin{cases} dX_t = b_t(\theta, X_t) dt + \sigma_t(\theta, X_t) dW_t^\theta, & X_0 = x_0 \\ dY_t = h_t(\theta, X_t) dt + r_t(X_t) dV_t^\theta, & Y_0 = 0, \end{cases}$$

le problème de filtrage non-linéaire (estimation de l'état  $X_t$  au vu des observations  $\mathcal{Y}_t$ ) ne semble pas résolu actuellement. Par conséquent, les méthodes qui reposent sur le filtrage non-linéaire, ne permettent pas d'identifier un tel modèle.

### Asymptotique du type “petits bruits”

Considérons le modèle suivant

$$\begin{cases} dX_t = b_\theta(X_t) dt + \varepsilon dW_t^{\theta, \varepsilon}, & X_0 = \bar{x}_0 \\ dY_t = h_\theta(X_t) dt + \varepsilon dV_t^{\theta, \varepsilon}, & Y_0 = 0, \end{cases}$$

où  $W^{\theta, \varepsilon} = \{W_t^{\theta, \varepsilon}, 0 \leq t \leq T\}$  et  $V^{\theta, \varepsilon} = \{V_t^{\theta, \varepsilon}, 0 \leq t \leq T\}$  sont deux processus de Wiener standard indépendant sous la probabilité  $\mathbf{P}_{\theta, \varepsilon}$ . James-Le Gland [13] ont montré la consistance de l'estimateur du maximum de

vraisemblance  $\hat{\theta}_\varepsilon$  quand  $\varepsilon$  tend vers zéro. Leur méthode repose sur un résultat de grandes déviations, qui permet de montrer la convergence uniforme sur  $\Theta$  de la log-vraisemblance vers une fonction de contraste appropriée, en probabilité.

Kutoyants a étudié, pour ce modèle, l'estimateur de la distance minimale défini par :

$$\theta_\varepsilon^* \in \operatorname{argmin}_{\theta \in \Theta} \|Y - y^\theta\|$$

où  $y^\theta = \{y_t^\theta, 0 \leq t \leq T\}$  est la sortie du système déterministe limite

$$\begin{cases} \dot{x}_t^\theta = b_\theta(x_t^\theta), & x_0^\theta = \bar{x}_0 \\ \dot{y}_t^\theta = h_\theta(x_t^\theta), & y_0^\theta = 0. \end{cases}$$

Il a étudié aussi l'EDM construit à partir de la "sortie du filtre de Kalman", associé à ce modèle. Pour plus de détails, on consultera [16]. Les résultats de consistance et de normalité asymptotique quand  $\varepsilon$  tend vers zéro peuvent être prouvés pour l'estimateur  $\theta_\varepsilon^*$  avec des techniques beaucoup plus simple à mettre en œuvre que pour l'estimateur de maximum de vraisemblance  $\hat{\theta}^\varepsilon$ .

### Choix de la mesure positive $\mu$

Nous avons obtenu les résultats de consistance et de normalité asymptotique de l'EDM  $\theta_n^*$ , sans aucune restriction sur la mesure positive  $\mu$  sur  $[0, T]$ . Nous avons donc la possibilité de choisir cette mesure, de façon à optimiser les propriétés asymptotiques de l'EDM. On peut chercher à minimiser la norme de la matrice de covariance  $I(\theta_0)$  définie par (16), ou à rendre la fonction de contraste  $\|y(\theta) - y(\theta_0)\|$  "plus contrastée".

## Références

- [1] K.J. ÅSTRÖM. *Introduction to Stochastic Control Theory*. Academic Press, New York, 1970.
- [2] A. BAGCHI. Consistent estimates in continuous-time systems. In O.L.R. Jacobs, M.H.A. Davis, M.A.H. Dempster, C.J. Harris, and P.C. Parks, editors, *Analysis and Optimization of Stochastic Systems, Oxford 1978*, pages 437–450, London, 1980. Academic Press.

- 
- [3] A. BAGCHI. A note on asymptotically efficient estimates of parameters in continuous-time dynamical systems. *SIAM Journal on Control and Optimization*, 19(2):169–176, 1981.
  - [4] A.V. BALAKRICHNAN. *Stochastic Differential Systems. Filtering and Control: A Function Space Approach*, volume 84 of *Lecture Notes in Economics and Mathematical Systems*. Springer Verlag, Berlin, 1973.
  - [5] I. V. BASAWA and B.L.S. PRAKASA RAO. *Statistical Inference for Stochastic Processes*. Academic Press, London, 1980.
  - [6] P. BERTRAND. Adaptive control for partially observable linear stochastic systems. *Stochastics and Stochastics Reports*. (à paraître).
  - [7] P. BILLINGSLEY. *Convergence of Probability Measures*. John Wiley & sons, New York, 1968.
  - [8] V.S. BORKAR and A. BAGCHI. Parameter estimation in continuous-time stochastic processes. *Stochastics*, 8(3):193–212, 1982.
  - [9] F. CAMPILLO and F. LE GLAND. MLE for the partially observed diffusions: direct maximization vs. the EM algorithm. *Stochastic Processes and their Applications*, 33:245–274, 1989.
  - [10] D. DACUNHA-CASTELLE and M. DUFLO. *Probabilité et Statistiques. 2. Problèmes à Temps Mobile*. Masson, Paris, 1983.
  - [11] A. DEMBO and O. ZEITOUNI. Parameter estimation of partially observed continuous time stochastic processes via the EM algorithm. *Stochastic Processes and their Applications*, 23(1):91–113, 1986.
  - [12] I.A. IBRAGIMOV and R.Z. KHASHMINSKII. *Statistical Estimation. Asymptotic Theory*, volume 16 of *Applications of Mathematics*. Springer Verlag, New York, 1981.
  - [13] M.R. JAMES and F. LE GLAND. Consistent parameter estimation for partially observed diffusions with small noise. Rapport de Recherche 1223, INRIA, May 1990. (revised, October 1991, April 1993).

- 
- [14] M.R. JAMES and F. LE GLAND. Identification of partially observed diffusions with small noise. In M.H.A. Davis and R.J. Elliott, editors, *Applied Stochastic Analysis*, volume 5 of *Stochastics Monographs*, pages 561–568. Gordon and Breach, 1991.
- [15] G. KALLIANPUR and R.S. SELUKAR. Parameter estimation in linear filtering. *Journal of Multivariate Analysis*, 39:284–304, 1991.
- [16] Yu.A. KUTOYANTS. *Identification of dynamical systems with small noise*. Kluwer. (à paraître).
- [17] Yu.A. KUTOYANTS. *Parameter Estimation for Stochastic Processes*, volume 6 of *Research and Exposition in Mathematics*. Heldermann Verlag, Berlin, 1984.
- [18] Yu.A. KUTOYANTS. Minimum distance parameter estimation for diffusion type observations. *CRAS, Série I*, 312:637–642, 1991.
- [19] A. LE BRETON. *Sur l'Estimation de Paramètres dans les Modèles Différentiels Stochastiques Multidimensionnels*. Thèse d'Etat, Université Scientifique et Médicale de Grenoble, 1976.
- [20] R.Sh. LIPTSER and A.N. SHIRYAYEV. *Statistics of Random Processes I. General Theory*, volume 5 of *Applications of Mathematics*. Springer Verlag, New York, 1977.
- [21] P.W. MILLAR. The minimax principle in asymptotic statistical theory. In P.L. Hennequin, editor, *Ecole d'Eté de Probabilités de Saint-Flour XI – 1981*, volume 976 of *Lecture Notes in Mathematics*, pages 75–265, Berlin, 1983. Springer Verlag.
- [22] P.W. MILLAR. A general approach to the optimality of minimum distance estimators. *Transactions of the American Mathematical Society*, 286(1):377–418, November 1984.



---

Unité de recherche INRIA Lorraine, Technôpole de Nancy-Brabois, Campus scientifique,  
615 rue de Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY  
Unité de recherche INRIA Rennes, IRISA, Campus universitaire de Beaulieu, 35042 RENNES Cedex  
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1  
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex  
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

---

Éditeur

INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)

ISSN 0249-6399