Observations bruitées d'une diffusion. Estimation, filtrage, applications.

Benjamin Favetto

UFR de Mathématiques et Informatique
Ecole doctorale Mathématiques Paris Centre

THÈSE

pour obtenir le grade de

Docteur de l'Université Paris Descartes

Spécialité : Mathématiques

Présentée par

Benjamin FAVETTO

Observations bruitées d'une diffusion Estimation, filtrage, applications

Soutenue le 30 septembre 2010 devant le jury composé de :

Susanne DITLEVSEN	Rapportrice
Valentine GENON-CATALOT	Directrice de thèse
Arnaud GLOTER	Examinateur
Jean JACOD	Président
Mathieu KESSLER	Examinateur
Catherine LAREDO	Examinatrice
Adeline SAMSON	Examinatrice
Denis TALAY	Rapporteur

"Statistics are like a bikini.
What they reveal is suggestive,
but what they conceal is vital."
Aaron Levenstein

Remerciements

Au moment de rédiger les dernières lignes de cette thèse – qui seront paradoxalement les premières, et peut-être les seules à être lues – il m'a semblé essentiel de revenir sur ces trois dernières années d'aventure scientifique mais aussi humaine, et d'adresser mes plus vifs remerciements à celles et ceux qui ont contribué, directement ou indirectement, à l'achèvement de ce travail.

Mes premiers mots sont pour Valentine Genon-Catalot, qui a accepté de diriger cette thèse et qui a guidé mes premiers pas en recherche dans le domaine de la statistique des diffusions. Toujours prête à écouter mes idées même les plus saugrenues, d'une grande disponibilité, son enthousiasme communicatif et ses conseils judicieux m'ont toujours permis d'avancer, surtout dans les moments délicats. Je crois que ces quelques deux cents pages suffiraient difficilement à exprimer toute ma gratitude envers elle, pour le travail qu'elle m'a permis d'accomplir et le soutien constant et bienveillant qu'elle m'a apporté ... Merci encore !

Je remercie également Denis Talay pour avoir accepté la pénible tâche de lire ce manuscrit à une période où il est généralement plus agréable de lire un bon roman sous un parasol, pour les suggestions et les améliorations qu'il m'a proposées, et pour m'avoir honoré de sa présence dans le jury. I also adress many thanks to Susanne Ditlevsen, to have spent a part of her holidays to report this thesis, and for all the discussions we have had during my PhD about the applications of SDE in biology.

Je n'aurais sans doute pas eu le même intérêt pour les applications des diffusions à la biologie, et il manquerait une partie à cette thèse, sans les travaux menés en collaboration avec Adeline Samson. Elle a été pour moi un modèle de rigueur et de ténacité, toujours prête à faire parler un jeu de données récalcitrant, à prodiguer d'utiles conseils, ou à relever un pari ... Je la remercie d'avoir accepté d'être membre du jury, et pour ces trois ans passés à travailler avec elle.

Jean Jacod m'a fait l'honneur de présider le jury de soutenance, et j'en suis particulièrement touché. Il avait accepté de répondre à mes questions au moment de mon stage de Master, je crois que la boucle est désormais bouclée ! Catherine Laredo a suivi mes travaux depuis le Master, et m'a invité à exposer au groupe de

travail de l'INRA, je la remercie de participer au jury. Mathieu Kessler a accepté de faire un long trajet pour la soutenance, et je lui en suis reconnaissant. Je remercie enfin Arnaud Gloter d'avoir participé au jury, et d'avoir répondu à mes questions sur les diffusions bruitées.

Je me dois également de remercier Charles-André Cuénod pour m'avoir permis de travailler sur ses données, et d'avoir accueilli avec enthousiasme nos équations, et je remercie Yves Rozenholc pour m'avoir permis d'intégrer le projet "Cancer et outils mathématiques", mais aussi pour sa connaissance intime des moindres recoins de Matlab^{TM} et des bonnes tables des environs de Saint-Flour.

Cette thèse a été réalisée au laboratoire MAP5 de l'Université Paris-Descartes, et il est important de souligner le rôle qu'ont joué Christine Graffigne, Annie Raoult et l'ensemble des personnels administratifs et techniques, pour m'avoir assuré durant ces trois années de bonnes conditions matérielles de travail, et pour m'avoir permis d'aller ~~en vacances~~ en congrès dans des endroits aussi exotiques que Lipari, Berlin, Fréjus, Saint-Flour ou Angers.

J'en profite pour remercier l'ensemble des membres du laboratoire, que j'ai eu le plaisir de cotoyer et dont les bavardages informels de la pause déjeuner ou goûter m'ont peu à peu affranchi des pratiques du Milieu. J'adresse aussi un remerciement particulier à celles et ceux qui m'ont permis de jouer au professeur sans trop s'alarmer du sort des étudiants qui m'étaient confiés : Christine Graffigne, Antoine Chambaz, Avner Bar-Hen et François Patte, ainsi que Jean-Claude Fort pour les discussions pédagogiques nombreuses et variées, et Mireille Chaleyat-Maurel qui a été tutrice de mon monitorat. Je dois aussi une partie de mon savoir-faire pédagogique à Vincent Mercat, et je lui suis reconnaissant de m'avoir accordé sa confiance en m'intronisant *Monsieur Maple* de sa classe.

Je me dois aussi de réitérer un remerciement tout particulier à Avner Bar-Hen, non seulement car il est toujours bon de remercier Monsieur le Président de la Société Française de Statistique en pareille situation, mais aussi pour m'avoir donné ce qui a peut être été le meilleur conseil de toute ma thèse ...

Je suis aussi reconnaissant à celles et ceux qui m'ont encouragé sur la voie des mathématiques : Claude Phillioux et Christian Ulrich, Alain Chillès, Michel Pierre et Arnaud Debussche, Florent Malrieu, Arthur Charpentier, Laure Elie, et Philippe Berthet pour m'avoir présenté avec une rigueur toute bourbachique la difficulté et la beauté de la Statistique.

Parmi les personnes qui ont contribué, même occasionnellement, à ce que le bureau H409-2 ne se transforme pas complètement en cour des miracles de l'intelligence – ou en annexe d'un asile d'aliénés selon la saison – je tiens à remercier les admirables cheffes des thésards que furent Claire Lacour et Cécile Louchet, Arno Jegousse pour avoir assuré la sonorisation du bureau aux heures tardives

et pour les discussions sur le championnat de football de National, Maël Primet, Stefan Wolfsheimer, Nathalie Akakpo et Jean-Patrick Baudry pour leur placidité en toute circonstance et leur aide en Latex, et enfin Mahendra Mariadassou, dont la connaissance du site de l'INSEE nous a souvent tiré de situations délicates lors du quart d'heure d'analyse économique de comptoir-psychanalyse de groupe de la pause café, et dont les bavardages électroniques silencieux furent tour à tour informatifs, humoristiques et réconfortants. Je remercie aussi *les gens du septième* : Sylvain Pelletier, Mohamed Hachama, Julien Stirnemann – qui aura le bon goût d'être deux fois docteur – Maxime Bérar, Vanessa Dumeaux, Alice Démarez, Solange Whegang, Georges Djimefo, Baptiste Coulange, Christophe Denis, Jean-Pascal Jacob, pour leur hospitalité méridienne ainsi que les petits nouveaux qui sont prêts à reprendre le flambeau : Mélina Bec, Djeneba Thiam, Gaëlle Chagny et Camille Garcin. Enfin, je remercie aussi Emeline Schmisser, pour nous avoir, malgré tout, bien fait rire pendant trois ans.

J'en profite enfin pour remercier celles et ceux qui, de près ou de loin, m'ont assuré de leur soutien pendant ces trois dernières années. Tout d'abord, les Lyonnais avec qui je partage le souvenir de mes jeunes et insouciantes années : Mounir, Jean, ~~Ludovic~~ Mollah, Peyo, Hélène, Audrey, Emmanuel et Papou dont l'humour décapant et le goût des bonnes tables m'ont souvent encouragé à sécher la cantine ; les copains de l'AMT : Christophe, Bernadette, Doudou, Rafik et Momo ... Puis les Rennais rendirent mes trois années d'expatrié breton beaucoup plus amusantes que je ne l'avais imaginé : Jon, Mikl, Petit Ludo, ProutProut, Aurélien, Sébastien, Marie, Adeline, Nichouire, Cécile, François, Ronan, et tous ceux que j'ai eu plaisir à cotoyer à l'Ecole, que ce soit pendant l'année de BDE, sur le tournage de *Bouche Touche Mouche*, ou à la BU en train de réviser l'agreg ! Une mention spéciale à Marc et Lise-Marie, dont les conseils pleins de sagesse me furent utiles plus d'une fois, à mes colocataires passés et présent, Grand Ludo, Kiki et Ben, qui m'ont supporté – dans tous les sens du terme – jusqu'ici et qui ne sont pas étrangers à une amélioration modeste mais réelle de mes talents de cuisinier, et à Mikael, pour avoir partagé avec moi l'idée que faire des mathématiques était un combat (*Force et honneur* !), pour son goût prononcé pour les cordons-bleus et les abricots, mais aussi pour le dériveur, et enfin pour avoir m'avoir permis de tester en sa compagnie les hauts lieux de la vie nocturne européenne, de Londres à Berlin en passant par Saint-Flour et l'île d'Arz. Enfin, les Parisiens m'ont écouté me plaindre de la difficile condition de thésard avec bienveillance et bonhommie : tout d'abord Joseph S., qui peut témoigner de la véracité de mes dires (sic) ; puis les camarades du Pré Saint-Gervais, Denis, Gwendoline et Mathias, qui souvent me demandèrent timidement "Alors, ta thèse ... ça va ?" en ayant l'air d'avoir un peu peur de ma réponse ; et les amis du Genepi avec qui

j'ai eu plaisir à faire *tout autre chose que des mathématiques*, en y consacrant parfois un temps que d'aucuns auraient trouvé bien trop important par rapport à celui dévolu à ma thèse, et avec le seul regret de ne pas en avoir fait plus ... Ce fut un plaisir d'oeuvrer avec Charlène, Oriane et Gaëlle à Villepinte, d'intervenir aux côtés d'Elise et de Juan, de rencontrer Laure, Chirine, Gaspard, Pénélope, les Josephs, Rapahël, Caroline, et tant d'autres enfants de Cayenne !

J'ai un dernier mot pour mes parents et ma famille, chez qui j'ai toujours pu retrouver un havre de tranquillité lyonnais quand le travail le nécéssitait. Et, à l'heure d'écrire ces remerciements, j'ai une pensée pleine d'affection pour toutes celles qui m'ont permis de jouer au docteur avant même d'avoir soutenu.

De ces trois années, il me restera, je crois, un peu plus que quelques bouteilles de champagne vides, six cent grammes d'encre et de papier, quelques théorèmes et un carton jaune porteur du sceau de l'Université ... *Telle est la vie des hommes. Quelques joies, très vite effacées par d'inoubliables chagrins. Il n'est pas nécessaire de le dire aux enfants. (Marcel Pagnol)*

Table des matières

1 Introduction **1**
1.1 Introduction 1
1.2 Présentation des travaux 3
1.2.1 Première partie : estimation paramétrique pour un processus d'Ornstein-Uhlenbeck partiellement observé 3
1.2.1.1 Chapitre 2 : Résultats théoriques 3
1.2.1.2 Chapitre 3 : Application à un problème médical . 4
1.2.2 Deuxième partie : estimation des paramètres d'une diffusion bruitée avec données haute fréquence 7
1.2.2.1 Chapitre 4 : Estimation par minimisation de contraste 8
1.2.2.2 Chapitre 5 : Fonctions d'estimation et normalité asymptotique des estimateurs 10
1.2.3 Troisième partie : étude de la variance dans le théorème central limite pour le filtre particulaire 11
1.2.3.1 Chapitre 6 : Tension de la variance asymptotique 11

I Estimation paramétrique pour un processus d'Ornstein-Uhlenbeck caché dans un contexte biomédical **15**

2 Parameter estimation **17**
2.1 Introduction 18
2.2 Study of the stochastic differential equation 20
2.3 Parameter estimation by maximum likelihood 21
2.3.1 Computation of the exact likelihood 22
2.3.1.1 Kalman filter 22
2.3.1.2 Computation of the exact likelihood of the observations 23
2.3.2 Computation of the maximum likelihood estimator 23
2.3.2.1 New parametrization 24

2.3.2.2 Computation of the exact gradient and hessian of the log-likelihood . . . 24
2.3.2.3 Maximisation of the exact likelihood . . . 25
2.4 Parameter estimation by Expectation Maximization algorithm . . 26
2.5 Properties in the stationary case . . . 27
2.5.1 Link with an ARMA model . . . 27
2.5.2 Asymptotic properties of maximum likelihood estimator . 29
2.6 Simulation study . . . 30
2.7 Conclusion . . . 33

Appendices **35**

A Appendix **37**
A.1 Physiological model . . . 37
A.2 Proof of Proposition 2.1 . . . 38
A.3 Link between the two parametrisations . . . 38
A.4 Gradient and hessian of the log-likelihood . . . 39
A.5 Smoother algorithm . . . 41
A.6 ARMA property of multidimensional process . . . 41
A.7 First order identifiability . . . 43

3 Blood micro-circulation parameters **47**
3.1 Introduction . . . 48
3.2 Models . . . 49
3.2.1 MRI data extraction . . . 49
3.2.2 Pharmacokinetic models . . . 50
3.2.3 The ordinary differential pharmacokinetic model (ODE) . 50
3.2.4 The stochastic differential pharmacokinetic model (SDE) . 51
3.3 Estimation methods . . . 52
3.3.1 Estimation in the ODE model . . . 52
3.3.2 Estimation in the SDE model . . . 53
3.3.3 Numerical implementation of the two estimation methods . 54
3.4 Estimation on real data . . . 55
3.5 Simulated study . . . 63
3.6 Discussion . . . 63
3.7 Appendix . . . 68
3.7.1 The Kalman filter . . . 68
3.7.2 The Kalman smoother algorithm . . . 69
3.7.3 Computation of the Fisher Information Matrix . . . 69

3.7.4 Model shapes on simulated data 70

II Estimation des paramètres d'une diffusion bruitée pour des données haute fréquence 73

4 Contrast minimization 75
4.1 Introduction 76
4.2 Assumptions and Notations 78
4.3 Small sample properties of the local means sample 80
4.4 Uniform convergence in probability results 82
4.5 Statistical estimation by contrast minimization 83
4.5.1 Definition of the contrasts 83
4.5.2 Estimation with unknown observation noise level under **(B1)** 84
4.5.3 Link with the case of noisy observations of integrated diffusions 85
4.6 Examples 85
4.6.1 The Ornstein-Uhlenbeck process (simulation) 85
4.6.2 Comparison with a discretely observed Ornstein-Uhlenbeck process 88
4.6.3 The Ornstein-Uhlenbeck process (neuronal data) 89
4.6.4 The Cox-Ingersoll-Ross process 92
4.6.5 The hyperbolic diffusion 94
4.7 Concluding remarks 95
4.8 Proofs 96

Appendices 111

B appendix 113

5 Estimating equations 117
5.1 Introduction 118
5.2 Model and Assumptions 120
5.3 Rate-optimal estimating functions for local means 121
5.4 Applications and examples 127
5.4.1 The Ornstein-Uhlenbeck diffusion 130
5.4.2 The hyperbolic diffusion 130
5.5 Concluding remarks 131
5.6 Auxiliary tools 131
5.7 Proofs 131

III Filtre particulaire **145**

6 Asymptotic variance **147**
6.1 Introduction 148
6.2 Notations, assumptions and preliminary results 150
6.3 Tightness of the asymptotic variances 154
6.4 Discussion and examples 157
6.4.1 Checking of **(B3)** 157
6.4.2 The case of a diffusion on a compact manifold 158
6.4.3 A toy-example 159
6.5 Numerical simulations 160
6.5.1 Simulations based on the toy-example 160
6.6 Acknowledgements 162

Appendices **163**

C Appendix **165**
C.1 Bootstrap particle filter 165
C.2 Tightness lemma 165
C.3 The Gaussian case 166
C.3.1 Preliminary computations 167
C.3.2 Solving recursions for the stabilized operators 168
C.4 Simulations based on a Gaussian AR(1) model 170
C.4.1 Inequality between Gaussian distributions 170

Liste des tableaux

2.1 Mean estimated values (with empirical standard errors in bracket) obtained with the exact MLE and the EM algorithms and exact standard errors obtained from the Fisher information matrix, evaluated on 1000 simulated data with $n = 200$ and $n = 1000$ observations and $\sigma^2 = 1$ or $\sigma^2 = 3$ (σ^2 and θ_5 fixed to their true values). 32

2.2 Mean estimated values (with empirical standard errors in bracket) obtained with the exact MLE and the EM algorithms and exact standard errors obtained from the Fisher information matrix, evaluated on 1000 simulated data with $n = 200$ and $n = 1000$ observations and $\sigma^2 = 1$ or $\sigma^2 = 3$ (σ^2 and θ_5 fixed to their true values). 33

3.1 Estimated parameters for the first dataset, using the ODE and the SDE models. The values in parenthesis are the standard deviations evaluated using a numerical computation of the Fisher Information matrix. For predictions, see Figure 3.2. 56

3.2 Estimated parameters for dataset 2, using the ODE and the SDE models. The values in parenthesis are the standard deviations evaluated using a numerical computation of the Fisher Information matrix. For predictions, see Figure 3.3. 56

3.3 Estimated parameters for dataset 3, using the ODE and the SDE models. The values in parenthesis are the standard deviations evaluated using a numerical computation of the Fisher Information matrix. For predicted curves, see Figure 3.4. 59

3.4 Estimated parameters for dataset 4, using the ODE model (column 2), the SDE model (column 3) and using the ODE and SDE models after removing the last 3 (columns 4 and 5) and the last 5 (column 6 and 7) observations. The values in parenthesis are the standard deviations evaluated using a numerical computation of the Fisher Information matrix. The differences in parameters estimation by ODE influence drastically the enhancement curves (Figure 3.5). . 61

3.5 Estimation results for three different sets of fixed parameters. Estimation when simulating under the ODE model (top part of the Table), when simulating under the SDE model ($\sigma_1 = \sigma_2 = 2$) (middle part of the Table) and when simulating with a small PS ($PS = 1$) under the SDE model ($\sigma_1 = \sigma_2 = 1$) (bottom part of the Table). Empirical means and standard deviations (in parenthesis) are computed for each estimated parameter from 100 simulated datasets analyzed by the ODE and the SDE estimations. The NA values express that the associated quantities are not available in the considered model. 64

4.1 Influence of the size of blocks on the estimators, Ornstein-Uhlenbeck model. ($N = 5000$ observations, $\delta = 0.01$, 500 replications) $\kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$ 86

4.2 Influence of the size of blocks on the estimators, Ornstein-Uhlenbeck model. ($N = 20000$ observations, $\delta = 0.005$, 500 replications) $\kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$ 86

4.3 Influence of the size of blocks on the estimators, Ornstein-Uhlenbeck model. ($N = 100000$ observations, $\delta = 0.001$, 500 replications) $\kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$ 87

4.4 Influence of the observation noise variance on the estimators, Ornstein-Uhlenbeck model. (500 replications, $N = 10^5$, $\delta = 0.001$, $\alpha = \frac{3}{2}$) . 87

4.5 Influence of the diffusion coefficient on the estimators, Ornstein-Uhlenbeck model. (500 replications, $N = 10^5$, $\delta = 0.001$, $\alpha = \frac{3}{2}$) . 88

4.6 Influence of the distribution of the noise on the estimation, Ornstein-Uhlenbeck model. ($N = 10^5$ observations, $\delta = 0.0001$, $\alpha = \frac{3}{2}$) . . . 88

4.7 Parameter estimation with direct observations of the Ornstein-Uhlenbeck model, for several numbers of observations. (500 replications, $\alpha = 1.5, \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$) 89

4.8 Parameter estimation for neuronal data (with measurement error). 91

4.9 Parameter estimation for neuronal data (without measurement error). 91

4.10 94
4.11 Parameter estimation for the Cox-Ingersoll-Ross process with direct observations for different values of α. (500 replications, $\kappa_0 = -2, \kappa_0' = 3, \lambda_0 = 4, \rho^2 = 0.5, \alpha = \frac{3}{2}$ 94
4.12 Parameter estimation for the hyperbolic diffusion, with a Gaussian noise, for different values of N. (500 replications, $\alpha = \frac{3}{2}, \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$) 95
4.13 Parameter estimation for the hyperbolic diffusion, with a Laplace noise, for different values of N. (500 replications, $\alpha = \frac{3}{2}, \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$) 96

6.1 Computation of (6.16) for different values of N and k. 161

Table des figures

1.1 Modèle de Markov caché 2

1.2 Modèle pharmacocinétique utilisé pour l'agent de contraste 5

2.1 Noisy observations of one simulated data set ($n = 200$, $\Delta = 0.2$, $\sigma^2 = 3$) are plotted with stars. True simulated trajectories (thin solid lines), mean estimated trajectories (thick solid lines) and estimated 95% confidence intervals (dotted lines) obtained with the Kalman algorithm are plotted with dark lines for $S(t)$, light lines for $P(t)$ and very light line for $I(t)$. 31

3.1 Two-compartment physiological pharmacokinetic model used to describe the distribution of the contrast agent. The hatched compartments representing the red blood cells and the tissue cells are not reached by the contrast agent. 51

3.2 Predictions for the first dataset, obtained with the ODE model (left) and the SDE model (right) : black stars ($*$) are the tissue observations (y_i), the AIF observations are represented by the red line, crosses ($\times$) are the residuals. The plain blue, dashed pink and dash-dotted green lines are respectively the predictions for $S(t)$, $Q_P(t)$ and $Q_I(t)$. Estimated parameters are from Table 3.1. 57

3.3 Top two figures : predictions for dataset 2, obtained with the ODE model (left) and the SDE model (right) : black stars ($*$) are the tissue observations (y_i), crosses ($\times$) are the residuals. The plain blue, dashed pink and dashdotted green lines are respectively the predictions for $S(t)$, $Q_P(t)$ and $Q_I(t)$. For the SDE model, each prediction curve is surrounded by its 95% confidence intervals. Bottom two figures : first 15 partial autocorrelations of the residuals obtained with the ODE model (left) and the SDE model (right). The horizontal red dashed lines provide the 90% confidence interval around 0 for each partial autocorrelation. The horizontal violet dashed lines provide the Bonferroni confidence interval of the test that all partial autocorrelations are 0 together. The norm of these partial autocorrelations is 0.28 for the ODE estimation and 0.26 for the SDE estimation. Estimated parameters are from Table 3.2. 58

3.4 Top two figures : predictions for dataset 3, obtained with the ODE model (left) and the SDE model (right) : black stars ($*$) are the tissue observations (y_i), crosses ($\times$) are the residuals. The plain blue, dashed pink and dash-dotted green lines are respectively the predictions for $S(t)$, $Q_P(t)$ and $Q_I(t)$. For the SDE model, each prediction curve is surrounded by its 95% confidence intervals. Bottom two figures : first 15 partial autocorrelations of the residuals obtained with the ODE model (left) and the SDE model (right). The horizontal red dashed lines provide the 90% confidence interval around 0 for each partial autocorrelation. The horizontal violet dashed lines provide the Bonferroni confidence interval of the test that all partial autocorrelations are 0. The norm of these partial autocorrelations is 0.39 for the ODE estimation and 0.37 for the SDE estimation. Estimated parameters are from Table 3.3. 60

3.5 Top figures : predictions for dataset 4, obtained with the ODE model (left) and the SDE model (right) : black stars ($*$) are the tissue observations (y_i), crosses ($\times$) are the residuals. The plain blue, dashed pink and dashdotted green lines are respectively the predictions for $S(t)$, $Q_P(t)$ and $Q_I(t)$. For the SDE model, each prediction curve is surrounded by its 95% confidence intervals. Middle figures : predictions obtained with the ODE model (left) and the SDE model (right) removing the last 3 observations. Bottom figures : predictions obtained with the ODE model (left) and the SDE model (right) removing the last 5 observations (right). The adjustment differences come from the differences of parameter estimations (Table 3.4). 62

3.6 Typical trajectories of the ODE and SDE models for the ideal case $\sigma = 0$ (no measurement error). An ideal AIF was built artificially (plain line on the left figure). Parameter values are $F_T = 70$ ml min^{-1} 100 ml^{-1}, $V_b = 20$ %, $PS = 25$ ml min^{-1} 100 ml^{-1}, $V_e = 15$ %, $\delta = 15$ s. The functions Q_P, Q_I, S of the ODE system (3.2) are given on the left figure by the magenta, green and blue curves, respectively. On the right figures, five typical realizations of the processes Q_I (bottom), Q_P (middle), S (top) of the SDE model (3.3) are plotted for $\sigma_1 = \sigma_2 = 1$. 65

4.1 The neuronal dataset ($N = 35000$ observations, $\delta = 0.02 \times 10^{-2}$ seconds). 90

4.2 The observations with the estimated mean $\hat{\mu}_N$. 91

4.3 An example of Cox-Ingersoll-Ross process observed with a mutliplicative noise. 92

6.1 Densities involved in the toy-example. Functions u (solid), v (bigdash dot) and $\frac{d\mu}{dx}$ (dash dot). 160

6.2 Toy-example ($\alpha = 0.4$). Hidden Markov chain (plain, $\square$ marks). Observations (longdashed line, diamond marks). Particle filter (dashed line, + marks) 161

6.3 Histograms of the random variables (6.16) for $k = 50$ (top left), 100 (top right), 150 (bottom left), 200 (bottom right). 162

C.1 Kalman model. Hidden Markov chain (plain, square marks). Observations (longdashed line, diamond marks). Particle filter (dashed line, plus marks) 171

Chapitre 1

Introduction et synthèse des résultats

1.1 Introduction

On considère l'équation différentielle stochastique

$$dX_t = b(X_t, \theta)dt + \sigma(X_t, \theta)dB_t, \qquad X_0 = \eta \tag{1.1}$$

où (B_t) est un mouvement brownien standard, b, σ sont des fonctions dépendant d'un paramètre inconnu θ appartenant à une partie Θ de $\mathbb{R}^d$. Les modèles basés sur la diffusion (X_t) sont couramment utilisés en finance ou en biologie, et l'on considère alors des observations discrétisées X_{t_i} d'une trajectoire, aux instants $0 = t_0 < t_1 < \cdots < t_n = T$. De plus, afin de tenir compte d'éventuelles erreurs de mesure dans ces modèles, on suppose que l'on observe $Y_{t_i}, \quad i = 0, \ldots, n$ telles que

1. Conditionnellement à la trajectoire (X_t), les variables aléatoires Y_{t_i} sont indépendantes,
2. La distribution conditionnelle $\mathcal{L}(Y_{t_i}|(X_t))$ de Y_{t_i} sachant la trajectoire (X_t) ne dépend que de X_{t_i},
3. Sachant $X_{t_i} = x$, la distribution conditionnelle de Y_{t_i} ne dépend pas de i.

C'est le cas, par exemple, lorsque la diffusion est bruitée additivement. On observe alors

$$Y_{t_i} = X_{t_i} + \varepsilon_{t_i} \tag{1.2}$$

avec (ε_{t_i}) une suite de variables aléatoires indépendantes et identiquement distribuées. Sous ces hypothèses, la suite de variables aléatoires (X_{t_i}, Y_{t_i}) est un modèle de Markov caché (voir Cappé et al. (2005)), dont la chaîne cachée est issue de

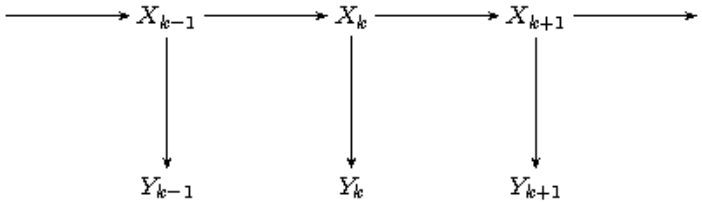

FIGURE 1.1 – Modèle de Markov caché

la discrétisation d'une diffusion. Deux questions statistiques sont abordées dans cette thèse :

1. l'estimation de θ, sur un plan théorique – propriétés asymptotiques des estimateurs – et appliqué – implémentation des estimateurs sur des données biomédicales – pour des diffusions observées avec erreur de mesure ;

2. certaines propriétés asymptotiques liées au filtrage particulaire.

De plus, on s'intéresse au cas où la diffusion (X_t) admet une probabilité stationnaire ν_0. Dans ce cas, on dispose du théorème ergodique

$$\frac{1}{T}\int_0^T f(X_s)ds \underset{T\to\infty}{\longrightarrow} \nu_0(f)$$

où la convergence est presque sûre, pour toute fonction $f \in L^1(\nu_0)$. On dira alors que (X_t) est une diffusion ergodique.

Le cadre asymptotique des différentes études est celui d'observations sur un long intervalle de temps $[0, T]$, c'est à dire lorsque $T = t_n$ tend vers l'infini en même temps que le nombre n d'observations. L'alternative pour la fréquence d'échantillonnage des données est la suivante :

1. le pas de temps $t_{i+1} - t_i = \Delta$ est fixe, on a alors $n\Delta = t_n$, et la chaîne de Markov $(X_{i\Delta})$ est ergodique ;

2. le pas de temps $t_{i+1} - t_i = \delta_n$ tend vers 0 lorsque le nombre d'observations tend vers l'infini, et $n\delta_n = t_n$ tend vers l'infini : c'est le contexte des *données haute fréquence*.

1.2 Présentation des travaux

1.2.1 Première partie : estimation paramétrique pour un processus d'Ornstein-Uhlenbeck partiellement observé

Cette première partie est composée de deux chapitres, motivés par un problème de statistique médicale, fruit d'un travail en collaboration entre le laboratoire MAP5 de l'Université Paris Descartes et le service de radiologie de l'Hopital Européen Georges Pompidou.

1.2.1.1 Chapitre 2 : Résultats théoriques

Ce chapitre reprend l'article en collaboration avec Adeline Samson (Favetto and Samson (2010)).

On considère un processus d'Ornstein-Uhlenbeck bidimensionnel (U_t), où $U_t = (U_t^{(1)}, U_t^{(2)})^T$ est défini par l'équation différentielle stochastique

$$dU_t = (G_\theta U_t + F_\theta)dt + \Sigma_\theta dB_t, \qquad X_0 = \eta. \tag{1.3}$$

Aux instants $t_i = i\Delta$, $i = 0, \ldots, n$, les observations Y_i sont données par

$$Y_i = JU_{t_i} + \sigma\varepsilon_i \tag{1.4}$$

avec $J = (1, 0)$ et (ε_i) une suite de variables aléatoires indépendantes et identiquement distribuées de loi commune $\mathcal{N}(0, 1)$. L'observation de la diffusion cachée est donc partielle – seulement la première des deux coordonnées de U_{t_i} – et bruitée.

L'estimation du paramètre θ est motivée par une application médicale présentée dans le chapitre suivant, le but étant de valider les méthodes proposées sur des simulations préalablement.

La première méthode envisagée est basée sur le calcul exact de la vraisemblance $L_n(\theta)$, grâce à l'algorithme de Kalman (voir, par exemple, le livre de Cappé et al. (2005) pour une présentation de l'algorithme, et le rapport technique de Pedersen (1994) pour une méthode d'approximation de l'estimateur du maximum de vraisemblance dans un contexte similaire). Ainsi,

$$L_n(\theta) = \prod_{i=1}^{n} p(Y_i | Y_{i-1}, \ldots, Y_0; \theta) \tag{1.5}$$

où $p(Y_i|Y_{i-1}, \ldots, Y_0; \theta)$ est la vraisemblance conditionnelle de Y_i sachant les observations précédentes. Dans une base de vecteurs propres de G_θ de matrice de

passage P, on considère $X_{i\Delta} = P^{-1}U_{i\Delta}$. L'algorithme de Kalman permet de calculer de façon exacte la vraisemblance conditionnelle $p(X_{i\Delta}|Y_{i-1}, \ldots, Y_0; \theta)$ de la variable cachée $X_{i\Delta}$, puisque la discrétisation du modèle continu mène à

$$X_{(i+1)\Delta} = AX_{i\Delta} + B + \eta_i, \tag{1.6}$$
$$Y_i = HX_i + \sigma\varepsilon_i \tag{1.7}$$

avec $H = (1, 1)$, et (η_i) une suite de variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{N}(0, R)$. La vraisemblance conditionnelle de Y_i s'en déduit directement.

De plus, le calcul exact du gradient et de la hessienne de la log-vraisemblance des observations s'obtient de façon itérative en différentiant les relations de Kalman. Ainsi, le calcul de l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ se fait à l'aide d'un algorithme de gradient conjugué, dans lequel la valeur de la log-vraisemblance, son gradient et sa hessienne sont calculés de façon exacte.

Les propriétés asymptotiques de l'estimateur $\hat{\theta}_n$ sont étudiées en lien avec celles des processus $ARMA$: on établit que, dans le cas stationnaire, le processus (Y_i), une fois recentré, est un processus ARMA(2,2) Gaussien. Ainsi, $\hat{\theta}_n$ est asymptotiquement normal, et efficace (voir, par exemple, le livre de Brockwell and Davis (1991)).

La méthode d'estimation basée sur le maximum de vraisemblance est comparée, sur des simulations, à celle basée sur l'algorithme E-M proposé par Dempster et al. (1977) (voir aussi Shumway and Stoffer (1982) et Segal and Weinstein (1989), par exemple). D'autre part, on établit que quatre des cinq paramètres au plus de la chaîne cachée issue de la diffusion discrétisée sont identifiables.

1.2.1.2 Chapitre 3 : Application à un problème médical

Ce chapitre repend l'article en collaboration avec Daniel Balvay, Charles-André Cuenod, Valentine Genon-Catalot, Yves Rozenholc, Adeline Samson et Isabelle Thomassin (Favetto et al. (2009)).

Ce travail porte sur l'étude d'un modèle pharmacocinétique stochastique à deux compartiments, et a pour but d'estimer les paramètres de microvascularisation à partir de données médicales – issues de l'imagerie par résonnance magnétique. Après l'injection d'un agent de contraste dans une artère d'un patient, on observe les variations de niveaux de gris sur une séquence d'images, en différentes parties de l'image – artère, tissus, organe – et pour différents *voxels* (pixels tridimensionnels).

Cette étude s'inscrit dans le contexte d'un projet plus large, menant au développement d'outils mathématiques, conjointement avec des médecins, pour me-

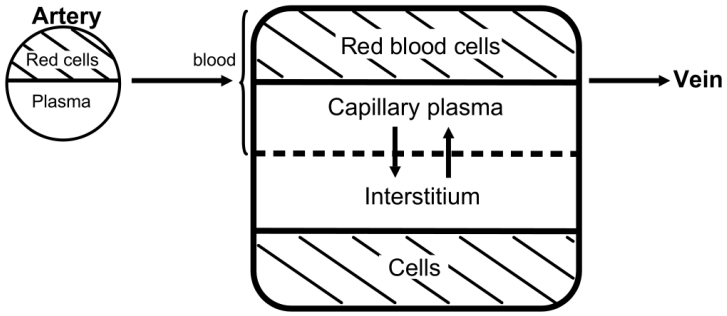

FIGURE 1.2 – Modèle pharmacocinétique utilisé pour l'agent de contraste

surer l'impact de nouveaux traitements contre le cancer (dits *anti-angiogénèse*) qui limitent le développement des tumeurs en réduisant le développement de vaisseaux sanguins. En particulier, la compréhension du phénomène de microvascularisation ainsi que l'estimation des paramètres dans le modèle proposé peuvent potentiellement être utilisés pour évaluer l'efficacité de ces traitements.

Les résultats présentés dans ce chapitre, et dans Favetto et al. (2009), ont pour but de proposer une validation du modèle pharmacocinétique stochastique, avec une étude sur données simulées, complétée par une étude sur données réelles, issues de patientes saines.

Soient $Q_P(t)$ la quantité d'agent de contraste dans le plasma à l'instant t et $Q_I(t)$ la quantité d'agent de contraste dans le milieu intersticiel à l'instant t, dans un voxel de l'image. On appelle $S(t) = Q_P(t)+Q_I(t)$ la quantité totale d'agent de contraste dans un voxel de l'image. La variation de niveaux de gris de l'image à un instant t est alors supposée proportionnelle à $S(t)$. De plus, afin de prendre en compte l'erreur de mesure commise lors de l'acquisition des données, on suppose que l'on observe

$$y_i = S(t_i) + \sigma\varepsilon_i \tag{1.8}$$

aux instants $t_0 = 0 < \cdots < t_n = T$, avec (ε_i) une suite de variables aléatoires indépendantes et identiquement distribuées, de loi commune $\mathcal{N}(0, 1)$.

Le modèle doit prendre en compte la fonction d'input artériel, notée AIF. Les paramètres biologiques sont les suivants : $F_T \geq 0$ est la constante de transfert entre le sang et les tissus, par unité de volume de tissus, $V_b \geq 0$ est le pourcentage de plasma par unité de volume, $V_e \geq 0$ est le pourcentage d'espace extravasculaire, et $PS \geq 0$ est la constante de perméabilité du produit par unité de volume de tissus. Le taux d'hématocrite h est fixé à $h = 0.4$. Le délai avec lequel l'agent de contraste passe des artères au plasma est noté δ.

On dispose d'un modèle pharmacocinétique déterministe pour l'agent de contraste

(voir Brochot et al. (2006) et Figure **??**) résultant du bilan des quantités d'agent de contraste à un instant donné dans les deux compartiments formés par le plasma et le milieu intersticiel. C'est un système de deux équations différentielles couplées, avec un retard dans la fonction AIF pour prendre en compte le délai d'arrivée de l'agent de contraste dans le voxel considéré :

$$\begin{aligned} dQ_P(t) &= \left(\frac{F_T}{1-h}\, AIF(t-\delta) - \frac{PS}{V_b(1-h)} Q_P(t) + \frac{PS}{V_e} Q_I(t) - \frac{F_T}{V_b(1-h)} Q_P(t) \right) dt \\ dQ_I(t) &= \left(\frac{PS}{V_b(1-h)} Q_P(t) - \frac{PS}{V_e} Q_I(t) \right) dt \end{aligned}$$

Pour les praticiens, ce modèle donne un résultat trop lisse, ne tenant pas compte des fluctuations locales, liées par exemple aux mouvements du patient lors de l'expérience d'acquisition des données. On considère donc le modèle suivant, basé sur un système de deux équations différentielles stochastiques, dont la dérive (et donc le comportement moyen) est donnée par le système déterministe :

$$\begin{aligned} dQ_P(t) &= \left(\frac{F_T}{1-h}\, AIF(t-\delta) - \frac{PS}{V_b(1-h)} Q_P(t) + \frac{PS}{V_e} Q_I(t) - \frac{F_T}{V_b(1-h)} Q_P(t) \right) dt \\ &\quad + \sigma_1 dW_t^1 \qquad (1.9) \\ dQ_I(t) &= \left(\frac{PS}{V_b(1-h)} Q_P(t) - \frac{PS}{V_e} Q_I(t) \right) dt + \sigma_2 dW_t^2 \end{aligned}$$

On observe donc partiellement une diffusion bidimensionnelle à des instants discrets, avec un bruit de mesure : le modèle de ce travail a motivé les développements mathématiques du précédent chapitre.

On se propose alors de comparer la méthode d'estimation par maximum de vraisemblance pour le modèle de diffusion – développée dans le chapitre précédent – à la méthode de référence utilisée par les praticiens, basée sur le modèle d'équation différentielle ordinaire et une méthode de moindres carrés.

On effectue alors plusieurs simulations du modèle, en se basant sur une fonction AIF "idéale" (lisse). Sur des jeux de données simulées, on obtient une réduction significative du biais et de l'écart-type en faveur de la méthode basée sur le système d'équations différentielles stochastiques, et ceci même si les valeurs obtenues pour σ_1 et σ_2 sont faibles.

D'autre part, plusieurs jeux de données provenant de séquences d'images de pelvis féminins sains ont été traités, en vue de la validation expérimentale de la méthode. On présente dans ce chapitre les résultats les plus significatifs pour quatre patientes. Lors de l'implémentation effective des deux méthodes, l'un des principaux résultats obtenus concerne la stabilité de la méthode stochastique : l'estimation est, dans ce cas, beaucoup plus robuste que la méthode déterministe

lorsqu'on enlève trois, puis cinq observations, en particulier lorsque la constante de perméabilité PS est faible. Enfin, l'intérêt de la méthode est aussi visible sur la reconstruction des courbes retraçant l'évolution des quantités d'agent de contraste dans le plasma et l'interstitium.

Les résultats obtenus pour ces quatre patientes et présentés dans ce chapitre sont des exemples significatifs d'estimation de paramètres et de reconstruction de courbes obtenus avec les modèles déterministes et stochastiques. En particulier, il convient de souligner la meilleure stabilité de l'estimation dans le modèle stochastique. De plus, les résultats présentés ici sont issus d'une base de données de huit patientes, avec de huit à trente-deux voxels par patiente, analysés individuellement.

1.2.2 Deuxième partie : estimation des paramètres d'une diffusion bruitée avec données haute fréquence

Cette partie porte sur l'estimation paramétrique pour un modèle de diffusion bruitée, dans un contexte haute fréquence. Elle se compose de deux chapitres : le premier porte sur l'étude d'un contraste dérivé du schéma d'Euler et sur la consistance de l'estimateur du minimum de contraste qui lui est associé, et le second étudie la normalité asymptotique de cet estimateur. Plusieurs simulations ainsi qu'une étude de données neuronales illustrent les résultats mathématiques.

On considère une diffusion

$$dX_t = b(X_t, \kappa_0)dt + \sigma(X_t, \lambda_0)dB_t, \qquad X_0 = \eta \tag{1.10}$$

et le but est d'estimer $\theta_0 = (\kappa_0, \lambda_0) \in \Theta$, où Θ est un produit d'intervalles compacts, à partir des observations bruitées

$$Y_{i\delta} = X_{i\delta} + \rho\varepsilon_{i\delta}, \quad i = 0, \dots, N \tag{1.11}$$

où $\delta > 0$ est le pas de discrétisation et $(\varepsilon_{i\delta})$ une suite de variables aléatoires centrées indépendantes, identiquement distribuées, et indépendantes de la diffusion (X_t). On suppose que $\mathbb{E}((\varepsilon_{i\delta})^2) = 1$, de sorte que ρ^2 est la variance du bruit d'observation. Lorsque le pas de temps entre deux observations consécutives $\delta = \delta_N$ tend vers 0, le schéma d'Euler pour la diffusion (X_t) fait approcher la distribution de l'accroissement $X_{(i+1)\delta_N} - X_{i\delta_N}$ conditionnellement à $X_{i\delta_N}$ par une variable aléatoire de loi $\mathcal{N}(\delta_N b(X_{i\delta_N}, \kappa_0), \delta_N \sigma^2(X_{i\delta_N}, \lambda_0))$.

Ainsi, lorsque l'on observe directement les $(X_{i\delta_N}, i = 0, \ldots, N)$, on peut considérer le contraste suivant, basé sur la log-vraisemblance gaussienne (voir les articles de Genon-Catalot (1990), Yoshida (1992) et Kessler (1997) par exemple)

$$\mathcal{C}_N(\theta) = \sum_{i=0}^{N-1} \frac{(X_{(i+1)\delta_N} - X_{i\delta_N} - \delta_N b(X_{i\delta_N}, \kappa))^2}{\delta_N c(X_{i\delta_N}, \lambda)} + \log(c(X_{i\delta_N}, \lambda))$$

où $c(x, \lambda) = \sigma(x, \lambda)^2$, et l'estimateur de minimum de contraste $\hat{\theta}_N = (\hat{\kappa}_N, \hat{\lambda}_N)$ associé est consistant lorsque δ_N tend vers 0, avec N (le nombre d'observations) et $N\delta_N$ (le temps maximal d'observation) tendent vers l'infini. De plus, lorsque $N\delta_N^2$ tend vers 0, $\hat{\theta}_N$ est asymptotiquement Gaussien, avec des vitesses d'estimation différentes pour le paramètre de la dérive, et celui du coefficient de diffusion : on a, sous $\mathbb{P}_{\theta_0}$,

$$\begin{pmatrix} \sqrt{N\delta_N}(\hat{\kappa}_N - \kappa_0) \\ \sqrt{N}(\hat{\lambda}_N - \lambda_0) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_2\left(\mathbf{0}, \begin{pmatrix} \nu_0\left(\frac{(\partial_\kappa b(.,\kappa_0))^2}{c(.,\lambda_0)}\right) & 0 \\ 0 & 2\nu_0\left(\frac{(\partial_\lambda c(.,\kappa_0))}{c(.,\lambda_0)^2}\right) \end{pmatrix}\right). \tag{1.12}$$

On se propose dans cette partie d'étendre ces résultats au cas des observations bruitées.

1.2.2.1 Chapitre 4 : Estimation par minimisation de contraste

On cherche alors à construire un contraste basé sur les observations bruitées. Pour cela, on réduit le bruit des observations en effectuant des moyennes locales sur des blocs de données de longueur p_N : on pose $\Delta_N = p_N\delta_N$, de sorte que $N = p_N k_N$ et $N\delta_N = k_N\Delta_N$, et

$$\begin{aligned} Y_\bullet^j = \frac{1}{p_N}\sum_{i=0}^{p_N-1} Y_{j\Delta_N + i\delta_N} &= \frac{1}{p_N}\sum_{i=0}^{p_N-1} X_{j\Delta_N+i\delta_N} + \frac{\rho_N}{p_N}\sum_{i=0}^{p_N-1}\varepsilon_{j\Delta_N+i\delta_N}, \\ &= X_\bullet^j + \rho_N\varepsilon_\bullet^j \end{aligned} \tag{1.13}$$

avec pour ρ_N l'alternative suivante : soit $\rho_N = \rho > 0$ et l'écart-type du bruit d'observation est fixe, soit ρ_N tend vers 0 lorsque N tend vers l'infini. Cette deuxième possibilité s'explique par le cas où $\varepsilon_{i\delta_N} = V_{(i+1)\delta_N} - V_{i\delta_N}$ avec V un mouvement brownien indépendant de (X_t). On note donc $\rho_\infty = \rho$ dans le premier cas, et $\rho_\infty = 0$ dans le second. De plus, on suppose pour la suite que ρ_N est connu.

Ainsi, on est ramené à comparer les variables aléatoires $Y_\bullet^j$ aux k_N discrétisées $X_{j\Delta_N}$ de la diffusion cachée obtenues avec un pas Δ_N. Afin de prendre en compte la contribution apportée par le bruit d'observation (la variance de $\varepsilon_\bullet^j$ est $\frac{1}{p_N}$), on

choisit de considérer des tailles de paquets telles que $\delta_N = p_N^{-\alpha}$ avec $\alpha \in (1,2]$, de sorte que $\Delta_N = p_N^{1-\alpha}$ et $k_N = N\delta_N^{\frac{1}{\alpha}}$.

De ce fait, avec une hypothèse d'ergodicité sur la diffusion cachée, on obtient tout d'abord des résultats asymptotiques pour les fonctionnelles du processus $(Y_\bullet^j)$:

$$\begin{aligned}
\bar{\nu}_N(f(.,\theta)) &= \frac{1}{k_N}\sum_{j=0}^{k_N-1} f(Y_\bullet^{j-1},\theta),\\
\bar{I}_N(f(.,\theta)) &= \frac{1}{k_N\Delta_N}\sum_{j=0}^{k_N-1} f(Y_\bullet^{j-1},\theta)(Y_\bullet^{j+1}-Y_\bullet^j-\Delta_N b(Y_\bullet^{j-1},\kappa)),\\
\bar{Q}_N(f(.,\theta)) &= \frac{1}{k_N\Delta_N}\sum_{j=0}^{k_N-1} f(Y_\bullet^{j-1},\theta)(Y_\bullet^{j+1}-Y_\bullet^j)^2.
\end{aligned}$$

On note dans ces formules le décalage conduisant à utiliser $f(Y_\bullet^{j-1},\theta)$, afin d'éviter la prise en compte de la corrélation entre $Y_\bullet^{j+1}-Y_\bullet^j$ et $Y_\bullet^j$. Ainsi,

$$\begin{aligned}
\bar{\nu}_N(f(.,\theta)) &\longrightarrow \nu_0(f(.,\theta)),\\
\bar{I}_N(f(.,\theta)) &\longrightarrow 0,\\
\bar{Q}_N(f(.,\theta)) &\longrightarrow \frac{2}{3}\nu_0(f(.,\theta)c(.,\lambda_0)) + 2\rho_\infty^2\mathbf{1}_{\alpha=2}
\end{aligned}$$

où toutes les convergences ont lieu en probabilité, uniformément en θ. On note que, dans le cas de la variation quadratique, un terme additionnel apparaît dans le cas $\alpha = 2$ et $\rho_N = \rho$, faisant intervenir la variance du bruit d'observation. De plus, on note la présence du coefficient $\frac{2}{3}$ dans ce cas : il provient de l'étude locale de $Y_\bullet^{j+1}-Y_\bullet^j$, qui diffère de $X_{(j+1)\Delta_N}-X_{j\Delta_N}$ par un terme de corrélation.

On définit alors, pour $\alpha \in (1,2)$, ou lorsque $\alpha = 2$ et ρ_N tend vers 0, le contraste

$$\mathcal{E}_N(\theta) = \sum_{j=1}^{k_N-2}\left\{\frac{3}{2\Delta_N}\frac{(Y_\bullet^{j+1}-Y_\bullet^j-\Delta_N b(Y_\bullet^{j-1},\kappa))^2}{c(Y_\bullet^{j-1},\lambda)} + \log(c(Y_\bullet^{j-1},\lambda))\right\} \quad (1.14)$$

et, pour $\alpha \in (1,2]$ et $\rho_N = \rho$, le contraste

$$\mathcal{E}_N^\rho(\theta) = \sum_{j=1}^{k_N-2}\left\{\frac{3}{2\Delta_N}\frac{(Y_\bullet^{j+1}-Y_\bullet^j-\Delta_N b(Y_\bullet^{j-1},\kappa))^2}{c_{N,\rho}(Y_\bullet^{j-1},\lambda)} + \log(c_{N,\rho}(Y_\bullet^{j-1},\lambda))\right\} \quad (1.15)$$

avec $c_{N,\rho}(x,\lambda) = c(x,\lambda) + 3\Delta_N^{\frac{2-\alpha}{\alpha-1}}\rho^2$. On note dans ces formules le décalage dans $c(Y_\bullet^{j-1},\lambda)$ et $b(Y_\bullet^{j-1},\kappa)$. On dispose alors des estimateurs

$$\hat{\theta}_N = \operatorname*{arginf}_{\theta\in\Theta}\mathcal{E}_N(\theta) \text{ et } \hat{\theta}_N^\rho = \operatorname*{arginf}_{\theta\in\Theta}\mathcal{E}_N^\rho(\theta). \quad (1.16)$$

et l'on montre que ces estimateurs sont consistants. De plus, dans le second cas, lorsqu'on dispose d'un estimateur consistant $\hat{\rho}_N$ de ρ, l'estimateur $\hat{\theta}_N^{\hat{\rho}_N}$ l'est aussi.

Ces résultats sont illustrés sur plusieurs simulations (modèles d'Ornstein-Uhlenbeck, de Cox-Ingersoll-Ross, hyperbolique) avec différents bruits. Une illustration sur un jeu de données issu de l'étude des neurones (voir Idoux et al. (2006) pour une présentation de ces données) est aussi proposée.

1.2.2.2 Chapitre 5 : Fonctions d'estimation et normalité asymptotique des estimateurs

Dans ce chapitre, on considère des fonctions d'estimation pour les observations bruités d'une diffusion discrétisée : étant donnée une fonction

$$G_{N,\alpha}(\theta) = \sum_{j=1}^{k_N-2} g_\alpha(\delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta, \rho_N) \tag{1.17}$$

où $\alpha \in (1,2]$ est le paramètre de taille des paquets, on définit un estimateur $\hat{\theta}_N$ comme solution de l'équation $G_{N,\alpha}(\theta) = 0$, et l'on choisit des fonctions g_α telles que $G_{N,\alpha}$ soit approximativement une martingale (voir Sørensen (2009) pour une étude des fonctions d'estimation pour des diffusions directement observées).

Les estimateurs obtenus par minimum de contraste au Chapitre 4 sont, en particulier, un exemple d'estimateurs associés à une fonction particulière : le gradient du contraste.

La partie principale de ce chapitre est consacrée à l'obtention de plusieurs convergences en loi pour les fonctionnelles des $(Y_\bullet^j)$ déjà étudiées au Chapitre 4. On rappelle que $p_N = \delta_N^{-\alpha}$ et $\Delta_N = p_N^{1-\alpha} = \delta_N^{1-\frac{1}{\alpha}}$. Ainsi, lorsque N tend vers l'infini, avec $N\delta_N$ tendant vers l'infini, δ_N tendant vers 0 et $p_N\delta_N = \Delta_N$ tendant vers 0, on montre que, pour la variation,

$$\sqrt{N\delta_N}\bar{I}_N(f) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu_0(f^2 c(.,\lambda_0)) \tag{1.18}$$

sous l'hypothèse additionnelle : $N\delta_N^{3-\frac{2}{\alpha}}$ tend vers 0.

Pour la variation quadratique, lorsque $\alpha \in (1,2)$ ou $\alpha = 2$ et ρ_N tend vers 0, on a

$$\sqrt{N\delta_N^{\frac{1}{\alpha}}}\left(\bar{Q}_N(f) - \nu_N(f(\frac{2}{3}c + 2\rho_N^2\delta_N^{\frac{1}{\alpha}}))\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu_0(f^2 c(.,\lambda_0)^2) \tag{1.19}$$

lorsque N tend vers l'infini, avec $N\delta_N$ tendant vers l'infini, δ_N tendant vers 0, et sous la condition $N\delta_N^{2-\frac{1}{\alpha}}$ tend vers 0. Il convient de noter que cette condition est plus forte que la condition $N\delta_N^2$ tend vers 0, nécessaire pour obtenir un résultat

similaire dans le cas d'observations non bruitées. De plus, la vitesse $\sqrt{N\delta_N^{\frac{1}{\alpha}}}$ obtenue ici est plus lente que la vitesse $\sqrt{N}$ obtenue pour la variation quadratique des observations sans bruit d'une diffusion discrétisée. En particulier, le choix de la valeur α de la taille des paquets servant à calculer les moyennes locales est important.

Lorsque $\alpha = 2$ et $\rho_N = \rho$, la variance du bruit des observations apparaît dans la variance asymptotique associée à la variation quadratique :

$$\sqrt{N\delta_N^{\frac{1}{2}}}\left(\bar{Q}_N(f) - \nu_N(f(\frac{2}{3}c + 2\rho^2\delta_N^{\frac{1}{2}}))\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu_0(f^2(c(.,\lambda_0)^2 + 4c\rho^2 + 12\rho^4)) \tag{1.20}$$

lorsque N tend vers l'infini, avec $N\delta_N$ tendant vers l'infini, δ_N tendant vers 0, et sous la condition $N\delta_N^{\frac{3}{2}}$ tend vers 0.

Ainsi, l'estimateur $\hat{\theta}_N$ solution de l'équation $G_{N,\alpha}(\theta) = 0$ est consistant et asymptotiquement Gaussien.

Le gradient du contraste $\mathcal{E}_N^\rho(\theta)$ introduit au Chapitre 4 est une fonction d'estimation au sens de (1.17). On déduit alors la normalité asymptotique des estimateurs de minimum de contraste $\hat{\kappa}_N^\rho$ et $\hat{\lambda}_N^\rho$ associés à la dérive et au coefficient de diffusion, sous la condition $N\delta_N^{2-\frac{1}{\alpha}}$ tend vers 0, lorsque $1 < \alpha < 2$,

$$\begin{pmatrix} \sqrt{N\delta_N}(\hat{\kappa}_N - \kappa_0) \\ \sqrt{N\delta_N^{\frac{1}{\alpha}}}(\hat{\lambda}_N - \lambda_0) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \left\{\nu_0\left(\frac{(\partial_\kappa b(.,\kappa_0))^2}{c(.,\lambda_0)}\right)\right\}^{-1} & 0 \\ 0 & \frac{9}{4}\left\{\nu_0\left(\frac{(\partial_\lambda c(.,\lambda_0))^2}{c(.,\lambda_0)^2}\right)\right\}^{-1} \end{pmatrix}\right).$$

On note le coefficient $\frac{9}{4}$ dans la variance asymptotique associée à l'estimation du paramètre du coefficient de diffusion, supérieur au coefficient 2 intervenant lorsqu'on observe une diffusion directement. Lorsque le paramètre α vaut 2 et que $\rho_N = \rho$, la variance augmente encore en faisant intervenir ρ^2.

1.2.3 Troisième partie : étude de la variance dans le théorème central limite pour le filtre particulaire

Cette partie porte sur l'étude de certaines propriétés asymptotiques de la méthode de Monte-Carlo particulaire. Elle a été initialement motivée par la poursuite d'un travail commencé avant la thèse, sur cette méthode. Elle comporte un chapitre présentant les résultats obtenus, ainsi qu'une annexe comprenant quelques exemples éclairant ce travail.

1.2.3.1 Chapitre 6 : Tension de la variance asymptotique

Ce chapitre reprend Favetto (2009), accepté pour publication à ESAIM P&S.

On considère un modèle de Markov caché $(X_n, Y_n), n \in \mathbb{N}$, et l'on appelle $\pi_{n|n:0} = \mathcal{L}(X_n|Y_n, \ldots, Y_0)$ (resp. $\eta_{n|n-1:0} = \mathcal{L}(X_n|Y_{n-1}, \ldots, Y_0)$) la loi du filtre (resp. de la prédiction) associée aux observations $(Y_0, \ldots, Y_n)$. A l'exception de l'exemple historique du filtre de Kalman (Cappé et al. (2005)), et des cas particuliers de filtres *calculables* (Chaleyat-Maurel and Genon-Catalot (2006), Chaleyat-Maurel and Genon-Catalot (2009)), le calcul exact du filtre $\pi_{n|n:0}$ et de la prédiction $\eta_{n|n:0}$ est impossible.

C'est pour cela que la méthode de Monte-Carlo basée sur un algorithme particulaire présente l'intérêt de fournir une bonne approximation $\pi^N_{n|n:0}$ du filtre, et l'on note $\eta^N_{n|n-1:0}$ l'approximation particulaire de la prédiction (voir, par exemple, Del Moral et al. (2001) ou Del Moral and Guionnet (2001)). Ces mesures dépendent des observations $Y_{0:n}$, et d'un entier N : le nombre de particules, qui fournit la précision de l'approximation. Plus précisément, la construction des suites $(\pi^N_{n|n:0})_n$ et $(\eta^N_{n|n-1:0})_n$ repose sur un algorithme faisant intervenir la simulation de N variables aléatoires, appelées particules en interaction. On note $Q(x, dx')$ le noyau de transition de la chaîne cachée (X_n) et $g_n(x) = f(Y_n|x)$ la vraisemblance conditionnelle de l'observation Y_n sachant l'état caché $X_n = x$. Ces quantités sont supposées connues, et de plus, on suppose pouvoir simuler des variables aléatoires selon la loi initiale $\eta_0 = \eta_{0|-1:0}$ de la chaîne cachée ainsi que selon le noyau de transition Q.

Description du filtre particulaire

Etape 0 : Simuler $(X^j_0)_{1\leq j\leq N}$ i.i.d. de loi η_0 et calculer $\eta^N_{0|-1:0} = \frac{1}{N}\sum_{j=1}^N \delta_{X^j_0}$.

Etape 1-a : Simuler X'^j_0 i.i.d. de loi $\pi^N_{0|0:0} = \sum_{j=1}^N \frac{g_0(X^j_0)}{\sum_{j=1}^N g_0(X^j_0)}\delta_{X^j_0}$.

Etape 1-b : Simuler N variables aléatoires $(X^j_1)_j$ indépendantes telles que $X^j_1 \sim Q(X'^j_0, dx)$. Alors $\eta^N_{1|0:0} = \frac{1}{N}\sum_{i=1}^N \delta_{X^j_1}$.

Etape k-a : (mise à jour) On suppose $\eta^N_{k|k-1:0}$ connue. Simuler $(X^j_k)_{1\leq j\leq N}$ i.i.d. de loi $\eta^N_{k|k-1:0}$ et X'^j_k i.i.d. de loi $\pi^N_{k|k:0} = \sum_{j=1}^N \frac{g_k(X^j_k)}{\sum_{j=1}^N g_k(X^j_k)}\delta_{X^j_k}$.

Etape k-b : (prédiction) Simuler X^N_{k+1} indépendantes telles que $X^N_{k+1} \sim Q(X'^j_k, dx)$. Alors $\eta^N_{k+1|k:0} = \frac{1}{N}\sum_{j=1}^N \delta_{X^j_{k+1}}$.

Pour une fonction f bornée – ou de façon plus générale, à croissance polynômiale – et conditionnellement aux observations $Y_n, n \geq 0$, on a le théorème central limite suivant lorsque le nombre de particules N tend vers l'infini (voir, par exemple, Del Moral and Jacod (2001a) ou Chopin (2004))

$$\sqrt{N}(\eta^N_{k|k-1:0}(f) - \eta_{k|k-1:0}(f)) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}(0, \Delta_{k|k-1:0}(f)), \tag{1.21}$$

$$\sqrt{N}(\pi^N_{k|k:0}(f) - \pi_{k|k:0}(f)) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}(0, \Gamma_{k|k:0}(f)). \tag{1.22}$$

Dans Del Moral and Jacod (2001b), les auteurs s'intéressent au comportement de la suite des variances $\Gamma_{k|k:0}(f)$, en tant que suite de variables aléatoires lorsque le conditionnement par rapport à la suite des observations est relaxé, et dans le cas particulier où la chaîne cachée (X_n) est un processus AR(1) Gaussien (issu de la discrétisation d'un processus d'Ornstein-Uhlenbeck) et les observations sont données par $Y_n = X_n + \varepsilon_n$ avec (ε_n) une suite de variables aléatoires Gaussiennes indépendantes et identiquement distribuées.

Le principal résultat de Del Moral and Jacod (2001b) est celui de la tension de la suite des variances asymptotiques $(\Gamma_{k|k:0}(f))_k$, obtenu par des calculs explicites spécifiques au cas Gaussien.

Dans ce chapitre, on montre la tension des suites $(\Gamma_{k|k:0}(f))_k$ et $(\Delta_{k|k-1:0}(f))_k$ sous les hypothèses suivantes :

- la fonction f est bornée,
- le noyau Q de chaîne (X_n) vérifie, pour une probabilité μ et deux réels $\epsilon_- \leq \epsilon_+$

$$\forall x \in \mathcal{X}, \forall B \in \mathcal{B}(\mathcal{X}) \quad \epsilon_-\mu(B) \leq Q(x, B) \leq \epsilon_+\mu(B),$$

- les observations vérifient, pour un $\delta > 0$

$$\sup_{k\geq 0} \mathbf{E} \left|\log\left(\eta_{k|k-1:0}\left(g_k\right)\right)\right|^{1+\delta} < \infty,$$

  où l'espérance est prise par rapport à la distribution des observations.

Ces hypothèses sont discutées sur des exemples, et illustrées par des simulations numériques. En particulier, on construit une chaîne de Markov vérifiant la deuxième hypothèse, avec une mesure μ et des constantes ϵ_-, ϵ_+ explicitement calculables.

On reprend enfin, en appendice de ce chapitre, les calculs du cas Gaussien, pour la prédiction.

Première partie

Estimation paramétrique pour un processus d'Ornstein-Uhlenbeck caché dans un contexte biomédical

Chapitre 2

Parameter estimation for a bidimensional partially observed Ornstein-Uhlenbeck process with biological application

Abstract

We consider a bidimensional Ornstein-Uhlenbeck process to describe the tissue microvascularisation in anti-cancer therapy. Data are discrete, partial and noisy observations of this stochastic differential equation (SDE). Our aim is the estimation of the SDE parameters. We use the main advantage of a one-dimensional observation to obtain an easy way to compute the exact likelihood using the Kalman filter recursion, which allows to implement an easy numerical maximisation of the likelihood. Furthermore, we establish the link between the observations and an ARMA process and we deduce the asymptotic properties of the maximum likelihood estimator. We show that this ARMA property can be generalised to a higher dimensional underlying Ornstein-Uhlenbeck diffusion. We compare this estimator with the one obtained by the well-known EM algorithm on simulated data. Our estimation methods can be directly applied to other biological contexts such as drug pharmacokinetics or hormone secretions.

2.1 Introduction

Stochastic continuous-time models are a useful tool to describe biological or physiological systems based on continuous evolution (see *e.g.* Ditlevsen and De Gaetano (2005), Ditlevsen et al. (2005), Picchini et al. (2006)). The biological context of this work is the modeling of tissue microvascularisation in anti-cancer therapy. This microcirculation is usually modeled by a bidimensional deterministic differential system which describes the circulation of a contrast agent between two compartments (see Brochot et al. (2006) and appendix A.1). However, this deterministic model is unable to capture the random fluctuations observed along time. In this paper, we consider a stochastic version of this system to take into account random variations around the deterministic solution by adding a Brownian motion on each compartment. This leads to a bidimensional stochastic differential equation (SDE) defined as :

$$\left\{ \begin{array}{rcl} dP(t) & = & (\alpha a(t) - (\lambda+\beta)P(t) + (k-\lambda)I(t))dt + \sigma_1 dW_1(t) \\ dI(t) & = & (\lambda P(t) - (k-\lambda)I(t))dt + \sigma_2 dW_2(t) \end{array} \right. \tag{2.1}$$

where $P(t)$ and $I(t)$ represent contrast agent concentrations in each compartment, $a(t)$ is an input function assumed to be known, α, β, λ and k are unknown positive parameters, W_1 and W_2 are two independent Brownian motions on $\mathbb{R}$, and σ_1, σ_2 are the constant diffusion terms. We assume that $(P(0), I(0))$ is a random variable independent of (W_1, W_2). In our biological context, only the sum $S(t) = P(t) + I(t)$ can be measured. So (2.1) is changed into :

$$\left\{ \begin{array}{rcl} dS(t) & = & (\alpha a(t) - \beta S(t) + \beta I(t))\, dt + \sigma_1 dW_1(t) + \sigma_2 dW_2(t) \\ dI(t) & = & (\lambda S(t) - kI(t))\, dt + \sigma_2 dW_2(t) \end{array} \right. \tag{2.2}$$

Noisy and discrete measurements $(y_i, i = 0, \ldots, n)$ of $S(t)$ are performed at times $t_0 = 0 < t_1 < \ldots < t_n = T$. The observation model is thus :

$$y_i = S(t_i) + \sigma\varepsilon_i,\ \varepsilon_i \sim \mathcal{N}(0,1)$$

where $(\varepsilon_i)_{i=0,\ldots,n}$ are assumed to be independent and σ is the unknown standard deviation of the Gaussian noise. To evaluate the effect of the treatment on a patient, it is of importance to have a proper estimation of all unknown parameters from this data set. The aim of this paper is to investigate this problem both theoretically and numerically on simulated data.

Parametric inference for discretely observed general SDEs has been widely investigated. Genon-Catalot and Jacod (1993) and Kessler (1997) propose estimators based on minimization of suitable contrasts and study the asymptotic

distribution of these estimators when the sampling interval tends to zero as the number of observations tends to infinity. For fixed sampling interval, Bibby and Sørensen (1995) propose martingale estimating functions. In a biological context, Ditlevsen et al. (2005) propose an estimation method based on simulation. Picchini et al. (2008) propose estimators based on the Hermite expansion of the transition densities. When combining the case of discrete, partial and noisy observations, parameter estimation is a more delicate statistical problem. In this context, it is classical to estimate the unobserved signal (filtering) (see *e.g.* Cappé et al. (2005)). However, our aim is the estimation of SDE parameters. In this paper, we use the main advantage of a one-dimensional observation y and the Gaussian framework of all distributions to obtain an easy way to compute the exact likelihood. For this, we solve and discretize the SDE (3.3). Then we use the Kalman filter recursion to compute the exact likelihood as proposed by Pedersen (1994) and implemented in the Danish Technical University project CTSM. We also obtain a recursive computation of the exact gradient and hessian of the log-likelihood based on Kalman filtering, which allows us to implement an easy numerical maximisation of the likelihood using a gradient method and to compute the exact maximum likelihood estimator. The exact observed Fisher information matrix is also directly obtained. As our model is a hidden Markov model, we develop a second approach based on the EM algorithm, which is widely used in this context since the so-called complete likelihood (observed, unobserved) is generally explicit whereas the exact likelihood (observed) is generally not explicit. This method has been first proposed by Shumway and Stoffer (1982) and Segal and Weinstein (1989). Segal and Weinstein (1989) claim that the EM algorithm is computationally more efficient that the Kalman filters. Thus we compare the EM algorithm and the Kalman filter approach in our context. These two estimation methods can be directly applied to other biological contexts. For instance, partially observed Ornstein-Uhlenbeck processes are used for modeling drug pharmacokinetics (Ditlevsen et al. (2005)) or detecting pulsatile hormone secretions (Guo et al. (1999)).

The paper is organized as follows. In Section 2.2, we study the SDE. We detail in Section C.3 the computation of the exact likelihood, the score and hessian functions. We present the EM method in Section 2.4. In Section 2.5, we establish the link between the observations and an ARMA process. This allows to deduce the asymptotic properties of the maximum likelihood estimator. Section 2.6 contains numerical results based on simulated data. This allows to compare the two estimation methods. Appendix A.1 describes briefly the biological background. Appendix A.2, A.3, A.4, 3.7.2 and A.6 contain some proofs and auxiliary results. In particular, the ARMA property can be generalised to a higher dimen-

sional underlying Ornstein-Uhlenbeck diffusion.

2.2 Study of the stochastic differential equation

Introducing $U(t) = (S(t), I(t))'$ where $'$ denotes the transposed matrix, (3.3) can be written in a matrix form :

$$\left\{ \begin{array}{rcl} dU(t) & = & (F(t) + G\,U(t))dt + \Sigma dW(t), U(0) = U_0 \\ y_i & = & J\,U(t_i) + \sigma\varepsilon_i \end{array} \right.$$

where $J = (1\ \ 0)$ and

$$F(t) = \begin{pmatrix} \alpha a(t) \\ 0 \end{pmatrix}, G = \begin{pmatrix} -\beta & \beta \\ \lambda & -k \end{pmatrix}, dW(t) = \begin{pmatrix} dW_1(t) \\ dW_2(t) \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_1 & \sigma_2 \\ 0 & \sigma_2 \end{pmatrix}$$

The process $(U(t))$ is a bidimensional Ornstein-Uhlenbeck diffusion, which can be explicitly solved. From the biological context (see Appendix A.1), the parameters satisfy $\beta, k, \lambda > 0$ and $\lambda < k$. This implies that G is diagonalizable with two distinct negative eigenvalues. Setting $d = (\beta - k)^2 + 4\beta\lambda > 0$, the eigenvalues of G are distinct and equal to :

$$\mu_1 = \frac{-(\beta + k) - \sqrt{d}}{2} \quad \text{and } \mu_2 = \frac{-(\beta + k) + \sqrt{d}}{2}$$

The diagonal matrix D of eigenvalues and the matrix P of eigenvectors are :

$$D = \begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix},\ P = \begin{pmatrix} 1 & 1 \\ \frac{\beta - k - \sqrt{d}}{2\beta} & \frac{\beta - k + \sqrt{d}}{2\beta} \end{pmatrix} \text{ with } D = P^{-1}GP.$$

Proposition 2.1 *Let $X(t) = P^{-1}U(t)$ and $\Gamma = (\Gamma^{kj})_{1 \le k,j \le 2} = P^{-1}\Sigma$. Then, for $t, h \ge 0$, we have :*

$$X(t+h) = e^{Dh}X(t) + B(t, t+h) + Z(t, t+h) \tag{2.3}$$

where

$$B(t, t+h) = e^{D(t+h)} \int_t^{t+h} e^{-Ds} P^{-1} F(s) ds \tag{2.4}$$

$$Z(t, t+h) = e^{D(t+h)} \int_t^{t+h} e^{-Ds} \Gamma dW_s. \tag{2.5}$$

Therefore, the conditional distribution of $X(t+h)$ given $X(s), s \le t$ is

$$\mathcal{N}_2\left(e^{Dh}X(t) + B\left(t, t+h\right), R\left(t, t+h\right)\right)$$

where

$$R(t,t+h) = \left(\frac{e^{(\mu_k+\mu_l)h}-1}{\mu_k+\mu_l}\left(\Gamma\Gamma'\right)^{kl}\right)_{1\leq k,l\leq 2} \tag{2.6}$$

If $a(t) \equiv c \geq 0$ with c a constant, $(X(t))$ has a Gaussian stationary distribution with mean equal to

$$M = -D^{-1}P^{-1}F$$

and covariance matrix equal to

$$V = \left(\frac{1}{-(\mu_k+\mu_l)}\left(\Gamma\Gamma'\right)^{kl}\right)_{1\leq k,l\leq 2}$$

Proof. See Appendix A.2.

2.3 Parameter estimation by maximum likelihood

Our aim is to estimate the unknown parameters $\alpha, \beta, \lambda, k, \sigma_1, \sigma_2$ and σ from observations $y_{0:n} = (y_0, \ldots, y_n)$. As the law of $((X(t)), \varepsilon_i, i = 0, \ldots, n)$ is Gaussian, the likelihood of $y_{0:n}$ can be explicitly evaluated. However, the direct maximization of this likelihood requires the inversion of a matrix of dimension $2(n+1)\times 2(n+1)$ (the covariance matrix of $(X(t_i))$). This inversion can be numerically instable. In this section, we present an alternative method for the computation of the exact likelihood based on Kalman filtering, which does not require any matrix inversion. This is due to the fact that data are one-dimensional. Moreover, it is worth stressing that we need not come back to the initial process $(U(t))$ for computing the likelihood. Indeed, as $(U(t))$ is not observed, we can use either $(U(t))$ or any other transformation of $(U(t))$ even involving unknown parameters. As $(X(t))$ is simpler, we consider the following transformed model :

$$\left\{\begin{array}{rcl} dX(t) & = & (DX(t) + P^{-1}F(t))dt + \Gamma dW_t,\ X(0) = P^{-1}U_0 = X_0 \\ y_i & = & J\,PX(t_i) + \sigma\varepsilon_i \end{array}\right. \tag{2.7}$$

Given the particular form of our vector $J = (1\ 0)$ and the fact that the eigenvectors can be chosen up to a proportionnality constant, we have

$$H = JP = (1\ 1).$$

It is especially interesting for further computations of the gradient and hessian of the likelihood that H does not depend of any unknown parameter. From model (2.7) and (2.3)-(2.6), we deduce the following discrete-time evolution system

where $X_i = X(t_i)$:

$$\left\{ \begin{array}{rcl} X_i & = & A_i X_{i-1} + B_i + \eta_i, \quad \eta_i \sim \mathcal{N}(0, R_i) \\ y_i & = & H X_i + \sigma \varepsilon_i \end{array} \right. \tag{2.8}$$

where $A_i = \exp(D(t_i - t_{i-1}))$, $B_i = B(t_{i-1}, t_i)$, $R_i = R(t_{i-1}, t_i)$.

2.3.1 Computation of the exact likelihood

This discrete model is a hidden Markov model (HMM) (Cappé et al., 2005) : (X_i) is a hidden Markov chain on $\mathbb{R}^2$ and, conditionally on (X_i), observations (y_i) are independent. Genon-Catalot and Laredo (2006) study the maximum likelihood estimator for general HMM. They specialize the exact likelihood in the case where the unobserved Markov chain is a Gaussian one-dimensional AR(1) process. We generalize this computation to the case where the unobserved Markov chain is a bidimensional AR(1) process. Let ϕ denote the vector of unknown parameters and $y_{0:i} = (y_0, \ldots, y_i)$ the vector of observations until time t_i. By recursive conditioning, it is sufficient to compute the distribution of y_i given $y_{0:i-1}$:

$$L(\phi, y_{0:n}) = p(y_0; \phi) \prod_{i=1}^{n} p(y_i | y_{0:i-1}; \phi).$$

But the conditional law of y_i given $y_{0:i-1}$ can be evaluated by

$$p(y_i | y_{0:i-1}; \phi) = \int p(y_i | X_i; \phi) p(X_i | y_{0:i-1}; \phi) dX_i$$

Then, as the innovation noise η_i of the hidden Markov chain, and the observation noise ε_i are Gaussian variables, by elementary computations on Gaussian laws, we are able to get the law of y_i given $y_{0:i-1}$ if we know the mean and covariance of the Gaussian conditional law of X_i given $y_{0:i-1}$. This conditional distribution can be exactly computed using Kalman recursions as proposed by Pedersen (1994) and implemented in the Danish Technical University project CTSM. This computation is described below.

2.3.1.1 Kalman filter

To ease the reading, the parameter ϕ is omitted. The Kalman filter is an iterative procedure which computes recursively the following conditional distributions

$$\begin{array}{rcll} \mathcal{L}(X_i | y_{0:i-1}) & = & \mathcal{N}_2(\hat{X}_i^-, P_i^-) & (prediction) \\ \mathcal{L}(X_i | y_{0:i}) & = & \mathcal{N}_2(\hat{X}_i, P_i) & (filter) \end{array}$$

where

$$\hat{X}_i^- = \mathbb{E}(X_i|y_{0:i-1}) \quad \text{and} \quad P_i^- = \mathbb{E}((X_i - \hat{X}_i^-)(X_i - \hat{X}_i^-)')$$
$$\hat{X}_i = \mathbb{E}(X_i|y_{0:i}) \quad \text{and} \quad P_i = \mathbb{E}((X_i - \hat{X}_i)(X_i - \hat{X}_i)')$$

Let us assume that the law of X_0 is Gaussian. Initial values for the algorithm are :

$$X_0 \sim \mathcal{N}(\hat{X}_0^-, P_0^-)$$

with $\hat{X}_0^- = 0$, $P_0^- = 0$ (from theoretical point of view, one can choose the stationary distribution $\hat{X}_0^- = M$, $P_0^- = V$ without changes). Next we have the recursive formulae obtained using (2.8) :

$$\begin{aligned} \hat{X}_i^- &= A_i\hat{X}_{i-1} + B_i, \quad P_i^- = A_iP_{i-1}A_i' + R_i, \quad i \geq 1 \\ \hat{X}_i &= \hat{X}_i^- + K_i(y_i - H\hat{X}_i^-), \quad P_i = (I - K_iH)P_i^-, \quad i \geq 0 \end{aligned} \tag{2.9}$$

where $K_i = P_i^-H'(HP_i^-H' + \sigma^2)^{-1}$ (see *e.g.* Cappé et al. (2005)).

2.3.1.2 Computation of the exact likelihood of the observations

The conditional distribution of y_i given $y_{0:i-1}$ is Gaussian and one-dimensional. Let $m_i(\phi) = \mathbb{E}_\phi(y_i|y_{0:i-1})$ and $V_i(\phi) = Var_\phi(y_i|y_{0:i-1})$ denote its mean and variance which are given using (2.8) by

$$m_i(\phi) = H\hat{X}_i^-, \quad V_i(\phi) = HP_i^-H' + \sigma^2$$

where $\hat{X}_i^-$ and P_i^- depend on ϕ. The exact likelihood of $y_{0:n}$ is thus equal to

$$L(\phi, y_{0:n}) = \prod_{i=0}^{n} \frac{1}{\sqrt{2\pi V_i(\phi)}} \exp\left(-\frac{1}{2}\frac{(y_i - m_i(\phi))^2}{V_i(\phi)}\right). \tag{2.10}$$

Relations (2.9) imply that there exist two functions F_ϕ and G_ϕ such that

$$m_i(\phi) = F_\phi(m_{i-1}(\phi)), \; V_i(\phi) = G_\phi(V_{i-1}(\phi)) \tag{2.11}$$

These iterative relations are used to compute the derivatives of the log-likelihood.

2.3.2 Computation of the maximum likelihood estimator

Pedersen (1994) and the Danish Technical University project CTSM propose to approximate the maximum likelihood estimator (MLE) using a quasi-Newton maximisation method based on the approximation of the gradient and the hessian of the log-likelihood. In our model, we show that it is possible to compute the exact MLE. We use a conjugate gradient method, which relies on the explicit knowledge of the gradient and hessian of the log-likelihood. Both can be exactly computed using formula (2.10) and observing that the derivatives of $m_i(\phi)$ and $V_i(\phi)$ can be explicitly and recursively computed by derivating formulae (2.11).

2.3.2.1 New parametrization

In order to simplify the derivatives of (2.11), from now on, we assume that observation times are equally spaced and set

$$\Delta = t_i - t_{i-1},\ \forall i = 1, \dots, n.$$

Hence we have $A_i = A$, $R_i = R$. For the sake of simplicity we set $a(t) \equiv c \geq 0$, where c is a known constant, corresponding to an intravenous injection in our biological framework. Hence we have $B_i = B = -(I - A)D^{-1}P^{-1}F = (I - A)M$. Set $Z_i = X_i - M$ and $m = HM$. Therefore, model (2.8) becomes

$$\left\{ \begin{array}{rcl} Z_i & = & AZ_{i-1} + \eta_i,\ \ \eta_i \sim \mathcal{N}(0, R) \\ y_i & = & HZ_i + m + \sigma\varepsilon_i \end{array} \right. \tag{2.12}$$

The exact likelihood (2.10) of $y_{0:n}$ is thus equal to

$$L(\phi, y_{0:n}) = \prod_{i=0}^{n} \frac{1}{\sqrt{2\pi V_i(\phi)}} \exp\left(-\frac{1}{2} \frac{(y_i - m - H\hat{Z}_i^-(\phi))^2}{V_i(\phi)} \right). \tag{2.13}$$

Moreover, instead of $\phi = (\alpha, \beta, \lambda, k, \sigma_1, \sigma_2, \sigma^2)$, we propose a new parametrization fitted with the discretization. We consider $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)$ where $\theta_i = e^{\mu_i \Delta}$, $i = 1, 2$ and θ_3, θ_4 and θ_5 are explicit functions of $\mu_1, \mu_2, \sigma_1, \sigma_2$ and Δ such that

$$A = A(\theta) = \begin{pmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{pmatrix} \quad \text{and} \quad R = R(\theta) = \begin{pmatrix} \theta_3 & \theta_5 \\ \theta_5 & \theta_4 \end{pmatrix}$$

and $\theta_6 = m$. We set $\vartheta = (\theta, \sigma^2)$. Our aim is to maximize the likelihood $L(\vartheta, y_{0:n})$ with respect to ϑ. Given an estimation $\hat{\vartheta}$, $\hat{\phi}$ can be obtained by solving numerically the equation $f(\hat{\phi}) = \hat{\vartheta}$ where f is the mapping $\phi \to f(\phi) = \vartheta$ (see Appendix A.3 for details). Note that, later on, we will see that only six out of the seven parameters can be consistently estimated.

2.3.2.2 Computation of the exact gradient and hessian of the log-likelihood

Let $W_i(\vartheta) = y_i - \vartheta_6 - H\hat{Z}_i^-(\vartheta)$ and $l_{0:i}(\vartheta) = \log L(\vartheta, y_{0:i})$. Using (2.10), it comes :

$$l_{0:i}(\vartheta) = l_{0:i-1}(\vartheta) - \frac{1}{2}\log(2\pi V_i(\vartheta)) - \frac{1}{2}\frac{W_i(\vartheta)^2}{V_i(\vartheta)}. \tag{2.14}$$

Thus for $i = 1, \dots, n$, $q = 1, \dots, 7$:

$$\frac{\partial l_{0:i}}{\partial \vartheta_q}(\vartheta) = \frac{\partial l_{0:i-1}}{\partial \vartheta_q}(\vartheta) - \frac{1}{2}\frac{1}{V_i(\vartheta)}\frac{\partial V_i}{\partial \vartheta_q}(\vartheta) - \frac{W_i(\vartheta)}{V_i(\vartheta)}\frac{\partial W_i(\vartheta)}{\partial \vartheta_q} + \frac{1}{2}\frac{W_i(\vartheta)^2}{V_i(\vartheta)^2}\frac{\partial V_i(\vartheta)}{\partial \vartheta_q} \tag{2.15}$$

where

$$\frac{\partial V_i(\vartheta)}{\partial \vartheta_q} = H\frac{\partial P_i^-(\vartheta)}{\partial \vartheta_q}H', 1 \leq q \leq 6, \ \frac{\partial V_i(\vartheta)}{\partial \sigma^2} = H\frac{\partial P_i^-(\vartheta)}{\partial \sigma^2}H' + 1$$
$$\frac{\partial W_i(\vartheta)}{\partial \vartheta_q} = -\frac{\partial m}{\partial \vartheta_q} - H\frac{\partial \hat{X}_i^-(\vartheta)}{\partial \vartheta_q}, 1 \leq q \leq 7$$

Furthermore, the derivatives of $\hat{X}_i^-(\vartheta)$ and $P_i^-(\vartheta)$ can be obtained using Kalman recursions (see appendix A.4). With a more cumbersome computation, second order derivatives of $l_{0:n}(\vartheta)$ can be analogously deduced from (2.15). For $i = 1, \ldots, n$, $q, r = 1, \ldots, 7$,

$$\begin{aligned}\frac{\partial^2 l_{0:i}}{\partial \vartheta_r \partial \vartheta_q}(\vartheta) &= \frac{\partial^2 l_{0:i-1}}{\partial \vartheta_r \partial \vartheta_q}(\vartheta) - \frac{1}{2}\frac{1}{V_i(\vartheta)}\frac{\partial^2 V_i}{\partial \vartheta_r \partial \vartheta_q}(\vartheta) + \frac{1}{2}\frac{1}{V_i^2(\vartheta)}\frac{\partial V_i}{\partial \vartheta_r}(\vartheta)\frac{\partial V_i}{\partial \vartheta_q}(\vartheta) \\ &\quad -\frac{1}{2}\left(2\frac{W_i(\vartheta)}{V_i(\vartheta)}\frac{\partial^2 W_i(\vartheta)}{\partial \vartheta_r \partial \vartheta_q} - \frac{W_i(\vartheta)^2}{V_i(\vartheta)^2}\frac{\partial^2 V_i(\vartheta)}{\partial \vartheta_r \partial \vartheta_q}\right) \\ &\quad -\left(\frac{\partial W_i(\vartheta)}{\partial \vartheta_r}\frac{\partial W_j(\vartheta)}{\partial \vartheta_q}\frac{1}{V_i(\vartheta)} - \frac{W_i(\vartheta)}{V_i(\vartheta)^2}\frac{\partial W_i(\vartheta)}{\partial \vartheta_r}\frac{\partial V_i(\vartheta)}{\partial \vartheta_q}\right) \\ &\quad +\left(\frac{\partial W_i(\vartheta)}{\partial \vartheta_q}\frac{\partial V_i(\vartheta)}{\partial \vartheta_r}\frac{W_i(\vartheta)}{V_i(\vartheta)^2} - \frac{W_i(\vartheta)^2}{V_i(\vartheta)^4}\frac{\partial V_i(\vartheta)}{\partial \vartheta_r}V_i(\vartheta)\frac{\partial V_i(\vartheta)}{\partial \vartheta_q}\right)\end{aligned} \tag{2.16}$$

where (see appendix A.4 for details)

$$\frac{\partial^2 V_i(\vartheta)}{\partial \vartheta_r \partial \vartheta_q} = H\frac{\partial^2 P_i^-(\vartheta)}{\partial \vartheta_r \partial \vartheta_q}H' \text{ and } \frac{\partial^2 W_i(\vartheta)}{\partial \vartheta_r \partial \vartheta_q} = -H\frac{\partial^2 \hat{X}_i^-(\vartheta)}{\partial \vartheta_r \partial \vartheta_q}.$$

Hence, we obtain an explicit expression of $\left(-\frac{\partial^2 l_{0:n}}{\partial \vartheta_r \partial \vartheta_q}(\vartheta)\right)_{1 \leq q,r \leq 7}$.

2.3.2.3 Maximisation of the exact likelihood

To compute the maximum likelihood estimator, the conjugate gradient algorithm is applied to minimize $l_{0:n}^-(\vartheta) = -l_{0:n}(\vartheta)$ (see Stoer and Bulirsch (1993)). Let ∇l^- denote the gradient of $l_{0:n}^-$ and $Hess\, l^-$ its hessian evaluated by (2.15)-(2.16). Starting with an arbitrary initial vector ϑ_0, we set as descent direction $u_0 = \vartheta_0$. At iteration k, given ϑ_k and u_k, the parameter and descent direction are updated by

$$\vartheta_{k+1} = \vartheta_k - \frac{\langle u_k, \nabla l^-(\vartheta_k)\rangle}{\langle u_k, Hess\, l^-(\vartheta_k)u_k\rangle}u_k, \ u_{k+1} = -\nabla l^-(\vartheta_{k+1}) + \frac{\|\nabla l^-(\vartheta_{k+1})\|}{\|\nabla l^-(\vartheta_k)\|}u_k.$$

Classical stopping conditions are used. The sequence $(\vartheta_k)_k$ converges towards the maximum of the likelihood $l_{0:n}(\vartheta)$.

2.4 Parameter estimation by Expectation Maximization algorithm

An alternative method to estimate $\vartheta = (\theta, \sigma^2)$ is the Expectation Maximization (EM) algorithm, proposed by Dempster et al. (1977), see also Shumway and Stoffer (1982) and Segal and Weinstein (1989). The EM algorithm is a classical approach to estimate parameters of models with non-observed or incomplete data, especially it is widely used for HMMs. In our case, the non-observed data are the (Z_i)'s, the complete data are the (y_i, Z_i)'s. The principle is to maximize

$$\vartheta \to Q(\vartheta|\vartheta_*) = \mathbb{E}(\log p(y_{0:n}, Z_{0:n}; \vartheta)|y_{0:n}; \vartheta_*)$$

with $Z_{0:n} = (Z_0, \ldots, Z_n)$. This is often easier than the maximization of the observed data log-likelihood since the log-likelihood of the complete data is generally simpler. Moreover, according to Wu (1983), as our model is an exponential family, the EM estimate sequence $(\vartheta_k)_k$ converges towards a (local) maximum of the data likelihood. The EM algorithm uses two steps : the Expectation step (E-step) and the Maximization step (M-step). Starting with an initial value (ϑ_0), the k-th iteration is

- E-step : evaluation of $Q_k(\vartheta) = Q(\vartheta \,|\, \vartheta_k)$
- M-step : update of ϑ_k by $\vartheta_{k+1} = \arg\max Q_k(\vartheta)$.

In our model, function Q has an explicit expression. Recall that we have assumed that observations times are equally spaced and that $a(t) \equiv c$. For simplicity, we also set for the initial variable $Z_0 = 0$ ($\hat{Z}_0^- = 0$ and $P_0^- = 0$). The complete data log-likelihood is thus equal to :

$$\log p(y_{0:n}, Z_{0:n}; \vartheta) = -\frac{n+1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=0}^{n}(y_i - \vartheta_6 - HZ_i)^2$$
$$-\frac{1}{2}\sum_{i=1}^{n}\log(2\pi|R(\vartheta)|) - \frac{1}{2}\sum_{i=1}^{n}(Z_i - A(\vartheta)Z_{i-1})'R(\vartheta)^{-1}(Z_i - A(\vartheta)Z_{i-1}).$$

Function Q consists in taking the conditional expectation given $y_{0:n}$ under $\mathbb{P}_{\vartheta_*}$. This conditional distribution is the so-called smoothing distribution at ϑ_*. In our model, it is Gaussian and characterized by $M_{i|0:n}(\vartheta_*) = \mathbb{E}_{\vartheta_*}(Z_i|y_{0:n})$ and

$$\Sigma_{i|0:n}(\vartheta_*) = Var_{\vartheta_*}(Z_i|y_{0:n}),\ \Sigma_{i-1,i|0:n}(\vartheta_*) = Cov_{\vartheta_*}(Z_{i-1}, Z_i|y_{0:n})$$

These can be obtained through a forward-backward algorithm (see Appendix 3.7.2). Thus function Q is equal to :

$$Q(\vartheta|\vartheta_*) = -\frac{n+1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=0}^{n}\left[(y_i - \vartheta_6 - HM_{i|0:n}(\vartheta_*))^2 + H\Sigma_{i|0:n}(\vartheta_*)H'\right]$$
$$-\frac{n}{2}\log(2\pi|R(\vartheta)|) - \frac{1}{2}Tr\left\{R(\vartheta)^{-1}\left[C(\vartheta_*) - T(\vartheta_*)A'(\vartheta) - A(\vartheta)T'(\vartheta_*) + A(\vartheta)S(\vartheta_*)A'(\vartheta)\right]\right\}$$

where

$$\begin{aligned}
T(\vartheta_*) &= \sum_{i=1}^{n}\left(\Sigma_{i-1,i|0:n}(\vartheta_*) + M_{i|0:n}(\vartheta_*)M'_{i-1|0:n}(\vartheta_*)\right)\\
S(\vartheta_*) &= \sum_{i=1}^{n}\left(\Sigma_{i-1|0:n}(\vartheta_*) + M_{i-1|0:n}(\vartheta_*)M'_{i-1|0:n}(\vartheta_*)\right)\\
C(\vartheta_*) &= \sum_{i=1}^{n}\left(\Sigma_{i|0:n}(\vartheta_*) + M_{i|0:n}(\vartheta_*)M'_{i|0:n}(\vartheta_*)\right).
\end{aligned}$$

The matrices A, R, θ_6 and σ^2 are updated as

$$\begin{aligned}
A(\vartheta_k) &= diag(T(\vartheta_{k-1})S^{-1}(\vartheta_{k-1}))\\
R(\vartheta_k) &= \frac{1}{n}(C(\vartheta_{k-1}) - T(\vartheta_{k-1})S^{-1}(\vartheta_{k-1})T'(\vartheta_{k-1}))\\
\vartheta_{6k} &= \frac{1}{n+1}\sum_{i=0}^{n}(y_i - HM_{i|0:n}(\vartheta_{k-1}))\\
\sigma_k^2 &= \frac{1}{n+1}\sum_{i=0}^{n}\left[(y_i - \vartheta_{6k-1} - HM_{i|0:n}(\vartheta_{k-1}))^2 + H\Sigma_{i|0:n}(\vartheta_{k-1})H'\right]
\end{aligned}$$

2.5 Properties of the exact maximum likelihood estimator in the stationary case

Recall that we have assumed that $a(t) \equiv c$. Moreover in this paragraph we assume that the initial variable X_0 has the stationary distribution $\mathcal{N}_2(M, V)$ given in Proposition 2.1. This implies that the joint process (X_i, y_i) is strictly stationary. Let $(y_i)_{i\in\mathbb{Z}}$ be its extension to a process indexed by $\mathbb{Z}$.

2.5.1 Link with an ARMA model

We generalize the result of Genon-Catalot et al. (2003) to the bidimensional case and also to the multidimensional case (see Appendix A.6).

Proposition 2.2 *Let $(\tilde{y}_i) = (y_i - \theta_6)$ define the centered process. The process $(\tilde{y}_i)_{i\in\mathbb{Z}}$ is centered Gaussian and ARMA(2,2).*

Proof. Evidently $(\tilde{y}_i)$ is centered Gaussian. We easily check that

$$\tilde{y}_i - (\theta_1+\theta_2)\tilde{y}_{i-1} + \theta_1\theta_2\tilde{y}_{i-2} = \xi_i \tag{2.17}$$

where ξ_i is defined by

$$\xi_i = HA(\theta)\eta_{i-1} + H\eta_i + \sigma\varepsilon_i - (\theta_1+\theta_2)H\eta_{i-1} - (\theta_1+\theta_2)\sigma\varepsilon_{i-1} + \theta_1\theta_2\sigma\varepsilon_{i-2}$$

As the $(\eta_i)_i$ and $(\varepsilon_i)_i$ are mutually independent, we get that :

$$Cov(\xi_i, \xi_{i+k}) = 0, \ \forall k \geq 3.$$

This implies that (ξ_i) is MA(2). Hence the result. □

Proposition 2.3 *The spectral density $f(u,\vartheta)$ of $(\tilde{y}_i)$ has the explicit form :*

$$f(u,\vartheta) = \sigma^2 + \frac{H(ARA' + (1+(\theta_1+\theta_2)^2)R)H' + 2\cos(u)(HA - (\theta_1+\theta_2)H)RH'}{1+(\theta_1+\theta_2)^2+\theta_1^2\theta_2^2 - 2(\theta_1+\theta_2)(1+\theta_1\theta_2)\cos(u) + 2\cos(2u)\theta_1\theta_2} \tag{2.18}$$

with $A = A(\theta), R = R(\theta)$.

Proof. Let $\gamma(k) = Cov(\xi_i, \xi_{i+k})$. Elementary computations show that

$$\begin{aligned}
\gamma(0) &= HARA'H' + HRH'(1+(\theta_1+\theta_2)^2) + \sigma^2(1+(\theta_1+\theta_2)^2+\theta_1^2\theta_2^2) \\
\gamma(1) &= (HA - (\theta_1+\theta_2)H)RH' - \sigma^2(\theta_1+\theta_2)(1+\theta_1\theta_2) \\
\gamma(2) &= \sigma^2\theta_1\theta_2 \\
\gamma(k) &= 0 \ \forall\, k \geq 3
\end{aligned}$$

The spectral density (with respect to $\frac{du}{2\pi}$) $h(u,\vartheta)$ of (ξ_i) is

$$h(u,\vartheta) = \sum_{n\in\mathbb{Z}} \gamma(n)\exp(-inu) = \gamma(0) + \gamma(1)2\cos(u) + \gamma(2)2\cos(2u).$$

For the AR(2) part, we set : $p(x) = 1 - (\theta_1+\theta_2)x + \theta_1\theta_2x^2$ (recall that $\theta_1, \theta_2 < 1$). Then

$$\begin{aligned}
f(u,\vartheta) &= \frac{h(u,\vartheta)}{|p(\exp(-iu))|^2} \\
&= \frac{\gamma(0) + \gamma(1)2\cos(u) + \gamma(2)2\cos(2u)}{1+(\theta_1+\theta_2)^2+\theta_1^2\theta_2^2 - 2(\theta_1+\theta_2)(1+\theta_1\theta_2)\cos(u) + 2\cos(2u)\theta_1\theta_2}
\end{aligned}$$

See Brockwell and Davis (1991) for technical details. □
The number of parameters which are identifiable on the spectral density is precised by the following proposition

Proposition 2.4 *The identifiable quantities are σ^2, $\theta_1+\theta_2$ and $\theta_1\theta_2$, and at most two out of three parameters among θ_3, θ_4 and θ_5.*

When Δ is small, exactly two out of the three paramaters θ_3, θ_4 and θ_5 are identifiable.

Proof. See Appendix A.7.

2.5.2 Asymptotic properties of maximum likelihood estimator

We now deduce the asymptotic properties of the maximum likelihood estimators when (y_i) is stationary.

We denote by ϑ^0 the true value of parameters vector $(\theta_1, \ldots, \theta_6, \sigma^2)$. We assume that the parameter set Θ is an open subset of $\mathbb{R}^7$. We denote by $\vartheta_- = (\theta_1, \ldots, \theta_5, \sigma^2) \in \Theta_-$ the projection of ϑ on $\mathbb{R}^6$.

The following result is classical to estimate the parameter ϑ_6^0 (see *e.g.* in Brockwell and Davis (1991)).

Proposition 2.5 (Mean estimator) *Let $\bar{y}_n = \frac{1}{n}\sum_{i=1}^n y_i$ be the empirical mean. Under the assumption of stationarity of (y_i), $\bar{y}_n \to \theta_6^0$ a.s. as $n \to \infty$. Moreover, $\sqrt{n}(\bar{y}_n - \theta_6^0)$ converges in distribution :*

$$\sqrt{n}(\bar{y}_n - \theta_6^0) \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0, J(\vartheta_-^0))$$

where $J(\vartheta_-^0) = f(0, \vartheta_-^0) = \frac{\gamma(0)+2\gamma(1)+2\gamma(2)}{((1-\theta_1)(1-\theta_2))^2}$

We now consider the centered process $(\tilde{y}_i)$. Consider the two assumptions (which can be checked up to some technicities)

A1 $(u, \vartheta_-) \longmapsto f(u, \vartheta_-)$ is a $\mathcal{C}^3$-function on a neighborhood of $[-\pi, \pi] \times \Theta_-$

A2 $\vartheta_- \longmapsto f(., \vartheta_-)$ is one to one

As $(\tilde{y}_i)$ is a ARMA(2,2) process, its spectral density is positive for every $(u, \vartheta_-) \in [-\pi, \pi] \times \Theta_-$.

Proposition 2.6 (Information matrix) *Let $\tilde{l}_{0:n}(\vartheta_-) = \log L(\vartheta_-, \tilde{y}_{0:n})$ be the log-likelihood of the centered process $(\tilde{y}_{0:n})$. Under the assumption A1, we have $\mathbb{P}_{\vartheta_-^0}$-a.s.*

$$\lim_{n\to\infty} \left(-\frac{1}{n} \frac{\partial^2}{\partial\theta_i \partial\theta_j} \tilde{l}_{0:n}\left(\vartheta_-^0\right) \right)_{1\le i,j\le 6} = I(\vartheta_-^0) \tag{2.19}$$

Proposition 2.7 (Consistency and asymptotic normality of the MLE) *Let $\hat{\theta}_n$ be a maximum likelihood estimator of ϑ_-^0 based on $\tilde{y}_{0:n}$. Under the assumptions A1*

and A2, $\hat{\theta}_n \to \vartheta^0_-$ *a.s. as* $n \to \infty$*. Moreover, if* $I(\vartheta^0_-)$ *is invertible,* $\sqrt{n}(\hat{\theta}_n - \vartheta^0_-)$ *converges in distribution :*

$$\sqrt{n}(\hat{\theta}_n - \vartheta^0_-) \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0, I^{-1}(\vartheta^0_-))$$

Proof. This result may be found *e.g.* in Brockwell and Davis (1991). □

Remarque – As it is done usually, for further numerical considerations, the empirical mean estimator $\bar{y}_n$ is plugged in the likelihood for parameter θ_6 in the Kalman-recursion approach. The EM algorithm estimates this parameter during the algorithm iterations.

2.6 Simulation study

We compare the performances of the two estimation methods on simulated data sets. The exact maximum likelihood estimators and the EM estimators are computed as described in Section C.3 and Section 2.4, respectively. Data are simulated using equally spaced observation times ($\Delta = 0.2$) and $n = 200$ or $n = 1000$ observations. Values of parameter θ are deduced from values of biological parameters $(\alpha, \beta, \lambda, k)$ estimated on real data in Thomassin (2008), and $\sigma_1^2 = 0.5$, $\sigma_2^2 = 0.125$, $c = 50$:

$$\theta_1 = 0.6,\ \theta_2 = 0.9,\ \theta_3 = 0.7,\ \theta_4 = 0.2,\ \theta_5 = 0.1, \theta_6 = 20$$

Two levels of observation noise are used : $\sigma^2 = 1$ or $\sigma^2 = 3$. Thousand replications are performed for each design ($n = 200$ or $n = 1000$ observations, $\sigma^2 = 1$ or $\sigma^2 = 3$). The influence of the time scale Δ is evaluated on simulated data with $\Delta = 0.04$ and $n = 1000$ observations with parameter values equal to

$$\theta_1 = 0.91,\ \theta_2 = 0.99,\ \theta_3 = 0.2,\ \theta_4 = 0.03,\ \theta_5 = 0.01, \theta_6 = 20$$

Thousand replications are performed for each observation noise ($\sigma^2 = 1$ or $\sigma^2 = 3$) in this case. Mean estimates and their empirical standard errors are computed on the 1000 replications of each design. The exact standard errors obtained from the asymptotic information matrix (computed by Kalman-based recursions) are also provided.

Identifiability results of Section 2.5 show that parameters θ_6 and σ^2 are identifiable, and that at most 4 among the five parameters $(\theta_1, \ldots, \theta_5)$ are identifiable. In our biological context, it is reasonnable to assume that σ^2 is known. Therefore σ^2 is fixed to its true value. One parameter among $(\theta_1, \ldots, \theta_5)$ has to be fixed :

we choose to fix θ_5 to its true value and to estimate $\theta_1, \theta_2, \theta_3, \theta_4$. For the MLE estimation, parameter θ_6 is previously estimated by the empirical mean. Then the other parameters are estimated based on the empirically centered observations. Results are given in Table 2.1 for $\Delta = 0.2$ and Table 2.2 for $\Delta = 0.2$ and $\Delta = 0.04$. Results obtained on one simulated data set ($n = 200$, $\Delta = 0.2$, $\sigma^2 = 3$) are plotted in Figure 2.1 and show the ability of the method to estimate the trajectories $(S(t), P(t), I(t))$. The EM algorithm is around three times quicker than the MLE algorithm. The mean computation time to estimate parameters of a data set with $n = 200$ observations is around 4 seconds (CPU time) for the EM algorithm and around 13 seconds (CPU time) for the MLE for the same precision of convergence, on an Apple MacPro 2×3 Ghz with 5 Go of RAM. The Matlab code are given at http ://www.mi.parisdescartes.fr/˜favetto.

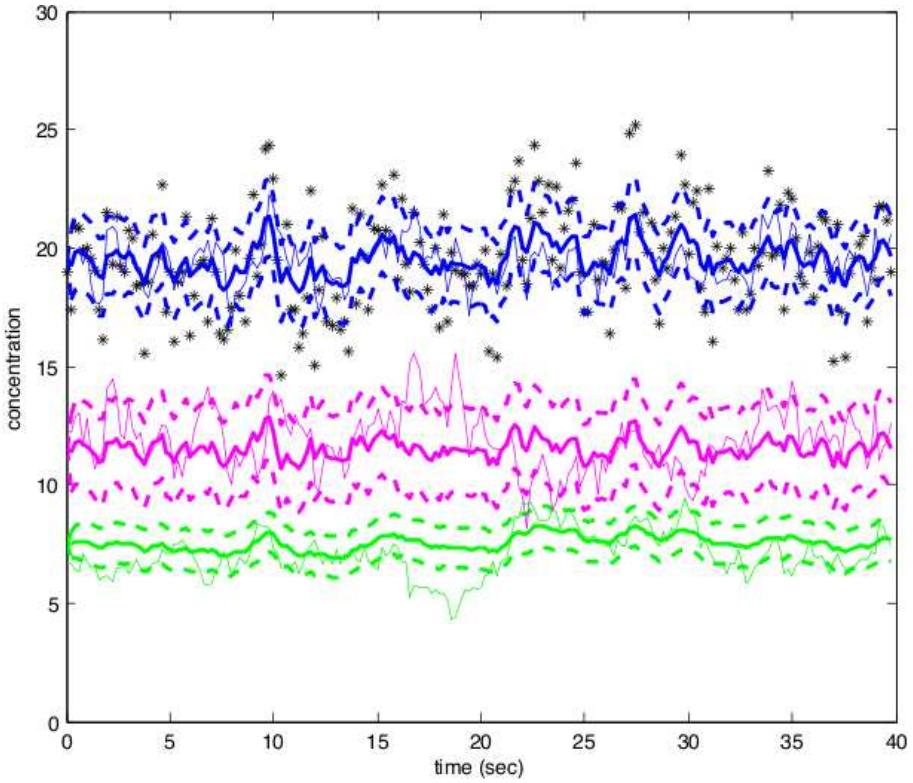

FIGURE 2.1 – Noisy observations of one simulated data set ($n = 200$, $\Delta = 0.2$, $\sigma^2 = 3$) are plotted with stars. True simulated trajectories (thin solid lines), mean estimated trajectories (thick solid lines) and estimated 95% confidence intervals (dotted lines) obtained with the Kalman algorithm are plotted with dark lines for $S(t)$, light lines for $P(t)$ and very light line for $I(t)$.

The EM estimates are often less biased for parameters θ_1 and θ_2 than the

MLE estimates. Variance parameters (θ_3, θ_4) are estimated with less bias by the MLE method, θ_4 is always estimated with a large bias by the EM algorithm. The standard errors of the EM estimates are lower than those of the MLE estimates. The standard errors reduce with the increase of the number of observations n. The bias and standard errors of all parameters decrease with the decrease of the observation noise σ. When Δ decreases, bias and standard errors decrease for a small observation noise $\sigma = 1$. MLE estimates are very satisfactory in this case. The exact Fisher information matrix provides standard errors of the estimates which are close to the empirical ones, especially those obtained with the MLE approach. Note that the theoretical study of the exact MLE allows to deduce the identifiable parameters. On the contrary, the EM algorithm misses completely the problem.

				$\sigma^2 = 1$			
		$n = 200$, $\Delta = 0.2$			$n = 1000$, $\Delta = 0.2$		
Par.	true value	EM (SE)	MLE (SE)	Fisher SE	EM (SE)	MLE (SE)	Fisher SE
θ_1	0.60	0.62 (0.11)	0.64 (0.17)	(0.15)	0.62 (0.05)	0.68 (0.11)	(0.11)
θ_2	0.90	0.89 (0.06)	0.79 (0.15)	(0.04)	0.92 (0.02)	0.87 (0.07)	(0.11)
θ_3	0.70	0.85 (0.20)	0.78 (0.26)	(0.25)	0.86 (0.09)	0.76 (0.15)	(0.37)
θ_4	0.20	0.10 (0.01)	0.12 (0.16)	(0.14)	0.10 (0.01)	0.10 (0.09)	(0.36)
θ_6	20.00	20.00 (0.39)	20.00 (0.38)		20.01 (0.17)	20.01 (0.17)	
$\sigma_1^2 + \sigma_2^2$	0.32	0.35 (0.09)	0.34 (0.14)		0.34 (0.04)	0.30 (0.07)	
				$\sigma^2 = 3$			
		$n = 200$, $\Delta = 0.2$			$n = 1000$, $\Delta = 0.2$		
Par.	true value	EM (SE)	MLE (SE)	Fisher SE	EM (SE)	MLE (SE)	Fisher SE
θ_1	0.60	0.58 (0.13)	0.59 (0.20)	(0.17)	0.57 (0.06)	0.59 (0.13)	(0.24)
θ_2	0.90	0.89 (0.06)	0.76 (0.19)	(0.05)	0.92 (0.02)	0.88 (0.07)	(0.04)
θ_3	0.70	0.96 (0.27)	0.85 (0.43)	(0.53)	0.96 (0.10)	0.88 (0.23)	(0.22)
θ_4	0.20	0.10 (0.01)	0.15 (0.23)	(0.11)	0.10 (0.01)	0.14 (0.09)	(0.13)
θ_6	20.00	20.00 (0.41)	20.01 (0.39)		19.98 (0.18)	19.98 (0.18)	
$\sigma_1^2 + \sigma_2^2$	0.32	0.41 (0.14)	0.42 (0.31)		0.39 (0.05)	0.38 (0.12)	

TABLE 2.1 – Mean estimated values (with empirical standard errors in bracket) obtained with the exact MLE and the EM algorithms and exact standard errors obtained from the Fisher information matrix, evaluated on 1000 simulated data with $n = 200$ and $n = 1000$ observations and $\sigma^2 = 1$ or $\sigma^2 = 3$ (σ^2 and θ_5 fixed to their true values).

$\sigma^2 = 1$								
	$n = 200$, $\Delta = 0.2$				$n = 1000$, $\Delta = 0.04$			
Par.	true value	EM (SE)	MLE (SE)	Fisher SE	true value	EM (SE)	MLE (SE)	Fisher SE
θ_1	0.60	0.62 (0.11)	0.64 (0.17)	(0.15)	0.91	0.86 (0.03)	0.87 (0.08)	(0.06)
θ_2	0.90	0.89 (0.06)	0.79 (0.15)	(0.04)	0.99	0.97 (0.01)	0.94 (0.12)	(0.02)
θ_3	0.70	0.85 (0.20)	0.78 (0.26)	(0.25)	0.20	0.26 (0.02)	0.06 (0.26)	(0.23)
θ_4	0.20	0.10 (0.01)	0.12 (0.16)	(0.14)	0.03	0.09 (0.01)	0.03 (0.13)	(0.17)
θ_6	20.00	20.00 (0.39)	20.00 (0.38)		20.00	20.01 (0.56)	20.01 (0.56)	
$\sigma_1^2 + \sigma_2^2$	0.32	0.35 (0.09)	0.34 (0.14)		0.05	0.09 (0.01)	0.02 (0.09)	
$\sigma^2 = 3$								
	$n = 200$, $\Delta = 0.2$				$n = 1000$, $\Delta = 0.04$			
Par.	true value	EM (SE)	MLE (SE)	Fisher SE	true value	EM (SE)	MLE (SE)	Fisher SE
θ_1	0.60	0.58 (0.13)	0.59 (0.20)	(0.17)	0.91	0.74 (0.06)	0.79 (0.15)	(0.01)
θ_2	0.90	0.89 (0.06)	0.76 (0.19)	(0.05)	0.99	0.96 (0.02)	0.89 (0.17)	(0.01)
θ_3	0.70	0.96 (0.27)	0.85 (0.43)	(0.53)	0.20	0.24 (0.02)	0.20 (1.39)	(0.04)
θ_4	0.20	0.10 (0.01)	0.15 (0.23)	(0.11)	0.03	0.08 (0.01)	0.14 (1.85)	(0.01)
θ_6	20.00	20.00 (0.41)	20.01 (0.39)		20.00	19.99 (0.53)	19.99 (0.54)	
$\sigma_1^2 + \sigma_2^2$	0.32	0.41 (0.14)	0.42 (0.31)		0.05	0.13 (0.01)	0.39 (5.40)	

TABLE 2.2 – Mean estimated values (with empirical standard errors in bracket) obtained with the exact MLE and the EM algorithms and exact standard errors obtained from the Fisher information matrix, evaluated on 1000 simulated data with $n = 200$ and $n = 1000$ observations and $\sigma^2 = 1$ or $\sigma^2 = 3$ (σ^2 and θ_5 fixed to their true values).

2.7 Conclusion

The Kalman filter is classical in the field of noisy, discretely and partially observed stochastic differential equations. In this paper, we have shown that it can be used for the estimation of the parameters by maximum likelihood. In the particular case of an Ornstein-Uhlenbeck process, this method computes the exact likelihood, its gradient and hessian. We have also shown that the EM algorithm, which is classical in the field of hidden Markov models, combined with a smoother algorithm can be used for the parameter estimation.

We study some theoretical properties of the model. We show that only six out of the seven parameters are identifiable and we deduce the asymptotic properties of the maximum likelihood estimate. We illustrate the two methods on simulated data. The identifiability problem is confirmed on the simulation study : the observed Fisher information matrix computed by the Kalman method is not

invertible when we estimate the seven parameters.

The next step of this work is its application to real data in anti-cancer therapy. This work could also be extended to the case of non-Gaussian observation errors. For a unidimensional Markov chain (X_i) observed with non Gaussian errors, Ruiz (1994) proposes a quasi-maximum likelihood estimator based on the Kalman filter and shows the normality asymptotic distribution of this estimator. This approach can be extended to our bidimensional model.

Acknowledgments

The authors thank V. Genon-Catalot for her constructive advices and help. The authors thank C.A. Cuenod, D. Balvay and Y. Rozenholc for their helpful discussions on the biological problem. This project was supported by a grant BQR from University Paris Descartes managed by Y. Rozenholc.

Appendices

Annexe A

Appendix

A.1 Physiological model

We focus on the evaluation of anti-angiogenesis treatments in anti-cancer therapy. These treatments take effect on the vascularization of tissue. The *in vivo* evaluation of their efficacy is based on the estimation of the tissue micro-vascularization parameters. The experiment consists in the injection of a contrast agent to the patient, followed by the recording of a medical images sequence which measures the evolution of the concentration of contrast agent along time. The contrast agent pharmacokinetic is modeled by a bidimensional differential system. The contrast agent pulsates in the plasma and interstitium cells. Let $a(t)$, $P(t)$ and $I(t)$ denote respectively the quantity of contrast agent at time t in the artery, the plasma and the interstitium and $1-h$, V_P and V_I the volume of artery, plasma and interstitium (h is the hematocrit rate). The initial condition at time $t_0 = 0$ is $P(0) = 0, I(0) = 0$. The contrast agent is injected in vein at time t_0, transits in the artery and arrives in plasma, with a tissue perfusion flow equal to F_{tp}. The contrast agent is eliminated from plasma with the perfusion flow F_{tp}, proportionally to the concentration of contrast agent in plasma. The quantity of contrast agent exchanging from plasma through interstitium is equal to K_{trans} times the concentration of contrast agent in plasma, where K_{trans} is the volume transfer constant. Inversely, the quantity of contrast agent exchanging from interstitium through plasma is equal to K_{trans} times the concentration of contrast agent in interstitium. Lastly, the two-compartment model is :

$$\left\{ \begin{array}{rcl} \frac{dP(t)}{dt} & = & \frac{F_{tp}}{1-h}a(t) - \frac{K_{trans}}{V_P}P(t) + \frac{K_{trans}}{V_I}I(t) - \frac{F_{tp}}{V_P}P(t) \\ \frac{dI(t)}{dt} & = & \frac{K_{trans}}{V_P}P(t) - \frac{K_{trans}}{V_I}I(t) \end{array} \right. \tag{A.1}$$

For statistical accommodations, we use a new parameterization and set :

$$\alpha = \frac{F_{tp}}{1-h}, \beta = \frac{F_{tp}}{V_P}, \lambda = \frac{K_{trans}}{V_P}, k = \frac{K_{trans}}{V_P} + \frac{K_{trans}}{V_I}$$

Model (A.1) can thus be transformed as follows :

$$\left\{ \begin{array}{rcl} \frac{dP(t)}{dt} & = & \alpha a(t) - \lambda P(t) + (k-\lambda) I(t) - \beta P(t) \\ \frac{dI(t)}{dt} & = & \lambda P(t) - (k-\lambda) I(t) \end{array} \right. \tag{A.2}$$

A.2 Proof of Proposition 2.1

The process $X(t) = P^{-1}U(t)$ is solution of :

$$dX(t) = (DX(t) + P^{-1}F(t))dt + P^{-1}\Sigma dW_t,\ X(0) = X_0 = P^{-1}U_0.$$

Applying Ito's formula, we obtain

$$X(t) = e^{Dt}X_0 + e^{Dt}\int_0^t e^{-Ds}P^{-1}F(s)ds + e^{Dt}\int_0^t e^{-Ds}\Gamma dW_s.$$

From this equation, we deduce :

$$X(t+h) = e^{Dh}X(t) + B(t, t+h) + Z(t, t+h)$$

where $B(t, t+h)$ and $Z(t, t+h)$ are given in Proposition 2.1. Using that W_1, W_2 are independent and that X_0 is independent of (W_1, W_2), we obtain the conditional law of $X(t+h)|(X(s), s \leq t)$.

The stationary distribution can be deduced from equation (2.3) with $a(t) = c$. As the two elements of D are negative, we have

$$\begin{array}{rcl} \lim\limits_{t\to+\infty} \mathbb{E}(X(t)) & = & \lim\limits_{t\to+\infty} e^{Dt}\mathbb{E}(X_0) + \lim\limits_{t\to+\infty} B(0,t) = -D^{-1}P^{-1}F = M \\ \lim\limits_{t\to+\infty} Var(X(t)) & = & \lim\limits_{t\to+\infty} R(0,t) = \left(\frac{1}{-(\mu_k + \mu_{k'})} (\Gamma\Gamma')^{kk'} \right)_{1\leq k,k'\leq 2} = V. \end{array}$$

If $X_0 \sim \mathcal{N}_2(M, V)$, an elementary computation shows that $(X(t))$ is strictly stationary. □

A.3 Link between the two parametrisations

New parameters $(\theta_1, \ldots, \theta_5)$ are given as functions of initial parameters $(\beta, \lambda, k, \sigma_1, \sigma_2)$ in Section 2.3.2.1. Now we deduce $(\beta, \lambda, k, \sigma_1, \sigma_2)$ from $(\theta_1, \ldots, \theta_5)$. From the definition of μ_1 and μ_2 (Section 2.2), we have

$$\mu_2 - \mu_1 = \sqrt{d}, \quad \mu_1 + \mu_2 = -(\beta + k), \quad \mu_1\mu_2 = \beta(k - \lambda)$$

Thus we rewrite the matrix P and its inverse

$$P = \begin{pmatrix} 1 & 1 \\ \frac{\mu_1}{\beta}+1 & \frac{\mu_2}{\beta}+1 \end{pmatrix}, \quad P^{-1} = \begin{pmatrix} \frac{\mu_2+\beta}{\mu_2-\mu_1} & \frac{-\beta}{\mu_2-\mu_1} \\ -\frac{\mu_1+\beta}{\mu_2-\mu_1} & \frac{\beta}{\mu_2-\mu_1} \end{pmatrix}.$$

Then the covariance matrix Γ is

$$\Gamma = \begin{pmatrix} \sigma_1 \frac{\mu_2+\beta}{\mu_2-\mu_1} & \sigma_2 \frac{\mu_2}{\mu_2-\mu_1} \\ -\sigma_1 \frac{\mu_1+\beta}{\mu_2-\mu_1} & -\sigma_2 \frac{\mu_1}{\mu_2-\mu_1} \end{pmatrix}$$

and finally comes the matrix R

$$R = \begin{pmatrix} \frac{e^{2\mu_1\Delta}-1}{2\mu_1}(\sigma_1^2(\frac{\mu_2+\beta}{\mu_2-\mu_1})^2 + \sigma_2^2(\frac{\mu_2}{\mu_2-\mu_1})^2) & \frac{e^{(\mu_1+\mu_2)\Delta}-1}{\mu_1+\mu_2}(-\sigma_1^2 \frac{\mu_1+\beta}{\mu_2-\mu_1}\frac{\mu_2+\beta}{\mu_2-\mu_1} - \sigma_2^2 \frac{\mu_1\mu_2}{\mu_2-\mu_1}) \\ \frac{e^{(\mu_1+\mu_2)\Delta}-1}{\mu_1+\mu_2}(-\sigma_1^2 \frac{\mu_1+\beta}{\mu_2-\mu_1}\frac{\mu_2+\beta}{\mu_2-\mu_1} - \sigma_2^2 \frac{\mu_1\mu_2}{\mu_2-\mu_1}) & \frac{e^{2\mu_2\Delta}-1}{2\mu_2}(\sigma_1^2(\frac{\mu_1+\beta}{\mu_2-\mu_1})^2 + \sigma_2^2(\frac{\mu_1}{\mu_2-\mu_1})^2) \end{pmatrix}$$

Define

$$\tilde{\theta}_3 = \sigma_1^2(\mu_2+\beta)^2+\sigma_2^2\mu_2^2, \quad \tilde{\theta}_4 = \sigma_1^2(\mu_1+\beta)^2+\sigma_2^2\mu_1^2, \quad \tilde{\theta}_5 = -\sigma_1^2(\mu_1+\beta)(\mu_2+\beta)-\sigma_2^2\mu_1\mu_2$$

we have

$$\tilde{\theta}_3 = \theta_3 \frac{2\mu_1(\mu_2-\mu_1)^2}{\exp(2\mu_1\Delta)-1}, \quad \tilde{\theta}_4 = \theta_4 \frac{2\mu_2(\mu_2-\mu_1)^2}{\exp(2\mu_2\Delta)-1}, \quad \tilde{\theta}_5 = \theta_5 \frac{(\mu_1+\mu_2)(\mu_2-\mu_1)^2}{\exp((\mu_1+\mu_2)\Delta)-1}.$$

Notice that $\mu_1 = \log(\theta_1)/\Delta$ and $\mu_2 = \log(\theta_2)/\Delta$. The parameters β, λ, k, σ_1^2 and σ_2^2 are solution of the system

$$\begin{cases} k &= -(\mu_1+\mu_2+\beta), \\ \lambda &= -(\frac{\mu_1\mu_2}{\beta}+k), \\ (\mu_2+\beta)^2\sigma_1^2 + \mu_2^2\sigma_2^2 &= \tilde{\theta}_3, \\ (\mu_1+\beta)^2\sigma_1^2 + \mu_1^2\sigma_2^2 &= \tilde{\theta}_4, \\ \sigma_1\beta^2 + \sigma_1(\mu_1+\mu_2)\beta + \tilde{\theta}_5 + (\sigma_1^2+\sigma_2^2)\mu_1\mu_2 &= 0. \end{cases}$$

A.4 Gradient and hessian of the log-likelihood

The gradient and the hessian of the loglikelihood are computed with explicit recursions. We denote $\vartheta_7 = \sigma^2$. The first order derivatives of $A(\vartheta)$ are equal to :

$$\frac{\partial A(\vartheta)}{\partial \vartheta_1} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \frac{\partial A(\vartheta)}{\partial \vartheta_2} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \frac{\partial A(\vartheta)}{\partial \vartheta_q} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, q = 3,4,5,6,7.$$

for $R(\vartheta)$ we get :

$$\frac{\partial R(\vartheta)}{\partial \vartheta_1} = \frac{\partial R(\vartheta)}{\partial \vartheta_2} = \frac{\partial R(\vartheta)}{\partial \vartheta_6} = \frac{\partial R(\vartheta)}{\partial \vartheta_7} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

$$\frac{\partial R(\vartheta)}{\partial \vartheta_3} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \frac{\partial R(\vartheta)}{\partial \vartheta_4} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \frac{\partial R(\vartheta)}{\partial \vartheta_5} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

and for m we get : $\frac{\partial m}{\partial \vartheta_q} = 0, q = 1, \ldots, 5, 7$ and $\frac{\partial m}{\partial \vartheta_6} = 1$. The second order derivatives of $A(\vartheta)$ and $R(\vartheta)$ are null. The first order derivatives of $\hat{X}_i^-(\vartheta)$ and $P_i^-(\vartheta)$ with respect to ϑ_q, $q = 1, \ldots, 7$ can be deduced :

$$\frac{\partial \hat{X}_i^-(\vartheta)}{\partial \vartheta_q} = \frac{\partial A(\vartheta)}{\partial \vartheta_q} \hat{X}_{i-1}(\vartheta) + A(\vartheta) \frac{\partial \hat{X}_{i-1}(\vartheta)}{\partial \vartheta_q}$$
$$\frac{\partial P_i^-(\vartheta)}{\partial \vartheta_q} = \frac{\partial A(\vartheta)}{\partial \vartheta_q} P_{i-1}(\vartheta) A(\vartheta)' + A(\vartheta) \frac{\partial P_{i-1}(\vartheta)}{\partial \vartheta_q} A(\vartheta)' + A(\vartheta) P_{i-1}(\vartheta) \frac{\partial A(\vartheta)'}{\partial \vartheta_q} + \frac{\partial R(\vartheta)}{\partial \vartheta_q}.$$

Then we get the derivatives of the mean and the covariance of the filter :

$$\frac{\partial \hat{X}_{i-1}(\vartheta)}{\partial \vartheta_q} = \frac{\partial \hat{X}_{i-1}^-(\vartheta)}{\partial \vartheta_q} + \frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_q} \frac{H' W_{i-1}(\vartheta)}{V_{i-1}(\vartheta)} + \frac{P_{i-1}^-(\vartheta) H'}{V_{i-1}(\vartheta)} \left(\frac{\partial W_{i-1}(\vartheta)}{\partial \vartheta_q} - \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_q} \frac{W_{i-1}(\vartheta)}{V_{i-1}(\vartheta)} \right)$$
$$\frac{\partial P_{i-1}(\vartheta)}{\partial \vartheta_q} = \left(I - \frac{P_{i-1}^- H' H}{V_{i-1}(\vartheta)} \right) \frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_q} - \left(\frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_q} \frac{H'}{V_{i-1}(\vartheta)} - \frac{P_{i-1}^-(\vartheta) H'}{V_{i-1}(\vartheta)^2} \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_q} \right) H P_{i-1}^-(\vartheta)$$

The computation of second order derivatives can be deduced. Because second order derivatives of A and R are null, we have for $q, r = 1, \ldots, 7$, :

$$\frac{\partial^2 \hat{X}_i^-(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} = A(\vartheta) \frac{\partial^2 \hat{X}_{i-1}(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} + \frac{\partial A(\vartheta)}{\partial \vartheta_r} \frac{\partial \hat{X}_{i-1}(\vartheta)}{\partial \vartheta_q} + \frac{\partial A(\vartheta)}{\partial \vartheta_q} \frac{\partial \hat{X}_{i-1}(\vartheta)}{\partial \vartheta_r}$$
$$\frac{\partial^2 P_i^-(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} = \left(\frac{\partial A(\vartheta)}{\partial \vartheta_r} \frac{\partial P_{i-1}(\vartheta)}{\partial \vartheta_q} + \frac{\partial A(\vartheta)}{\partial \vartheta_q} \frac{\partial P_{i-1}(\vartheta)}{\partial \vartheta_r} \right) A(\vartheta)' + \frac{\partial A(\vartheta)}{\partial \vartheta_r} P_{i-1}(\vartheta) \frac{\partial A(\vartheta)'}{\partial \vartheta_q}$$
$$+ A(\vartheta) \left(\frac{\partial^2 P_{i-1}(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} A(\vartheta)' + \frac{\partial P_{i-1}(\vartheta)}{\partial \vartheta_r} \frac{\partial A(\vartheta)'}{\partial \vartheta_q} + \frac{\partial P_{i-1}(\vartheta)}{\partial \vartheta_q} \frac{\partial A(\vartheta)'}{\partial \vartheta_r} \right) + \frac{\partial A(\vartheta)}{\partial \vartheta_q} P_{i-1}(\vartheta) \frac{\partial A(\vartheta)'}{\partial \vartheta_r}$$

The second order derivatives of the mean and the covariance of Kalman filter are :

$$\frac{\partial^2 \hat{X}_{i-1}(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} = \frac{\partial^2 \hat{X}_{i-1}^-(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} + \frac{\partial^2 P_{i-1}^-(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} \frac{H' W_{i-1}(\vartheta)}{V_{i-1}(\vartheta)} + P_{i-1}^-(\vartheta) \frac{H'}{V_{i-1}(\vartheta)} \left(\frac{\partial^2 W_{i-1}(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} \right.$$
$$- \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_q} \frac{1}{V_{i-1}(\vartheta)} \frac{\partial W_{i-1}(\vartheta)}{\partial \vartheta_r} - \frac{\partial^2 V_{i-1}(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} \frac{W_{i-1}(\vartheta)}{V_{i-1}(\vartheta)} + 2 \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_r} \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_q} \frac{W_{i-1}(\vartheta)}{V_{i-1}^2(\vartheta)}$$
$$\left. - \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_r} \frac{1}{V_{i-1}(\vartheta)} \frac{\partial W_{i-1}(\vartheta)}{\partial \vartheta_q} \right) + \frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_r} \frac{H'}{V_{i-1}(\vartheta)} \left(\frac{\partial W_{i-1}(\vartheta)}{\partial \vartheta_q} - \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_q} \frac{W_{i-1}(\vartheta)}{V_{i-1}(\vartheta)} \right)$$
$$+ \frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_q} \frac{H'}{V_{i-1}(\vartheta)} \left(\frac{\partial W_{i-1}(\vartheta)}{\partial \vartheta_r} - \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_r} \frac{W_{i-1}(\vartheta)}{V_{i-1}(\vartheta)} \right)$$

$$\frac{\partial^2 P_{i-1}(\vartheta)}{\partial\vartheta_q\partial\vartheta_r} = \left(I - \frac{P_{i-1}^-(\vartheta)H'H}{V_{i-1}(\vartheta)}\right)\frac{\partial^2 P_{i-1}^-(\vartheta)}{\partial\vartheta_q\partial\vartheta_r} - \left[\frac{\partial^2 P_{i-1}^-(\vartheta)}{\partial\vartheta_q\partial\vartheta_r} - \frac{P_{i-1}^-(\vartheta)}{V_{i-1}(\vartheta)}\frac{\partial^2 V_{i-1}(\vartheta)}{\partial\vartheta_q\partial\vartheta_r}\right.$$
$$\left. - \left(\frac{\partial P_{i-1}^-(\vartheta)}{\partial\vartheta_r}\frac{\partial V_{i-1}(\vartheta)}{\partial\vartheta_q} + \frac{\partial P_{i-1}^-(\vartheta)}{\partial\vartheta_q}\frac{\partial V_{i-1}(\vartheta)}{\partial\vartheta_r}\right)\frac{1}{V_{i-1}(\vartheta)} + \frac{2P_{i-1}^-(\vartheta)}{V_{i-1}^2(\vartheta)}\frac{\partial V_{i-1}(\vartheta)}{\partial\vartheta_q}\frac{\partial V_{i-1}(\vartheta)}{\partial\vartheta_r}\right]\frac{H'HP_{i-1}^-(\vartheta)}{V_{i-1}(\vartheta)}$$
$$- \left(\frac{\partial P_{i-1}^-(\vartheta)}{\partial\vartheta_r}H' - \frac{P_{i-1}^-(\vartheta)H'}{V_{i-1}(\vartheta)}\frac{\partial V_{i-1}(\vartheta)}{\partial\vartheta_r}\right)\frac{H}{V_{i-1}(\vartheta)}\frac{\partial P_{i-1}^-(\vartheta)}{\partial\vartheta_q}$$
$$- \left(\frac{\partial P_{i-1}^-(\vartheta)}{\partial\vartheta_q}H' - P_{i-1}^-(\vartheta)\frac{H'}{V_{i-1}(\vartheta)}\frac{\partial V_{i-1}(\vartheta)}{\partial\vartheta_q}\right)\frac{H}{V_{i-1}(\vartheta)}\frac{\partial P_{i-1}^-(\vartheta)}{\partial\vartheta_r}$$

A.5 Smoother algorithm

The Kalman smoother is calculated recursively with a forward-backward algorithm (see *e.g.* Cappé et al. (2005)). The forward algorithm is the classical Kalman filter which computes $M_{i|0:i-1} = \hat{Z}_i^- = \mathbb{E}(Z_i|y_{0:i-1})$, $\Sigma_{i|0:i-1} = P_i^- = Var(Z_i|y_{0:i-1})$, $M_{i|0:i} = \hat{Z}_i = \mathbb{E}(Z_i|y_{0:i})$ and $\Sigma_{i|0:i} = P_i = Var(Z_i|y_{0:i})$. Then, in order to calculate $M_{i|0:n} = \mathbb{E}(Z_i|y_{0:n})$, $\Sigma_{i|0:n} = Var(Z_i|y_{0:n})$, $\Sigma_{i-1,i|0:n} = Cov(Z_{i-1}, Z_i|y_{0:n})$, one performs the set of backward recursions $i = n, n-1, \ldots, 1$:

$$\begin{aligned} J_{i-1} &= \Sigma_{i-1|0:i-1}A'(\Sigma_{i|0:i-1})^{-1} \\ M_{i-1|0:n} &= M_{i-1|0:i-1} + J_{i-1}(M_{i|0:n} - M_{i|0:i-1}) \\ \Sigma_{i-1|0:n} &= \Sigma_{i-1|0:i-1} + J_{i-1}(\Sigma_{i|0:n} - \Sigma_{i|0:i-1})J_{i-1}' \end{aligned}$$

To calculate $\Sigma_{i-1,i|0:n}$, we have

$$\Sigma_{n-1,n|0:n} = (I - K_nH)A\Sigma_{n-1|0:n-1}$$

and the following backward recursions, for $i = n-1, n-2, \ldots, 1$

$$\Sigma_{i-1,i|0:n} = \Sigma_{i|0:i}J_{i-1}' + J_i(\Sigma_{i,i+1|0:n} - A\Sigma_{i|0:i})J_{i-1}'$$

A.6 ARMA property of multidimensional process

In our model, $(y_i)_{i\in\mathbb{Z}}$ is an ARMA(2,2) process and asymptotic properties of the maximum likelihood estimator are derived. This result can be generalised to the case where X_i is p-dimensional under weak assumptions. We consider the model :

$$y_i = HX_i + \sigma\varepsilon_i, X_i = AX_{i-1} + \eta_i, \quad X_0 \sim \nu$$

where X_i is p-dimensional, A is a diagonal matrix with diagonal coefficients $(\theta_k, k = 1, \ldots, p)$ such that $\theta_k \neq \theta_l$ for $k \neq l$ and $\theta_i \in (0, 1)$ for $i = 1 \ldots p$, $(\eta_i)_{i\geq 0}$ is a sequence of independent $\mathcal{N}_p(0, R)$ random variables, H is a $(1, p)$-matrix and (ε_i) is a sequence of i.i.d $\mathcal{N}(0, 1)$ random variables. Up to a transformation of (X_i), H is equal to $H = (1 \ldots 1\ 0 \ldots 0)$ with its first d coordinates equal to 1 and its $p - d$ next coordinates equal to 0 $(1 \leq d \leq p)$. Consequently, we observe with additive noise the partial sum of the first d coordinates of X_i. Since A is diagonal and $\theta_i \in (0, 1)$ for $i = 1 \ldots p$, the process $(X_i)_{i\geq 0}$ admits a stationary distribution ν. Then with $X_0 \sim \nu$, the process $(X_i)_{i\geq 0}$ is stationary. Denote $(y_i)_{i\in\mathbb{Z}}$ the extension of y to $\mathbb{Z}$ by stationarity.

Proposition A.1 *The process $(y_i)_{i\in\mathbb{Z}}$ is ARMA(d,d).*

Proof. Denote S_j, $j = 1, \ldots, d$, the j-th symmetric function of $\theta_1, \ldots, \theta_d$:

$$S_j = \sum_{1\leq i_1<\cdots<i_j\leq d} \theta_{i_1} \ldots \theta_{i_j}$$

and $S_0 = 1$. Define the polynomial P such that $P(\theta_1) = \ldots = P(\theta_d) = 0$:

$$P(x) = \sum_{k=0}^{d} (-1)^k S_k x^{d-k}$$

Set L the one-lag operator : $Ly_i = y_{i-1}$, $L\varepsilon_i = \varepsilon_{i-1}$. Set $\xi_i = P(L)(y_i) = \sum_{k=0}^{d} (-1)^k S_k y_{i-k}$. By recursive computation for $0 \leq k \leq d$, we have

$$y_{i-k} = HA^{d-k}X_{i-d} + \sum_{j=0}^{d-k-1} HA^j\eta_{i-k-j} + \sigma\varepsilon_{i-k}$$

We deduce

$$\xi_i \;=\; HP(A)X_{i-d} + \sum_{k=0}^{d} (-1)^k S_k \left(\sum_{j=0}^{d-k-1} HA^j\eta_{i-k-j} \right) + \sigma P(L)(\varepsilon_i)$$

But $HP(A) = (P(\theta_1) \ldots P(\theta_d)\ 0 \ldots 0) = (0 \ldots\ldots 0)$. Thus ξ_i only depends on $(\eta_j, \varepsilon_j)_{j\leq i}$. Therefore (y_i) verifies an AR(d) equation. Moreover, as the $(\eta_i)_i$ and $(\varepsilon_i)_i$ are mutually independent, we get that :

$$Cov(\xi_i, \xi_{i+k}) = 0,\ \forall k \geq d.$$

This implies that (ξ_i) is MA(d). Hence the result.

A.7 First order identifiability

The spectral density can be rewritten as

$$f(u,\vartheta) = \sigma^2 + \frac{d(\vartheta)e^{iu} + c(\vartheta) + d(\vartheta)e^{-iu}}{(1-(\theta_1+\theta_2)e^{iu}+\theta_1\theta_2e^{2iu})(1-(\theta_1+\theta_2)e^{-iu}+\theta_1\theta_2e^{-2iu})}$$

where $c(\vartheta) = H(ARA' + (1+(\theta_1+\theta_2)^2)R)H'$ and $d(\vartheta) = (HA - (\theta_1+\theta_2)H)RH'$. The equality $f(u,\vartheta) = f(u,\vartheta')\ \forall u \in (0,2\pi)$ implies

$$\sigma^2 = \sigma'^2, \quad \theta_1+\theta_2 = \theta'_1+\theta'_2, \quad \theta_1\theta_2 = \theta'_1\theta'_2, \quad c(\vartheta) = c(\vartheta'), \quad d(\vartheta) = d(\vartheta')$$

We deduce that σ^2, $\theta_1+\theta_2$, $\theta_1\theta_2$ are identifiable and that at most two of three parameters among θ_3, θ_4 and θ_5 are identifiable from $c(\vartheta)$ and $d(\vartheta)$.

We can prove that there are exactly two parameters identifiable when Δ is small. Set

$$g(z,\vartheta) = d(\vartheta) + c(\vartheta)z + d(\vartheta)z^2.$$

so that for $z = e^{iu}$

$$f(z,\vartheta) = \sigma^2 + z\frac{g(z,\vartheta)}{(1-\theta_1 z)(\theta_1 - z)(1-\theta_2 z)(\theta_2 - z)}$$

Note that the product of the roots of g is equal to 1. Thus, if θ_1 and θ_2 are not roots of g, then no simplification is possible in the expression of f and exactly two out of the three parameters θ_3, θ_4, θ_5 are identifiable (from $c(\vartheta)$ and $d(\vartheta)$). The computation of $g(\theta_1)$ and $g(\theta_2)$ gives

$$\begin{aligned}
g(\theta_1) &= (\theta_1-\theta_2)(\theta_3+\theta_5) + 2\theta_1^3\theta_3 + \theta_1^3\theta_5 \\
&\quad +\theta_1\theta_2^2\theta_3 + \theta_1^2\theta_2\theta_3 + 2\theta_1\theta_2^2\theta_4 + 2\theta_1^2\theta_2\theta_4 + 2\theta_1\theta_2^2\theta_5 + 5\theta_1^2\theta_2\theta_5 \\
g(\theta_2) &= (\theta_2-\theta_1)(\theta_4+\theta_5) + 2\theta_2^3\theta_4 + \theta_2^3\theta_5 \\
&\quad +2\theta_1\theta_2^2\theta_3 + 2\theta_1^2\theta_2\theta_3 + \theta_1\theta_2^2\theta_4 + \theta_1^2\theta_2\theta_4 + 5\theta_1\theta_2^2\theta_5 + 2\theta_1^2\theta_2\theta_5
\end{aligned}$$

It is not straightforward to prove that g does not vanish in θ_1 and θ_2. But for small Δ, by developping $\theta_1, \ldots, \theta_5$ at first order we have

$$\begin{aligned}
\theta_1-\theta_2 &= \Delta(\mu_1-\mu_2) + O(\Delta) \\
\theta_3+\theta_5 &= \Delta(\sigma_1^2(\frac{\mu_2+\beta}{\mu_2-\mu_1})^2 + \sigma_2^2(\frac{\mu_2}{\mu_2-\mu_1})^2 - \sigma_1^2\frac{\mu_1+\beta}{\mu_2-\mu_1}\frac{\mu_2+\beta}{\mu_2-\mu_1} - \sigma_2^2\frac{\mu_1\mu_2}{\mu_2-\mu_1}) + O(\Delta) \\
\theta_4+\theta_5 &= \Delta(\sigma_1^2(\frac{\mu_1+\beta}{\mu_2-\mu_1})^2 + \sigma_2^2(\frac{\mu_1}{\mu_2-\mu_1})^2 - \sigma_1^2\frac{\mu_1+\beta}{\mu_2-\mu_1}\frac{\mu_2+\beta}{\mu_2-\mu_1} - \sigma_2^2\frac{\mu_1\mu_2}{\mu_2-\mu_1}) + O(\Delta)
\end{aligned}$$

Hence it comes

$$g(\theta_1) = O(\Delta^2) + 4C\Delta + o(\Delta), \quad g(\theta_2) = O(\Delta^2) + 4C\Delta + o(\Delta).$$

with

$$C = \theta_3 + \theta_4 + 2\theta_5$$

But C is a positive constant because

$$C = (1 \quad 1)R(1 \quad 1)'.$$

where R is a covariance matrix thus positive. Hence $g(\theta_1)$ and $g(\theta_2)$ are positive when Δ is small. We have thus a non degenerate ARMA(2,2) process and exactly five parameters are identifiable.

Chapitre 3

Blood micro-circulation parameters extraction from Dynamic Contrast Enhanced MRI data using stochastic differential equations

Abstract

Dynamic Contrast Enhanced imaging (DCE-imaging) following a contrast agent bolus allows the extraction of information on tissue micro-vascularization. The dynamic signals obtained from DCE-imaging are modeled by pharmacokinetic compartmental models whose parameters are tissue micro-vascular parameters. These models use ordinary differential equations (ODEs) to describe the exchanges between the arterial and capillary plasma and the extravascular-extracellular space. Their least squares fitting takes into account measurement noises but fails to deal with unpredictable fluctuations due to external/internal sources of variations (patients' moves or breathing, anxiety, time-varying parameters, etc). Adding Brownian components to the ODEs leads to stochastic differential equations (SDEs) which model these fluctuations. In DCE-imaging, SDEs are discretely observed with an additional measurement noise. We propose to estimate the parameters of these noisy SDEs by maximum likelihood, using the Kalman filter. The uncertainty of the estimated signals is computed with the Kalman smoother algorithm. Estimations based on the SDE and ODE pharmacokinetic models are compared to DCE-MRI data and illustrate the improvement of the SDE model. A simulation study confirms these results.

3.1 Introduction

Tissue micro-vascularization and angiogenesis can now be studied in vivo by several Dynamic Contrast Enhanced Imaging (DCE-imaging) techniques. These techniques are increasingly used in the medical imaging of brain (Wintermark et al., 2006; Ostergaard, 2005), heart (Vallee et al., 1997; Canet et al., 1993; Fritz-Hansen et al., 1998) and cancer (Goh et al., 2007, 2005; Miles, 2003; Padhani et al., 2005; Bisdas et al., 2007). DCE-imaging follows a bolus of contrast agent injected during a sequential imaging acquisition with Computed Tomography, Magnetic Resonance Imaging or Ultrasound imaging (DCE-CT, DCE-MRI or DCE-US) (Miles, 2003; Tofts, 1997). Recent experimental and clinical studies have shown that DCE-imaging can assess tumor aggressiveness and monitor the *in vivo* effects of treatments (Jain et al., 2007; Fournier et al., 2007; Rosen and Schnall, 2007; Zhu et al., 2008). Clinicians aim at building high resolution functional maps on a "voxel by voxel" basis (Sørensen et al., 1997). Functional maps are of great interest in heterogeneous cancerous tumors (Kiessling et al., 2004). They allow better treatment monitoring by optimizing *in vivo* the therapeutic strategy. Four quantities are usually used to characterize vasculature : the tissue blood flow (perfusion), the permeability surface area (Shames et al., 1993), the tissue blood fractional volume and the tissue extravascular-extracellular space fractional volume (Brix et al., 2004; de Bazelaire et al., 2005). Sequential imaging data are related to these parameters through pharmacokinetic compartment models, that describe the exchanges of the contrast agent between a central compartment (plasma) and a peripheral compartment (extracellular space or interstitial water). Recently, the need to integrate the Arterial Input Function (AIF) in the model to estimate microcirculation parameters more accurately has been emphasized (Port et al., 2001; Krishnamurthi et al., 2005; Brochot et al., 2006). However, deterministic pharmacokinetic models generally fail to capture the random fluctuations due to external or internal environmental causes (patients' moves or breathing, anxiety, small random variations of the micro-vascularization parameters along time, etc) which occur in supplement of the measuring noise due to electronic devices. These additional sources of variations are unpredictable and thus impossible to model in a deterministic way. Spatial averaging or large regions of interest are generally considered to minimize the failure of deterministic models on voxel data (Brochot et al., 2006). This implies mixing or averaging dynamics which may be heterogeneous, and lead to inaccurate parameter estimation. In this paper, we propose to model the unexplained sources of variations by adding random components to the compartment differential system. More precisely, we consider a pharmacokinetic model describing the kinetics of the contrast agent

in the voxel with two compartments (plasma and interstitial water). This deterministic model is transformed into a stochastic differential system by adding a Brownian motion to each differential equation. Observations obtained from DCE-imaging are noisy measurements of the total contrast agent quantity described by the stochastic differential system. The measurement noise differs from the random variations added to the pharmacokinetic model. It is due to the precision of the recording experiments and is thus an uncorrelated noise. Parameter estimation in this model is complex. Though the observations are one-dimensional, we propose an easy way to compute the exact likelihood using the Kalman filter recursion. This enables to implement an easy numerical maximization of the likelihood. The asymptotic properties of the maximum likelihood estimator have been studied by Favetto and Samson (2010). We propose to compute a confidence interval of the extracted signals from this non-linear model using the Kalman smoother algorithm. The aim of this paper is to apply this new stochastic model to estimate microcirculation parameters from DCE-MRI data.

The article is organized as follows. In Section 3.2, the data, the deterministic and stochastic differential systems are presented. The estimation methods are described in Section 3.3. The results of estimation on DCE-MRI data of normal female pelvises are given in Section 3.4. A simulation study is presented in Section 3.5. The discussion and conclusions are presented in Section 6.4. Statistical tools are reported in the Appendix.

3.2 Models

3.2.1 MRI data extraction

The acquisition of the MRI sequence was performed at discrete times $t_0 = 0 < t_1 < \ldots < t_n = T$. Two local sets of voxels were drawn manually on one of the MR images (one in the left iliac artery and the other in the pelvis) and propagated automatically over the entire image sequence. For a given voxel, the gray levels are denoted $(z_0, \ldots, z_n)$. A baseline gray level b_0 is derived by averaging the first times of the sequence before the injection of the contrast agent. Observations are defined as the gray level differences between the voxel gray levels z_i and the baseline gray level b_0 and are denoted $y_i = z_i - b_0$ for $i = 0, \ldots, n$. The measurements obtained from the arterial voxel after removing the baseline are denoted $(AIF(t_i))_{0 \leq i \leq n}$ (Arterial Input Function). We assume that the gray level variation y_i at time t_i is proportional to the total quantity of contrast agent inside the voxel up to some additive measurement errors due to the acquisition technique. Let $S(t)$ denote the total quantity of contrast agent inside the voxel.

The following relation is assumed

$$y_i = S(t_i) + \sigma\varepsilon_i, \tag{3.1}$$

where (ε_i) are the measurement error random variables, assumed to be Gaussian, centered, standardized, mutually independent and σ is the constant noise level. The whole vector of data is now denoted $y_{0:n}$.

3.2.2 Pharmacokinetic models

Two physiological models are considered to describe the evolution of $S(t)$ within each voxel after the contrast agent injection. They are both derived from the same pharmacokinetic model (Brix et al., 2004; de Bazelaire et al., 2005) based on a compartmental analysis (Figure 3.1). In that model, the contrast agent within a tissue voxel is assumed to be either in the plasma compartment of the micro-vessels (capillaries) or inside the interstitial compartment (extracellular-extravascular space). The contrast agent, a gadolinium chelate, cannot enter red or tissue cells. We assume that exchanges inside a voxel are 1/ from the arteries (input) into the blood plasma ; 2/ from the blood plasma into the veins (output) and 3/ between the blood plasma and the interstitial space through the capillary walls. The inter-voxel exchange through the interstitial compartment is assumed to be negligible with respect to the exchange between plasma and interstitium. The contrast agent quantities in a single unit voxel at time t are denoted $Q_P(t)$ and $Q_I(t)$ for plasma and interstitial compartments, respectively. In a single voxel of unit volume, we have $S(t) = Q_P(t) + Q_I(t)$. The biological parameters and constraints are as follows : $F_T \geq 0$ is the tissue blood perfusion flow per unit volume of tissue (in ml.min^{-1}.100ml^{-1}), $V_b \geq 0$ is the part of whole blood volume (in %), $V_e \geq 0$ is the part of extravascular extracellular space fractional volume (in %), and $PS \geq 0$ is the permeability surface area product per unit volume of tissue (in ml.min^{-1}.100ml^{-1}). The change rate constants are F_T and PS. Note that $V_b + V_e \leq 100$. The hematocrit rate h is set to $h = 0.4$. The delay with which the contrast agent arrives from the arteries to the plasma (or Bolus Arrival Time) is denoted δ. Both t and δ are measured in seconds.

3.2.3 The ordinary differential pharmacokinetic model (ODE)

Using the pharmacokinetic model presented in Figure 3.1 and assuming constant rates in the exchanges, the contrast agent kinetics can be modeled by the following

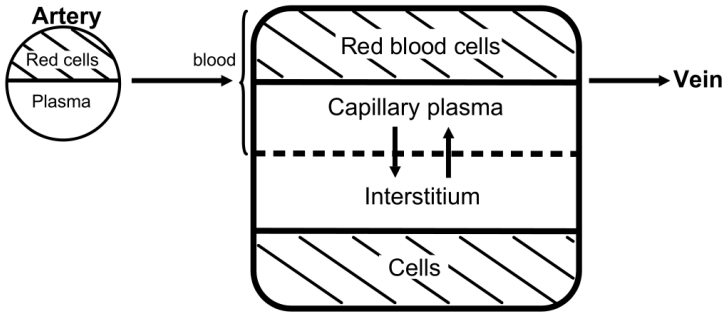

FIGURE 3.1 – Two-compartment physiological pharmacokinetic model used to describe the distribution of the contrast agent. The hatched compartments representing the red blood cells and the tissue cells are not reached by the contrast agent.

differential equations, called the Ordinary Differential Equation (ODE) model :

$$\begin{aligned} \frac{dQ_P(t)}{dt} &= \frac{F_T}{1-h}\, AIF(t-\delta) - \frac{PS}{V_b(1-h)}Q_P(t) + \frac{PS}{V_e}Q_I(t) - \frac{F_T}{V_b(1-h)}Q_P(t) \\ \frac{dQ_I(t)}{dt} &= \frac{PS}{V_b(1-h)}Q_P(t) - \frac{PS}{V_e}Q_I(t) \end{aligned} \tag{3.2}$$

We assume that no contrast agent exists inside the body before acquisition. Hence the initial conditions $Q_P(t_0) = Q_I(t_0) = AIF(t_0) = 0$ hold. This model only requires the biological parameters of interest $(F_T, V_b, PS, V_e, \delta)$ and the knowledge of $AIF(t)$. Given a set of parameters and the AIF, Q_P and Q_I are deterministic functions of time.

3.2.4 The stochastic differential pharmacokinetic model (SDE)

The ODE model (3.2) is a simplified model of the true contrast agent pharmacokinetics. For example, it fails to capture measurement errors in the arterial input function, or random fluctuations along time in the microcirculation parameters. These variations are unpredictable. Our main hypothesis is that a more realistic modeling can be obtained by a stochastic approach. We introduce a stochastic version of the ODE model, by adding random components :

$$\begin{aligned} dQ_P(t) &= \left(\frac{F_T}{1-h}\, AIF(t-\delta) - \frac{PS}{V_b(1-h)}Q_P(t) + \frac{PS}{V_e}Q_I(t) - \frac{F_T}{V_b(1-h)}Q_P(t)\right)dt \\ &\quad +\sigma_1 dW_t^1 \\ dQ_I(t) &= \left(\frac{PS}{V_b(1-h)}Q_P(t) - \frac{PS}{V_e}Q_I(t)\right)dt + \sigma_2 dW_t^2 \end{aligned} \tag{3.3}$$

where (W_t^1) and (W_t^2) are two independent real-valued standard Brownian motions, and σ_1, σ_2 are the standard deviations of the random perturbations (see Favetto and Samson (2010)). The initial conditions are the same as above. Although we use the same notations, given a set of parameters and the AIF, the solutions Q_P and Q_I of (3.3) are stochastic processes (see Liptser and Shiryaev (2001) for formal definitions and Appendix C.3 for details). This model is called the Stochastic Differential Equation (SDE) model. If $\sigma_1 = \sigma_2 = 0$, the ODE and SDE models are identical. Therefore SDE and ODE parameters have the same physiological interpretation. The behavior of the SDE model without measurement noise ($\sigma = 0$) is illustrated on simulations in Appendix. The random part of the SDE represents the random fluctuations around the deterministic ODE. The SDE trajectories are centered on the associated ODE trajectories. In particular, the random perturbation $\sigma_1 dW_t^1$ takes into account the measurement noise of the arterial input function among other sources of variations.

3.3 Estimation methods

3.3.1 Estimation in the ODE model

The estimation of the physiological parameters of interest $(F_T, V_b, PS, V_e, \delta)$ was obtained by applying standard least squares, *i.e.* minimizing

$$\sum_{i=0}^{n}(y_i - S(t_i))^2 = \sum_{i=0}^{n}(y_i - Q_P(t_i) - Q_I(t_i))^2 \tag{3.4}$$

with respect to $(F_T, V_b, PS, V_e, \delta)$ where $(Q_P(t), Q_I(t))$ are the solutions of (3.2). A plug-in of these estimated parameters in (3.2) provided an estimation of the functions $Q_P(t)$, $Q_I(t)$ and $S(t)$ denoted $\hat{Q}_P^{\text{ODE}}(t)$, $\hat{Q}_I^{\text{ODE}}(t)$ and $\hat{S}^{\text{ODE}}(t)$, respectively. To estimate σ, we set

$$\hat{\sigma}^{\text{ODE}} = \sqrt{\frac{1}{n-4}\sum_{i=0}^{n}(y_i - \hat{S}^{\text{ODE}}(t_i))^2}. \tag{3.5}$$

For short, we set $\theta^{\text{ODE}} = (F_T, V_b, PS, V_e, \delta, \sigma)$ and

$$\hat{\theta}^{\text{ODE}} = (\hat{F}_T^{\text{ODE}}, \hat{V}_b^{\text{ODE}}, \hat{PS}^{\text{ODE}}, \hat{V}_e^{\text{ODE}}, \hat{\delta}^{\text{ODE}}, \hat{\sigma}^{\text{ODE}}).$$

Standard deviations for each parameter estimate were obtained through the Fisher information matrix (see Appendix 3.7.3).

3.3.2 Estimation in the SDE model

For the statistical parameters in the SDE model, we set

$$\theta^{\text{SDE}} = (F_T, V_b, PS, V_e, \delta, \sigma_1, \sigma_2, \sigma).$$

Here, the physiological parameters and the noise standard deviations σ_1, σ_2 and σ have to be estimated. This was done by maximizing the likelihood $L(\theta^{\text{SDE}}; y_{0:n})$ of the observations (3.1) with $S(t)$ defined by the SDE model (3.3). Moreover we assumed that the random variables ε_i are standardized Gaussian variables, independent of $(S(t))$. The discretized process $(S(t_i))$ is a hidden Markov chain which yields to a Hidden Markov Model (HMM) for $(y_{0:n}, S(t_i))$. Due to the SDE model and assumptions, $(S(t))$ is a Gaussian process and $y_{0:n}$ is a Gaussian vector. Therefore, the likelihood can be calculated explicitly (Favetto and Samson, 2010). Let $p_{\theta^{\text{SDE}}}(y_i|y_{0:i-1})$ denote the conditional density of the observation y_i given the past $y_{0:i-1}$ assuming that the unknown true parameter is θ^{SDE}. Using the Bayes formula, the likelihood of the HMM can be written as the product of the conditional densities :

$$L(\theta^{\text{SDE}}; y_{0:n}) = p_{\theta^{\text{SDE}}}(y_0) \prod_{i=1}^{n} p_{\theta^{\text{SDE}}}(y_i|y_{0:i-1}) \tag{3.6}$$

At time 0, as no contrast agent exists in the body, y_0 follows a centered Gaussian random variable with variance σ^2. Moreover, as $y_{0:n}$ is a Gaussian vector, the conditional density $p_{\theta^{\text{SDE}}}(y_i|y_{0:i-1})$ is Gaussian and can be expressed using its conditional expectation $E_{\theta^{\text{SDE}}}(y_i|y_{0:i-1})$ and variance $V_{\theta^{\text{SDE}}}(y_i|y_{0:i-1})$ computed at the same unknown true parameter θ^{SDE}. Consequently, (3.6) can be written as

$$\begin{aligned} L(\theta^{\text{SDE}}; y_{0:n}) \;=\; & \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{{y_0}^2}{2\sigma^2}\right) \times \\ & \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi V_{\theta^{\text{SDE}}}(y_i|y_{0:i-1})}} \exp\left(-\frac{1}{2}\frac{(y_i - E_{\theta^{\text{SDE}}}(y_i|y_{0:i-1}))^2}{V_{\theta^{\text{SDE}}}(y_i|y_{0:i-1})}\right). \end{aligned} \tag{3.7}$$

It is possible to compute recursively the terms $E_{\theta^{\text{SDE}}}(y_i|y_{0:i-1})$ and $V_{\theta^{\text{SDE}}}(y_i|y_{0:i-1})$ using the well-known Kalman filter (see Appendix C.3). We denote $\hat{\theta}^{\text{SDE}}$ the estimate of θ^{SDE} obtained by maximizing the likelihood with respect to θ^{SDE} given the observation $y_{0:n}$:

$$\hat{\theta}^{\text{SDE}} = (\hat{F}_T^{\text{SDE}}, \hat{V}_b^{\text{SDE}}, \hat{PS}^{\text{SDE}}, \hat{V}_e^{\text{SDE}}, \hat{\delta}^{\text{SDE}}, \hat{\sigma}_1^{\text{SDE}}, \hat{\sigma}_2^{\text{SDE}}, \hat{\sigma}^{\text{SDE}}).$$

Standard deviations for each parameter estimate were obtained through the Fisher information matrix. The corresponding estimations $\hat{Q}_P^{\text{SDE}}(t)$, $\hat{Q}_I^{\text{SDE}}(t)$ and

$\hat{S}^{\text{SDE}}(t)$ of $Q_P(t)$, $Q_I(t)$ and $S(t)$ were defined as the conditional expectations of $Q_P(t)$, $Q_I(t)$ and $S(t)$ given the observation $y_{0:n}$ and assuming the unknown parameter θ^{SDE} to be equal to $\hat{\theta}^{\text{SDE}}$:

$$\hat{Q}_P^{\text{SDE}}(t) = E_{\hat{\theta}^{\text{SDE}}}(Q_P(t)|y_{0:n}),\ \hat{Q}_I^{\text{SDE}}(t) = E_{\hat{\theta}^{\text{SDE}}}(Q_I(t)|y_{0:n}),\ \hat{S}^{\text{SDE}}(t) = E_{\hat{\theta}^{\text{SDE}}}(S(t)|y_{0:n}).$$

These extracted stochastic signals are computed by the smoother algorithm (see Appendix 3.7.2). The uncertainty of these extracted signals can be quantified by computing their 95% confidence intervals. The smoother algorithm provides a computation of the variances of the estimated signals :

$$Var_{\hat{\theta}^{\text{SDE}}}(Q_P(t)|y_{0:n}),\ Var_{\hat{\theta}^{\text{SDE}}}(Q_I(t)|y_{0:n}),\ Var_{\hat{\theta}^{\text{SDE}}}(S(t)|y_{0:n}).$$

As the processes $(S(t))$, $(Q_P(t))$ and $(Q_I(t))$ are Gaussian, exact confidence intervals of confidence level $1-\eta$ can be deduced. For example, we have :

$$IC_{1-\eta}(S_{\hat{\theta}^{\text{SDE}}}(t)) = [\hat{S}^{\text{SDE}}(t) \pm t_{1-\eta}\sqrt{(Var_{\hat{\theta}^{\text{SDE}}}(S(t)|y_{0:n}))}]$$

where $t_{1-\eta}$ is the quantile of order $1-\eta$ of a standardized centered Gaussian distribution.

3.3.3 Numerical implementation of the two estimation methods

The computation of the ODE mean squares and the SDE likelihood required the integral of the arterial input function, which was approximated using the trapezoidal method : the integral $\int_{t_0}^{t_i} AIF(x)dx$ over the interval $[t_0, t_i]$ was approximated by $\sum_{k=1}^{i}(t_k - t_{k-1})(AIF(t_k) + AIF(t_{k-1}))/2$. The minimization of (3.4) for the ODE model and the maximization of (3.6) for the SDE model were performed under the constraints :

$$0 \leq PS, F_T; \quad 0 \leq V_b, V_e \leq 100; \quad V_b + V_e \leq 100; \quad 1 \leq \delta.$$

The two optimizations were done using the Matlab$^{\text{TM}}$ function `fmincon` available in the toolbox `Optimization` release 2008b. The initial values were $F_T = 50$, $V_b = 10$, $PS = 10$, $V_e = 10$, $\delta = 20$ for the ODE optimization adding the extra values $\sigma = 5$, $\sigma_1 = 1$ and $\sigma_2 = 0.5$ for the SDE optimization. Computational times were about 4 seconds for the ODE and 20 seconds for the SDE methods, on a Mac Pro 3 Ghz with 7 Go RAM, Mac OS X 10.5 and Matlab$^{\text{TM}}$ Release 2008b.

3.4 Estimation on real data

DCE-MRI data from a series of normal female pelvises DCE-MRI were used to test and optimize the processing method (Thomassin-Naggara et al., ress). MRI sequences were performed on a MR 1.5T magnet (Siemens, Sonata, Erlangen, Germany) with a phased-array pelvic coil. For DCE-MRI, a dynamic contrast-enhanced T1-weighted gradient-echo sequence (2D FLASH) was acquired in the axial plane (TR/TE (ms), 27/2.24 ; flip angle, 80° ; slice thickness, 5mm ; number of slices, 3 ; excitations, 1 ; field of view, 400-200 mm ; interpolated matrix 256 x 134 ; BW, 300). Superior and inferior saturation bands of 90 mm width were added to avoid inflow effects and vessel artifacts. A dose of 0.1 mmol/kg^{-1} body weight of DOTA gadolinium (Guerbet, Aulnay, France) was injected intravenously via a power injector (Medrad, Maastricht, The Netherlands) at a rate of 2 ml.s^{-1}, and flushed by 20 ml of saline. Images were obtained at 2.4 second intervals for a total of 320 seconds after injection, yielding a total of 130 time frames.

Four datasets were analyzed. The ODE and SDE estimations were successively applied to the signals to estimate the parameters and the associated predicted concentrations. The ODE and SDE residuals were computed as the difference between the observations $y_{0:n}$ and the predictions of the corresponding model. The partial autocorrelation of these residuals were also computed to compare the quality of the two model predictions. The k-th partial autocorrelation is obtained by computing the correlation between y_0 and y_k given $(y_1, \ldots, y_{k-1})$.

The first dataset is summarized in Table 3.1 and Figure 3.2. For this voxel, the ODE and SDE estimates of F_T, V_b, PS, V_e and δ were identical as well as the contrast agent quantity predictions.

For dataset 2, the ODE and SDE estimations were not meaningfully different as, for each parameter, the 95% confidence interval contained both estimated values. However, the predictions $\hat{Q}_P$, $\hat{Q}_I$ and $\hat{S}$ obtained by the ODE and the SDE estimations look quite different (Figure 3.3) : SDE estimation achieves a better fit (*e.g.* around times 90 and 130). Figure 3.3 shows the first 15 partial autocorrelations for ODE (left) and SDE (right) residuals. The partial autocorrelations are lower for SDE. Indeed, the 90% test of zero autocorrelation (red dashed lines) detected the 6-th ODE autocorrelation to be non-zero while all the SDE autocorrelations were accepted to be zero.

For dataset 3, the ODE and SDE estimations were statistically different at least for one parameter (Table 3.3) even though the predictions $\hat{Q}_P$, $\hat{Q}_I$ and $\hat{S}$ look similar (Figure 3.4). Nevertheless, there were differences between the two methods : the SDE estimated extra-vascular volume V_e ($\hat{V}_e^{\text{SDE}} = 45.9$) was meaningfully larger than the ODE estimate ($\hat{V}_e^{\text{ODE}} = 35.8$). Figure 3.4 shows that the

parameters		ODE model	SDE model
F_T	(ml min^{-1} 100 ml^{-1})	48.7 (2.5)	48.7 (3.0)
V_b	(%)	40.5 (5.3)	40.5 (5.9)
PS	(ml min^{-1} 100 ml^{-1})	13.3 (4.3)	13.3 (4.8)
V_e	(%)	29.4 (2.4)	29.4 (2.8)
δ	(s)	6.0 (0.7)	6.0 (0.6)
σ		8.02 (0.2)	7.86 (1.0)
σ_1		-	$< 10^{-3}$ (0.6)
σ_2		-	$< 10^{-3}$ (0.2)

TABLE 3.1 – Estimated parameters for the first dataset, using the ODE and the SDE models. The values in parenthesis are the standard deviations evaluated using a numerical computation of the Fisher Information matrix. For predictions, see Figure 3.2.

parameters		ODE model	SDE model
F_T	(ml min^{-1} 100 ml^{-1})	78.6 (1.7)	78.0 (2.8)
V_b	(%)	39.5 (3.5)	40.1 (4.7)
PS	(ml min^{-1} 100 ml^{-1})	10.1 (4.5)	9.37 (1.3)
V_e	(%)	14.9 (1.8)	14.9 (5.5)
δ	(s)	10.6 (0.9)	10.6 (0.9)
σ		9.10 (0.2)	8.42 (1.1)
σ_1			1.00 (0.7)
σ_2			$< 10^{-3}$ (0.6)

TABLE 3.2 – Estimated parameters for dataset 2, using the ODE and the SDE models. The values in parenthesis are the standard deviations evaluated using a numerical computation of the Fisher Information matrix. For predictions, see Figure 3.3.

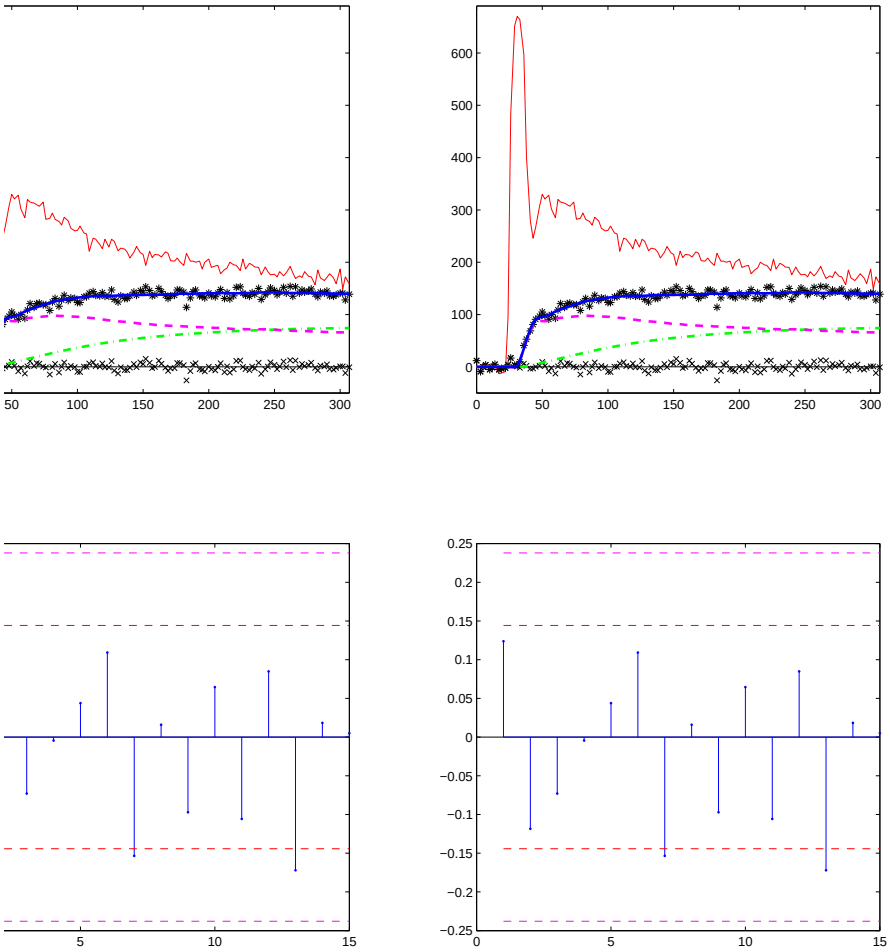

FIGURE 3.2 – Predictions for the first dataset, obtained with the ODE model (left) and the SDE model (right) : black stars ($*$) are the tissue observations (y_i), the AIF observations are represented by the red line, crosses ($\times$) are the residuals. The plain blue, dashed pink and dash-dotted green lines are respectively the predictions for $S(t)$, $Q_P(t)$ and $Q_I(t)$. Estimated parameters are from Table 3.1.

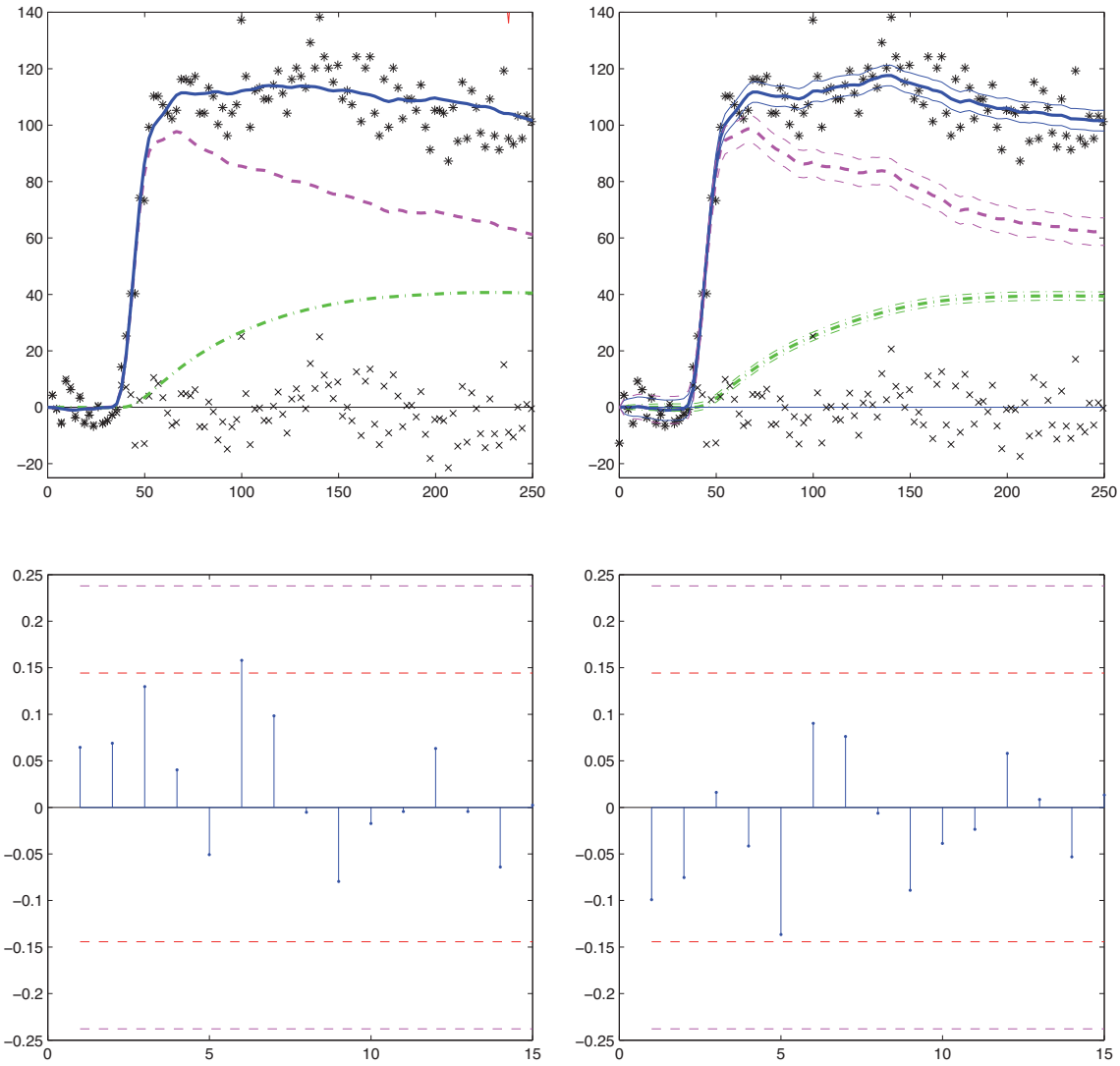

FIGURE 3.3 – Top two figures : predictions for dataset 2, obtained with the ODE model (left) and the SDE model (right) : black stars ($*$) are the tissue observations (y_i), crosses ($\times$) are the residuals. The plain blue, dashed pink and dashdotted green lines are respectively the predictions for $S(t)$, $Q_P(t)$ and $Q_I(t)$. For the SDE model, each prediction curve is surrounded by its 95% confidence intervals. Bottom two figures : first 15 partial autocorrelations of the residuals obtained with the ODE model (left) and the SDE model (right). The horizontal red dashed lines provide the 90% confidence interval around 0 for each partial autocorrelation. The horizontal violet dashed lines provide the Bonferroni confidence interval of the test that all partial autocorrelations are 0 together. The norm of these partial autocorrelations is 0.28 for the ODE estimation and 0.26 for the SDE estimation. Estimated parameters are from Table 3.2.

parameters		ODE model	SDE model
F_T	(ml min^{-1} 100 ml^{-1})	21.2 (3.3)	20.5 (4.0)
V_b	(%)	24.3 (3.7)	25.2 (6.0)
PS	(ml min^{-1} 100 ml^{-1})	5.44 (0.9)	4.87 (1.3)
V_e	(%)	35.8 (2.8)	45.9 (4.3)
δ	(s)	17.9 (1.9)	19.4 (2.3)
σ		6.68 (0.1)	6.23 (0.8)
σ_1			0.58 (0.4)
σ_2			$< 10^{-3}$ (0.4)

TABLE 3.3 – Estimated parameters for dataset 3, using the ODE and the SDE models. The values in parenthesis are the standard deviations evaluated using a numerical computation of the Fisher Information matrix. For predicted curves, see Figure 3.4.

SDE prediction $\hat{S}^{\text{SDE}}$ is non-zero for the first instants. This should not be interpreted as a default of the SDE estimation since the confidence interval curves are consistent with the absence of contrast agent inside the voxel for the first instants. Here, the flexibility of the SDE estimation enables to compensate for a probable shift in the baseline estimation. Furthermore, the ODE residuals showed correlations indicating a poor quality of the ODE estimation. The residual correlation structure is also illustrated in Figure 3.4 where the first 15 partial autocorrelation of the ODE and SDE residuals are plotted. The third ODE partial autocorrelation was detected to be non-zero at a level of 90% contrary to the third SDE partial autocorrelations.

For dataset 4, the SDE predicted quantity of contrast agent in the interstitial compartment $\hat{Q}_I^{\text{SDE}}(t)$ was always null ($\hat{Q}_I^{\text{SDE}}(t) = 0 \quad \forall t \geq 0$) while the ODE prediction $\hat{Q}_I^{\text{ODE}}(t)$ was not (Figure 3.5). The ODE model detected exchanges inside the voxel between the two compartments. The ODE residuals were correlated, especially between times $t = 40$ and $t = 75$ contrary to the SDE residuals. The parameter estimates obtained by the ODE and the SDE models were different (Table 3.4). The SDE estimated blood volume ($\hat{V}_b^{\text{SDE}} = 53.5$) was meaningfully larger than the ODE estimate ($\hat{V}_b^{\text{ODE}} = 41.3$). The SDE estimated permeability surface product ($\widehat{PS}^{\text{SDE}} = 0.81$) was much smaller than the ODE estimate ($\widehat{PS}^{\text{ODE}} = 2.96$). As $\hat{V}_b^{\text{ODE}} + \hat{V}_e^{\text{ODE}} = 100$, the ODE estimation stopped at a boundary of the optimization domain (see Section 3.3.3). This led us to investigate the analysis further. By computing the ratio Q_P/AIF, we suspected that the ODE estimation was strongly influenced by the last observation times. Therefore we

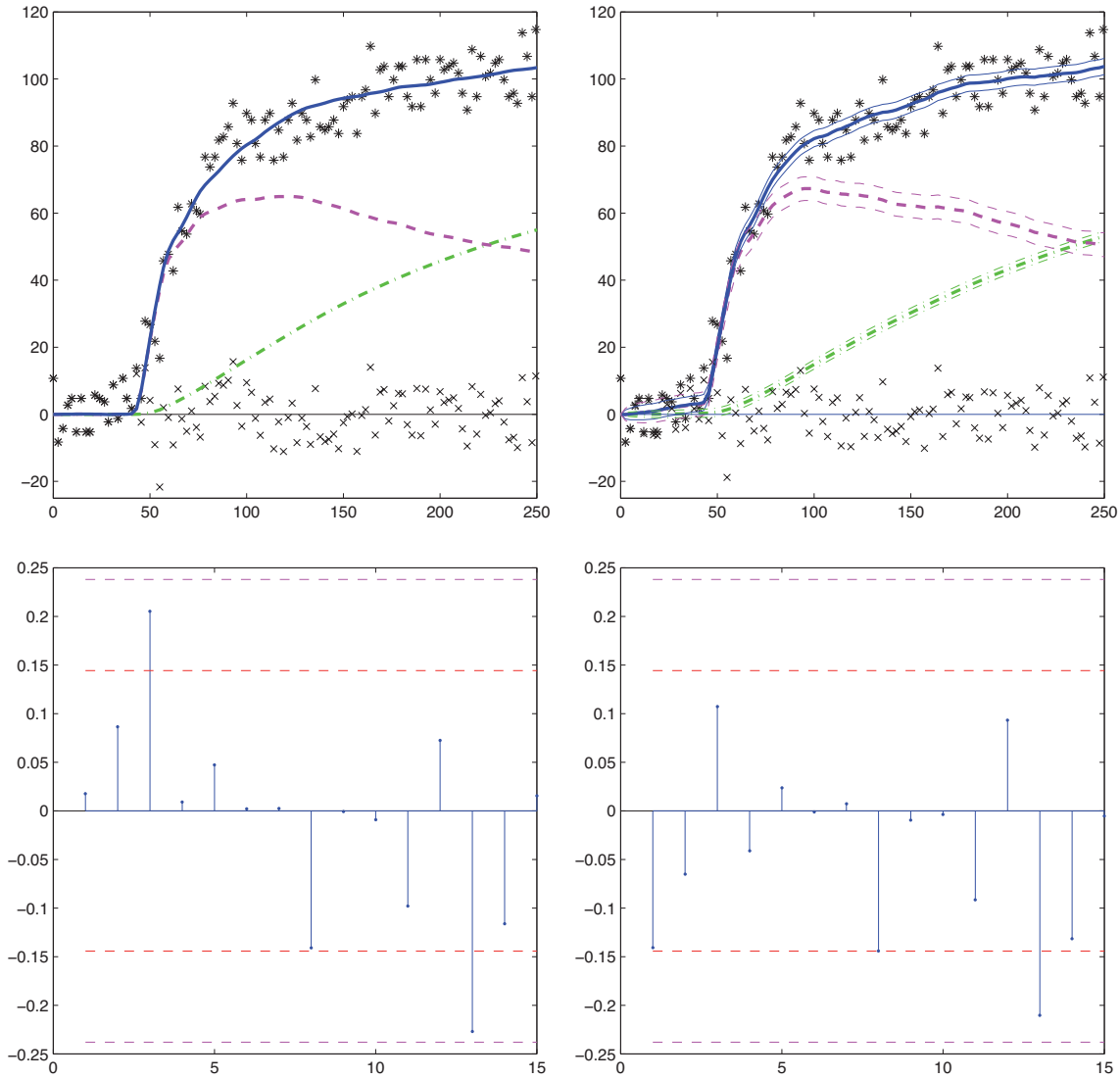

FIGURE 3.4 – Top two figures : predictions for dataset 3, obtained with the ODE model (left) and the SDE model (right) : black stars ($*$) are the tissue observations (y_i), crosses ($\times$) are the residuals. The plain blue, dashed pink and dash-dotted green lines are respectively the predictions for $S(t)$, $Q_P(t)$ and $Q_I(t)$. For the SDE model, each prediction curve is surrounded by its 95% confidence intervals. Bottom two figures : first 15 partial autocorrelations of the residuals obtained with the ODE model (left) and the SDE model (right). The horizontal red dashed lines provide the 90% confidence interval around 0 for each partial autocorrelation. The horizontal violet dashed lines provide the Bonferroni confidence interval of the test that all partial autocorrelations are 0. The norm of these partial autocorrelations is 0.39 for the ODE estimation and 0.37 for the SDE estimation. Estimated parameters are from Table 3.3.

	ODE model	SDE model	ODE without last 3 times	SDE without last 3 times	ODE without last 5 times	SDE without last 5 times
F_T	24.6 (2.1)	20.0 (3.8)	32.4	19.9	20.3	20.0
V_b	41.3 (3.3)	53.5 (8.0)	6.6	54.0	52.9	53.3
PS	2.96 (0.8)	0.81 (12)	43.2	0.38	0.04	0.23
V_e	58.7 (2.0)	0.04 (2.6)	27.9	0.02	0.002	0.01
δ	10.5 (1.5)	9.68 (3.2)	9.5	9.66	7.49	9.67
σ	7.55 (0.1)	6.51 (1.0)	8.4	6.46	8.19	6.38
σ_1		1.22 (0.7)		1.27		1.28
σ_2		0.02 (15)		0.02		0.02

TABLE 3.4 – Estimated parameters for dataset 4, using the ODE model (column 2), the SDE model (column 3) and using the ODE and SDE models after removing the last 3 (columns 4 and 5) and the last 5 (column 6 and 7) observations. The values in parenthesis are the standard deviations evaluated using a numerical computation of the Fisher Information matrix. The differences in parameters estimation by ODE influence drastically the enhancement curves (Figure 3.5).

removed the last 3 (and then the last 5) observations. While the SDE estimation remained stable when removing observations (up to changes in the last digits), the ODE estimation changed totally. The results with the last 3 or 5 observations removed are added in Table 3.4. Estimated parameters by ODE with last 5 data points removed (Table 3.4, last column) became close to estimated parameters by SDE with all data points (Table 3.4, 2nd column). Figure 3.5 shows enhancement curves corresponding to the estimated parameters of Table 3.4. Top curves are obtained with all data points by the ODE model (left) and the SDE model (right). In the SDE curves, the major part of the enhancement comes from the plasma while the interstitial enhancement is close to zero. Conversely, the ODE curves show enhancement from both plasma (dashed pink) and interstitium (dashdotted green). The middle curves are obtained by the ODE model (left) and the SDE model (right) without the last 3 data points. The bottom curves are obtained by the ODE model (left) and the SDE model (right) without the last 5 data points (right). On the middle ODE curves, there is an inversion of curves (compared with the top left curves) : the major part of enhancement is due to interstitium. On the bottom ODE curves, another inversion appears : enhancement is only due to plasma as with the SDE method. This illustrates the poor stability of the ODE method.

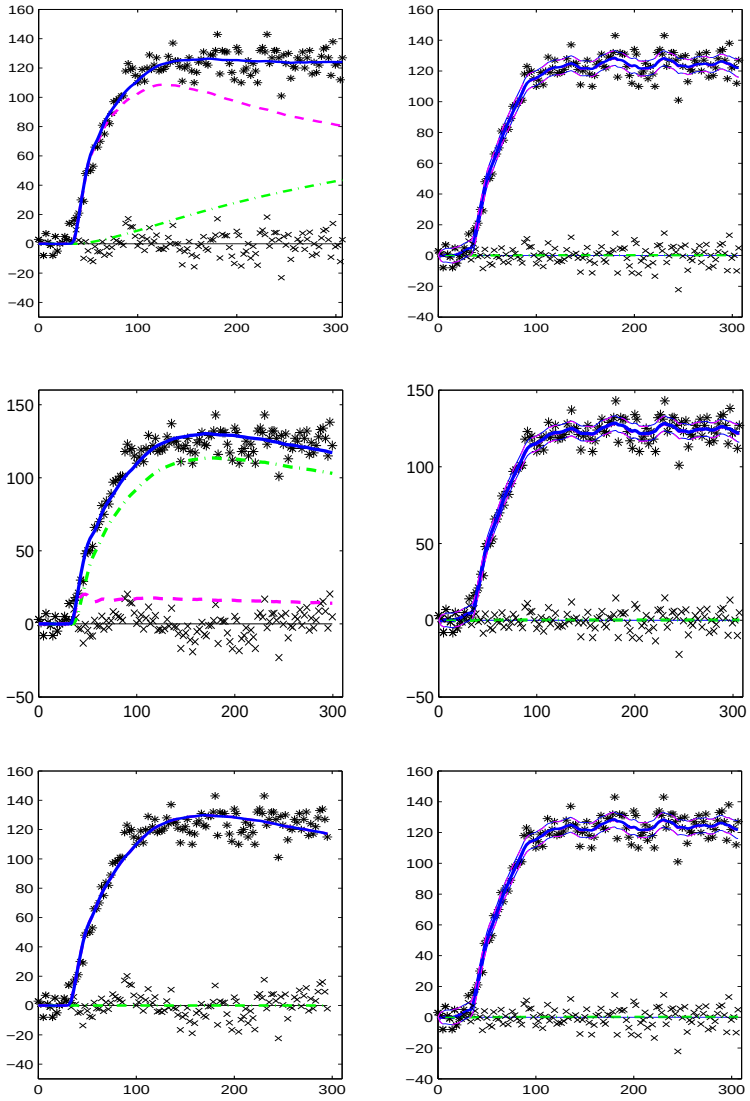

FIGURE 3.5 – Top figures : predictions for dataset 4, obtained with the ODE model (left) and the SDE model (right) : black stars ($*$) are the tissue observations (y_i), crosses ($\times$) are the residuals. The plain blue, dashed pink and dashdotted green lines are respectively the predictions for $S(t)$, $Q_P(t)$ and $Q_I(t)$. For the SDE model, each prediction curve is surrounded by its 95% confidence intervals. Middle figures : predictions obtained with the ODE model (left) and the SDE model (right) removing the last 3 observations. Bottom figures : predictions obtained with the ODE model (left) and the SDE model (right) removing the last 5 observations (right). The adjustment differences come from the differences of parameter estimations (Table 3.4).

3.5 Simulated study

Simulations were performed to illustrate the properties of the estimators of the ODE and SDE models. An ideal AIF was built artificially (red curve on the left sub-figure in Figure 3.6). The parameter values used for simulations were $F_T = 70$ ml min^{-1} 100 ml^{-1}, $V_b = 20$ %, $PS = 15$ ml min^{-1} 100 ml^{-1}, $V_e = 15$ %, $\delta = 10$ s and $\sigma = 7$ for the measurement error.

First, a hundred data sets were simulated using the ODE model which corresponds to $\sigma_1 = \sigma_2 = 0$ in the SDE model. Parameters were estimated by the two methods. As shown in Table 3.5 (top), estimations by both methods were identical. Not surprisingly, the parameters σ_1 and σ_2 had very small SDE estimations. Then, a hundred data sets were simulated using the SDE model with $\sigma_1 = \sigma_2 = 2$. Results (Table 3.5, middle part) show a clear reduction of bias and standard deviations enlightening the advantage of the SDE estimation method. At last, a hundred data sets were simulated with the SDE model with $\sigma_1 = \sigma_2 = 1$ and $PS = 1$ (Table 3.5, bottom part), which corresponds to the case where the exchange between plasma and interstitium is small. The ODE estimation of the parameters F_T and V_e were more biased than the SDE estimates. Again the SDE estimation outperformed the ODE estimation.

3.6 Discussion

To take into account noises which induce instability in microvascularization parameter extraction in DCE-MRI, a stochastic version of a two-compartment model described by stochastic differential equations was introduced. On voxels of normal female pelvises DCE-MRI data, both the stochastic and the standard deterministic two-compartment model (given by ordinary differential equations) were implemented. The stochastic model generally led to a more satisfactory description of enhancement curves and provided a more robust parameter estimation method. In the special case of aberrant data (e.g. due to patient movements or measurement disturbances), the SDE method outperformed the ODE method. When the permeability surface product (PS) was small, the ODE estimation method was unstable and gave different results when removing a few data points. Conversely, the stochastic method remained stable with or without these data. This proves that the stochastic method is less sensitive to special or aberrant data points.

These results were confirmed by the simulation study. For a given set of parameters $(F_T, V_b, PS, V_e, \delta, \sigma_1, \sigma_2, \sigma)$ and a given artificial arterial input function, a hundred trajectories were simulated under the stochastic model. For each of

	F_T	V_b	PS	V_e	δ	σ	σ_1	σ_2
True parameters	70	20	15	15	10	7	0	0
ODE estimates	72.2	19.9	15.1	15.1	10	7.07	NA	NA
	(10)	(2.4)	(2.2)	(1.3)	(0.54)	(0.35)	NA	NA
SDE estimates	72.2	19.9	15.1	15.1	10	6.99	0.03	0.002
	(10)	(2.4)	(2.2)	(1.3)	(0.54)	(0.35)	(0.12)	(0.015)
True parameters	70	20	15	15	10	7	2	2
ODE estimates	67.4	22	18.8	22.4	9.89	12.4	NA	NA
	(16)	(8.7)	(16)	(18)	(1.3)	(1.7)	NA	NA
SDE estimates	71.9	21.7	15.3	17.3	9.96	6.88	2.54	1.22
	(15)	(6.6)	(9.8)	(13)	(1)	(0.51)	(0.69)	(0.84)
True parameters	70	20	1	15	10	7	1	1
ODE estimates	77.8	18.4	2.42	26.7	10.3	9.69	NA	NA
	(12)	(3.4)	(3)	(33)	(0.83)	(2.5)	NA	NA
SDE estimates	72.4	19.7	2.08	17	10	6.89	1.32	0.716
	(10)	(2.5)	(2.5)	(26)	(0.65)	(0.32)	(0.52)	(0.43)

TABLE 3.5 – Estimation results for three different sets of fixed parameters. Estimation when simulating under the ODE model (top part of the Table), when simulating under the SDE model ($\sigma_1 = \sigma_2 = 2$) (middle part of the Table) and when simulating with a small PS ($PS = 1$) under the SDE model ($\sigma_1 = \sigma_2 = 1$) (bottom part of the Table). Empirical means and standard deviations (in parenthesis) are computed for each estimated parameter from 100 simulated datasets analyzed by the ODE and the SDE estimations. The NA values express that the associated quantities are not available in the considered model.

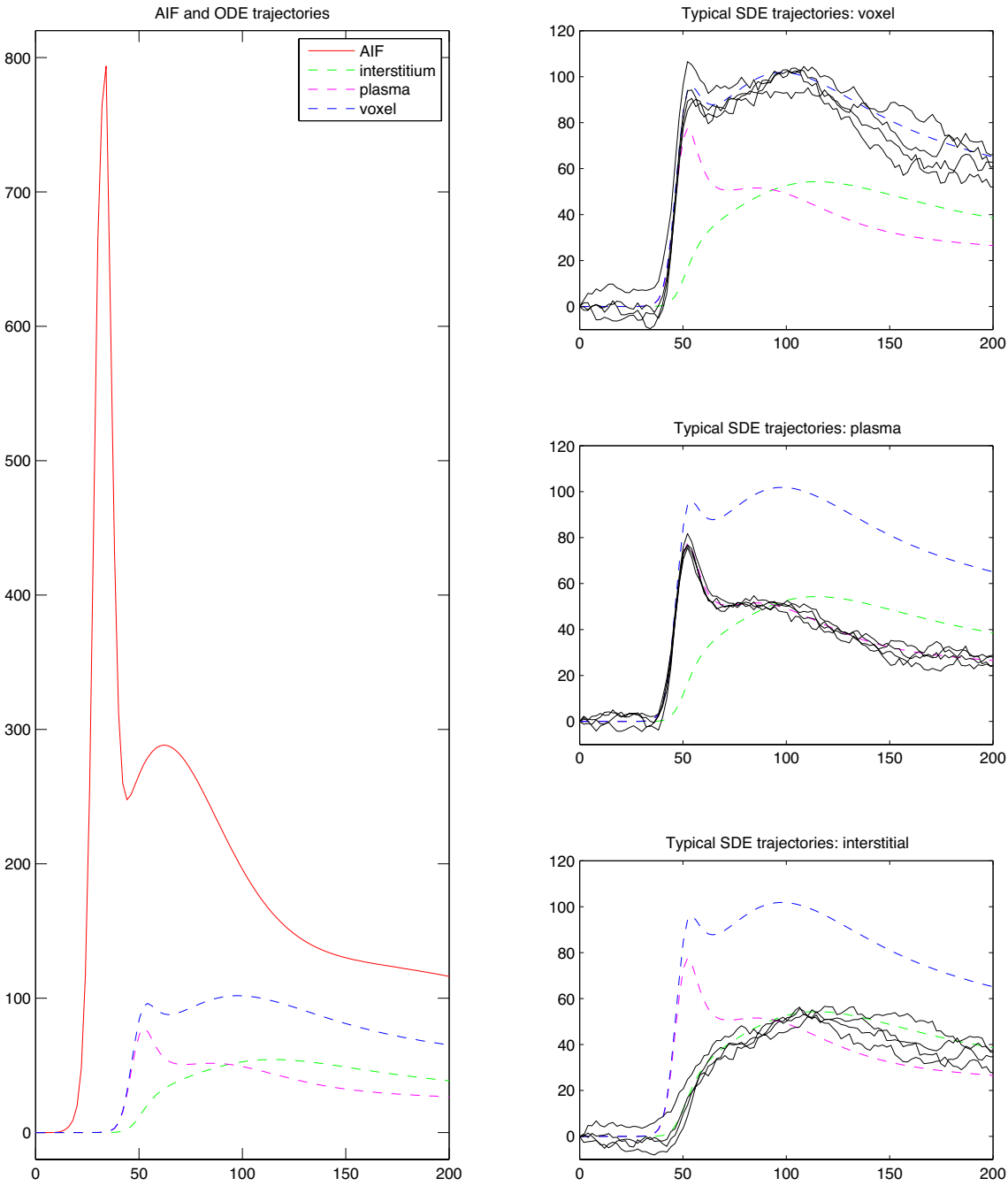

FIGURE 3.6 – Typical trajectories of the ODE and SDE models for the ideal case $\sigma = 0$ (no measurement error). An ideal AIF was built artificially (plain line on the left figure). Parameter values are $F_T = 70$ ml min^{-1} 100 ml^{-1}, $V_b = 20$ %, $PS = 25$ ml min^{-1} 100 ml^{-1}, $V_e = 15$ %, $\delta = 15$ s. The functions Q_P, Q_I, S of the ODE system (3.2) are given on the left figure by the magenta, green and blue curves, respectively. On the right figures, five typical realizations of the processes Q_I (bottom), Q_P (middle), S (top) of the SDE model (3.3) are plotted for $\sigma_1 = \sigma_2 = 1$.

the hundred simulated trajectories, the parameters were estimated by both the ODE and SDE methods. The means of estimated parameters were globally close to the true values for both methods, with smaller variances for the SDE method. The parameters PS and V_e were closer to the true values with the SDE method than with the ODE method. The estimation of V_b was very similar with both methods, whatever the PS value. The estimation of F_T was similar with both methods when PS was large. On the contrary, for small PS, the estimation of F_T became biased with the ODE method and not with the SDE method.

Tissue microcirculation assessment via DCE-MRI can be performed either in a descriptive manner (Buadu et al., 1996; Lucht et al., 2005; Tse et al., 2007), or by semi-quantitative analysis (time to peak, peak high, presence of wash-out, etc) (Kelcz et al., 2002; Fan et al., 2007; Nasel et al., 2000; Thomassin-Naggara et al., 2008), or by extracting quantitative parameters from models taking into account the arterial input function (Tofts, 1997; Brix et al., 2004; Pradel et al., 2003; Balvay et al., 2009). For clinical or biological use of microcirculation measurements by DCE-MRI, quantitative results are important for objective comparisons, especially when one wants to characterize lesions (tumor versus ischemia or inflammation), to test new drugs or follow treatments. It is therefore crucial to have robust and reliable results (Buckley, 2002). To improve robustness, several strategies were developed : 1/ use of simple pharmacokinetic models (Tofts, 1997), 2/ improvement of data recording (Cheng, 2008), 3/ preprocessing of data (Buonaccorsi et al., 2007; Delzescaux et al., 2001) or 4/ use of stochastic versions and more complex pharmacokinetic models.

Simple pharmacokinetic models are one-compartment models involving a small number of parameters (Tofts, 1997). They provide stable estimation results but have the drawback of using an oversimplified model. To improve data recording, it is essential to optimize the measurement devices and the protocol (measurement times and contrast agent doses) and to minimize motion artefacts. This however remains limited by external constraints. There are numerous preprocessing methods to improve signal-to-noise ratio : denoising by filtering or smoothing, spatial averaging over a manual region of interest or clustering methods (Calamante et al., 2003). The advantage of clustering is that it preserves dynamics homogeneity (Rozenholc et al., 2009). The other methods (especially the spatial averaging method) have the drawback of mixing different dynamics (for example, mixing necrotic and perfused voxels). During preprocessing, image registration is essential to correct patients' moves. At last, stochastic versions of physiological models provide a reliable way of using more complex models involving a larger number of parameters and closer to physiology than simpler models (Tofts, 1997). A combination of several of these approaches can be used.

The stochastic version of a classical multicompartment model is obtained by adding Brownian components to the deterministic model, leading to stochastic differential equations. Starting from a comprehensive two-compartment exchange model with four parameters (tissue blood perfusion, tissue blood volume, permeability surface product, interstitial volume) and the bolus arrival time (Balvay et al., 2009), we have built a SDE whose estimation method relies on maximum likelihood theory, maximum likelihood estimators being known to be the best estimators for a given statistical model (Favetto and Samson, 2010). SDEs have been recently applied for medical applications, to glucose dynamics (Picchini et al., 2008), neuron potential dynamics (Höpfner and Brodda, 2006) or pharmacokinetics (Ditlevsen et al., 2005; Donnet and Samson, 2008). These applications of SDEs show that stochastic versions of physiological models improve data fitting and stabilize parameter estimations. These results were confirmed in our study. Indeed, we obtained stable estimated parameters using a stochastic version of a complex parametric model while the deterministic model was giving results of high variability. The use of a stochastic model thus provides a serious improvement in data fitting and robustness in the estimation results even with a large number of parameters.

Yet, there are some limitations to our study. The two-compartment model does not take into account the blood propagation along capillaries, assumes instantaneous mixing into the interstitial compartment and equilibrium inside the interstitium of neighbor voxels. Moreover, the physiological parameters are assumed to be constant along time. Nevertheless, the good adjustment of enhancement curves, as measured by residuals, shows that this model is a reasonable physiological representation of tissue microcirculation. The model does not hold true for all kinds of tissues. For instance, in the liver, the sinusoids are irrigated by a dual vascular system (Cuenod et al., 2001; Materne et al., 2002) and in necrotic tissues without capillary network, the contrast enhancement is due to passive diffusion from adjacent irrigated tissues. However the present work could be extended by adapting models.

Some assumptions of the stochastic model may be criticized. For instance, constant variances for the Brownian motions may look unrealistic. It would be worth to extend this work to SDE with variance functions of times or of contrast agent quantities. This would be to the price of additional mathematical difficulties. This work could also be extended to the case of non-Gaussian observation errors. For a unidimensional Markov chain (X_i) observed with non Gaussian errors, Ruiz (1994) proposes a quasi-maximum likelihood estimator based on the Kalman filter and shows the normality asymptotic distribution of this estimator. This approach can be extended to our bidimensional model.

To conclude, this study shows that, in view of quantifying the tissue micro-circulation parameters, the stochastic approach makes it possible to reduce the instability observed with the deterministic two-compartment model. By taking into account the sources of variations in DCE-MRI data, the SDE approach provides a more robust parameter extraction.

Acknowledgments

This project was supported in 2008 by a grant "Bonus Qualité Recherche" (BQR) from Université Paris Descartes.

3.7 Appendix

3.7.1 The Kalman filter

In this section, we present the computation of the conditional expectation $E_{\theta^{\text{SDE}}}(y_i|y_{0:i-1})$ and variance $V_{\theta^{\text{SDE}}}(y_i|y_{0:i-1})$ that appear in Equation (3.7) using the Kalman filter Cappé et al. (2005); Favetto and Samson (2010). Let us denote

$$G = \begin{pmatrix} -\beta & \beta \\ \lambda & -k \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1 & \sigma_2 \\ 0 & \sigma_2 \end{pmatrix}, \quad F(t) = \begin{pmatrix} \alpha AIF(t) \\ 0 \end{pmatrix},$$

where $\alpha, \beta, \lambda, k$ are related to the biological parameters by the relations

$$\alpha = \frac{F_T}{1-h},\ \beta = \frac{F_T}{V_b(1-h)},\ \lambda = \frac{PS}{V_b(1-h)},\ k = \frac{PS}{V_b(1-h)} + \frac{PS}{V_e}.$$

The SDE system (3.3) can be rewritten in a matrix form as follows :

$$dU(t) = (GU(t) + F(t))dt + \Sigma dW(t), \quad U(0) = (0,0)' \tag{3.8}$$

where $U(t) = (S(t), Q_I(t))'$, $(W(t) = (W_1(t), W_2(t))', t \geq 0)$ is a standard 2-dimensional Brownian motion and X' denotes the transposed matrix of X. The observations are defined as $y_i = JU(t_i) + \sigma\epsilon_i$, where J is the line vector $(1,0)$. Solving (3.8) yields

$$U(t+h) = e^{Gh}U(t) + \int_t^{t+h} e^{G(t+h-s)}F(s)ds + \int_t^{t+h} e^{G(t+h-s)}\Sigma dW_s.$$

Hence the conditional distribution of $U(t_{i+1})$ given $U(t_i)$ is the Gaussian distribution $\mathcal{N}(A_{i+1}U(t_i) + B_{i+1}, R_{i+1})$ with

$$A_{i+1} = e^{G(t_{i+1}-t_i)},\ B_{i+1} = \int_{t_i}^{t_{i+1}} e^{G(t_{i+1}-s)}F(s)ds,\ R_{i+1} = \int_{t_i}^{t_{i+1}} e^{G(t_{i+1}-s)}\Sigma\Sigma' e^{G'(t_{i+1}-s)}ds.$$

Consider now that $t_i = i\Delta$ where $\Delta > 0$ is the discretization step, and denote $U_i = U(t_i)$. The Kalman algorithm enables to compute recursively the mean and the covariance of the conditional distributions

$$p(U_i|y_{0:i}) = \mathcal{N}(\hat{U}_i, P_i), \quad p(U_i|y_{0:i-1}) = \mathcal{N}(\hat{U}_i^-, P_i^-),$$

with the iterations

$$\begin{array}{rcll} \hat{U}_i^- = A_i\hat{U}_{i-1} + B_i & , & P_i = A_iP_{i-1}A_i' + R_i & i \geq 1 \\ \hat{U}_i = \hat{U}_i^- + K_i(y_i - J\hat{U}_i^-) & , & P_i = (I - K_iJ)P_i^- & i \geq 1 \end{array}$$

where $K_i = P_i^- J'(JP_i^- J' + \sigma^2)^{-1}$. It follows that the conditional distribution of y_i given $y_{0:i-1}$ is $\mathcal{N}(J\hat{U}_i^-, JP_i^- J' + \sigma^2)$.

3.7.2 The Kalman smoother algorithm

The Kalman smoother is calculated recursively with a forward-backward algorithm (see *e.g.* Cappé et al. (2005)). The forward algorithm is the classical Kalman filter which computes $M_{i|0:i-1} = \hat{U}_i^- = \mathbb{E}(U_i|y_{0:i-1})$, $\Sigma_{i|0:i-1} = P_i^- = Var(U_i|y_{0:i-1})$, $M_{i|0:i} = \hat{U}_i = \mathbb{E}(U_i|y_{0:i})$ and $\Sigma_{i|0:i} = P_i = Var(U_i|y_{0:i})$. Then, in order to calculate $M_{i|0:n} = \mathbb{E}(U_i|y_{0:n})$, $\Sigma_{i|0:n} = Var(U_i|y_{0:n})$, $\Sigma_{i-1,i|0:n} = Cov(U_{i-1}, U_i|y_{0:n})$, one performs the set of backward recursions $i = n, n-1, \ldots, 1$:

$$\begin{aligned} S_{i-1} &= \Sigma_{i-1|0:i-1}A_i'(\Sigma_{i|0:i-1})^{-1} \\ M_{i-1|0:n} &= M_{i-1|0:i-1} + S_{i-1}(M_{i|0:n} - M_{i|0:i-1}) \\ \Sigma_{i-1|0:n} &= \Sigma_{i-1|0:i-1} + S_{i-1}(\Sigma_{i|0:n} - \Sigma_{i|0:i-1})J_{i-1}' \end{aligned}$$

To calculate $\Sigma_{i-1,i|0:n}$, we have

$$\Sigma_{n-1,n|0:n} = (I - K_nJ)A_i\Sigma_{n-1|0:n-1}$$

and the following backward recursions, for $i = n-1, n-2, \ldots, 1$

$$\Sigma_{i-1,i|0:n} = \Sigma_{i|0:i}S_{i-1}' + S_i(\Sigma_{i,i+1|0:n} - A_i\Sigma_{i|0:i})S_{i-1}'.$$

Hence this gives a confidence interval for (U_t) at times t_i, based on $M_{i|0:n}$ and $\Sigma_{i|0:n}$. See Favetto and Samson (2010) for more details.

3.7.3 Computation of the Fisher Information Matrix

Theoretical properties of the maximum likelihood estimators were studied in Favetto and Samson (2010) when (y_i) is stationary. In this case, setting $l_{0:n}(\theta) =$

$\log L(\theta, y_{0:n})$ the log-likelihood of the observations $y_{0:n}$, we have, under some regularity conditions, $\mathbb{P}_\theta - a.s.$

$$\lim_{n\to\infty}\left(-\frac{1}{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}l_{0:n}(\theta)\right) = I(\theta)$$

where $I(\theta)$ is the asymptotic Fisher Information matrix (see Brockwell and Davis (1991) for theoretical developpements).

The computation of $\frac{\partial^2}{\partial\theta_i\partial\theta_j}l_{0:n}(\theta)$ is performed with a finite-difference method :

$$\frac{\partial^2 l_{0:n}}{\partial\theta_i\partial\theta_j}(\theta) \approx \frac{l_{0:n}(\theta+\Delta\theta_i+\Delta\theta_j)+l_{0:n}(\theta-\Delta\theta_i-\Delta\theta_j)-l_{0:n}(\theta+\Delta\theta_i-\Delta\theta_j)-l_{0:n}(\theta-\Delta\theta_i+\Delta\theta_j)}{4\Delta\theta_i\Delta\theta_j}.$$

The convergence rate of the method is improved by the Richardson method (see Stoer and Bulirsch (1993) for details). Hence, we can derive an asymptotic confidence interval for θ computing $I_n(\theta) = -\frac{1}{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}l_{0:n}(\theta)$, due to the associated Central Limit Theorem.

3.7.4 Model shapes on simulated data

The behavior of the ODE and SDE models without measurement noise ($\sigma = 0$) are illustrated on some simulations. An ideal AIF was built artificially (red curve on the left figure in Figure 3.6). Given the numerical values of $F_T, V_b, PS, V_e, \delta$, the ODE system is fully determined by the AIF and the functions Q_P, Q_I, S are determinist. The functions Q_P, Q_I, S of the ODE system (3.2) are given in the left figure by the magenta, green and blue curves, respectively, for the values $F_T = 70$ ml min^{-1} 100 ml^{-1}, $V_b = 20$ %, $PS = 25$ ml min^{-1} 100 ml^{-1}, $V_e = 15$ %, $\delta = 15$ s. In the right figures, five typical realizations of the processes Q_I (bottom), Q_P (middle), S (top) of the SDE model (3.3) are plotted for $\sigma_1 = \sigma_2 = 1$. The SDE trajectories are clearly centered on their associated ODE trajectories. This is a known property : the ODE model provides an average model for the SDE. These trajectories show that the SDE model can explain excursions from the ODE model and can also offer a good flexibility to take into account errors in baseline measurements.

Deuxième partie

Estimation des paramètres d'une diffusion bruitée pour des données haute fréquence

Chapitre 4

Parameter estimation by contrast minimization for noisy observations of a diffusion process

Abstract

We consider the estimation of unknown parameters in the drift and diffusion coefficients of a one-dimensional ergodic diffusion X when the observation Y is a discrete sampling of X with an additive noise, at times $i\delta, i = 1 \dots N$. Assuming that the sampling interval tends to 0 while the total length time interval tends to infinity, we prove limit theorems for functionals associated with the observations, based on local means of the sample. We apply these results to obtain a contrast function. The associated minimum contrast estimators are shown to be consistent. We provide an illustration on simulated and real data from neuronal activity.

4.1 Introduction

Statistical inference for continuous time models based on high frequency data has been the subject of a huge number of recent papers. On one hand, continuous time stochastic processes are increasingly used for modelling purposes. On the other hand, such kind of data is now commonly available in various fields of applications whether in finance or in biology and medicine.

Among continuous time models, one-dimensional diffusion processes have received a lot of attention. In particular, diffusion models have been introduced in the studies of neuronal activity (see *e.g.* Ditlevsen and Lansky (2005), Höpfner and Brodda (2006) and the references given in these papers). More precisely, let (X_t) be given by the stochastic differential equation :

$$dX_t = b(X_t, \kappa)dt + \sigma(X_t, \lambda)dB_t, \qquad X_0 = \eta \tag{4.1}$$

with B a standard Wiener process and η a random variable independent of B, and $b(., \kappa), \sigma(., \lambda)$ real valued functions, defined on $\mathbb{R}$, depending on unknown parameters $(\kappa, \lambda) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Suppose that, for some positive sampling interval δ, a sample $(X_{i\delta}, i \leq N)$ is observed and that is required to estimate $\theta = (\kappa, \lambda)$. As the exact likelihood of such observation is generally intractable, other methods have been developped to obtain explicit estimators : minimum contrast estimators, estimating equations, simulation based methods, ...See *e.g.* Florens-Zmirou (1989), Yoshida (1992), Kessler (1997), Bibby and Sørensen (1995), Sørensen (2009), Genon-Catalot (1990), Genon-Catalot and Jacod (1993), Pedersen (1995b), Pedersen (1995a), Aït-Sahalia (2002).

More recenty, especially in the case of high frequency data, other kinds of observations have been investigated among which the case of noisy observations. Suppose that, instead of observing exactly $X_{i\delta}$, the observation at time $i\delta$ is given by

$$Y_{i\delta} = X_{i\delta} + \rho\varepsilon_{i\delta} \tag{4.2}$$

with $(\varepsilon_{i\delta}, i \geq 0)$ a sequence of i.i.d. random variables, satisfying $\mathbb{E}(\varepsilon_{i\delta}) = 0$, $\mathbb{E}((\varepsilon_{i\delta})^2) = 1$, independent of the process (X_t). This kind of model takes into account measurement errors or, in the case of financial data, the so-called microstructure noise (see *e.g* Zhang et al. (2005), Jacod et al. (2009)). It provides a model fitted to data which show non Markovian features.

The exact likelihood of $(Y_{i\delta}, i \leq N)$ given by (4.1)-(4.2) is generally intractable except for few models (essentially for Gaussian diffusions with additive Gaussian noise, see *e.g.* Cappé et al. (2005), Pedersen (1994), Favetto and Samson (2010)). For data within a fixed length-time interval ($\delta = \delta_N = \frac{1}{N}$, $N\delta_N = 1$), estimation for a general diffusion with additive Gaussian noise is investigated in Gloter

and Jacod (2001). The authors use a contrast method. Only diffusion coefficient parameters can be consistently estimated in this case.

In this paper, we study observations given by (4.1)-(4.2) where $\delta = \delta_N \to 0$ while $N\delta_N \to \infty$, under ergodic properties for the hidden diffusion X and propose consistent estimators of both the drift and diffusion coefficient parameters (κ, λ). The noise distribution is unknown, the variance ρ^2 of the noise term may be known or unknown and we assume either that ρ is fixed or that $\rho = \rho_N \to 0$.

Our starting idea is to reduce the influence of the noise by splitting the sample into sub-samples and taking empirical means of the sub-samples. More precisely, we split the sample into k blocks of size p, with $N = pk$, where $p = p_N$ and $k = k_N$ tend to infinity with N. Then, setting $\Delta_N = p_N\delta_N$ where p_N and δ_N are chosen such that $\Delta_N \to 0$, we build the empirical mean of the j^{th} block :

$$Y_\bullet^j = X_\bullet^j + \rho_N \varepsilon_\bullet^j, \quad j = 0, 1 \dots k_N - 1, \tag{4.3}$$

where, for $Z = Y, X, \varepsilon$,

$$Z_\bullet^j = \frac{1}{p_N} \sum_{i=0}^{p_N - 1} Z_{j\Delta_N + i\delta_N}. \tag{4.4}$$

Thus, Δ_N defines a coarser sampling interval than δ_N, still tending to 0 while $N\delta_N = k_N\Delta_N \to \infty$. The empirical mean $X_\bullet^j$ is close to the mean $\bar{X}_j = \frac{1}{\Delta_N}\int_{j\Delta_N}^{(j+1)\Delta_N} X_s ds$ of X over $[j\Delta_N, (j+1)\Delta_N]$, which, in turn, is close to $X_{j\Delta_N}$. The variance of $\varepsilon_\bullet^j$ is reduced by a factor $\frac{1}{p_N}$.

Our statistical procedure is based on the k_N- sample $(Y_\bullet^j, j = 0 \dots k_N - 1)$ and follows a scheme analogous to the one in Gloter (2006). We study the differences $Y_\bullet^j - X_{j\Delta_N}$ (Proposition 4.2) and prove a regression type relation for the $Y_\bullet^j$'s (Proposition 4.3) which is the basement of the statistical applications. These results allow us to prove limit theorems for the variation and the quadratic variation of $(Y_\bullet^j)$ (Theorems 4.1 and 4.2). We precise the adequate calibration of δ_N and p_N for the limit theorems to hold. Then we introduce two different contrasts : according to the rates of p_N, δ_N, the noise variance has or has not to be taken into account. We set $\delta_N = p_N^{-\alpha}$, with $1 < \alpha \leq 2$. For $\alpha = 2$ and $\rho_N = \rho$, the value ρ^2 appears in the contrast definition. In each case, we prove the consistency of the associated minimum contrast estimators. As could be expected, we have to deal with two-rate contrasts, which indicate that drift parameters estimators must have rate $\sqrt{k_N\Delta_N}$, while diffusion coefficient parameters estimators must have rate $\sqrt{k_N}$. The study of the asymptotic distributions of the minimum contrast estimators is postponed to a further paper. Estimators are implemented on simulated data and on real data of neuronal activities provided by Idoux et al. (2006).

The paper is organised as follows. In Section 4.2, we give our notations and assumptions on the model. Section 4.3 is devoted to the small sample properties of the empirical means sample $(Y_\bullet^j)$ and Section 4.4 to uniform convergence in probability results. In Section 4.5, we introduce the contrasts and prove the consistency of the estimators. We also deal with the case $\rho_N = \rho$ unknown and prove that ρ^2 can be replaced by an estimator in the contrast formula.

Section 4.6 is devoted to examples and numerical results. For several models of hidden diffusions, we implement our estimators on simulated data for different choices of (N, δ_N, p_N) and of the noise level. We illustrate the estimation method on the set of neuronal data for one model of diffusion with estimated noise level. Section 4.7 contains some concluding remarks. Proofs are gathered in Section 4.8, and some auxiliary results are recalled in the Appendix.

4.2 Assumptions and Notations

Consider the one-dimensional stochastic differential equation

$$dX_t = b(X_t, \kappa_0)dt + \sigma(X_t, \lambda_0)dB_t, \qquad X_0 = \eta \tag{4.5}$$

where B is a standard Brownian motion and η is a real valued random variable independent of B. The functions $b(x, \kappa)$ and $\sigma(x, \lambda)$ are respectively defined on $\mathbb{R} \times \Theta_1$ and $\mathbb{R} \times \Theta_2$ where Θ_1 (resp. Θ_2) is a compact convex subset of $\mathbb{R}^{d_1}$ (resp. $\mathbb{R}^{d_2}$). For simplicity of notations, in proofs, we assume that $d_1 = d_2 = 1$. We denote by $\theta_0 = (\kappa_0, \lambda_0)$ the true value of the parameter and assume that $\theta_0 \in \overset{\circ}{\Theta}$ where $\Theta = \Theta_1 \times \Theta_2$.

From now on, we set $b(x) = b(x, \kappa_0)$ and $\sigma(x) = \sigma(x, \lambda_0)$ and make classical assumptions on functions b and σ ensuring that (4.5) admits an unique strong solution $(X_t)_{t\geq 0}$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and that this solution is positive recurrent on $\mathbb{R}$.

(A1) Functions b and σ belong to $\mathcal{C}^2(\mathbb{R})$, $\sigma(x) > 0$ for all x, and there exists $c > 0$ such that for all $x \in \mathbb{R}$:

$$|b(x)| + |b'(x)| + |b''(x)| \leq c(1 + |x|),$$
$$\sigma(x) + |\sigma'(x)| + |\sigma''(x)| \leq c(1 + |x|).$$

(A2) For $x_0 \in \mathbb{R}$, let $s(x) = \exp(-2\int_{x_0}^x \frac{b(u)}{\sigma^2(u)}du)$ denote the scale density and $m(x) = \frac{1}{\sigma^2(x)s(x)}$ the speed density. Assume $\int_{-\infty} s(x)dx = \int^{+\infty} s(x)dx = \infty$ and $\int_{-\infty}^{+\infty} m(x)dx = M < \infty$.

(A3) Let $\nu_0(dx) = \frac{1}{M}m(x)dx$. For all $k > 0$, ν_0 admits a finite moment of order k.

(A4) For all $k > 0$, $\sup_{t\geq 0} \mathbb{E}(|X_t|^k) < \infty$.

(A5) The common distribution of the random variables $\varepsilon_{i\delta_N}$ admits a 8th order moment, and is symmetric. We set $m_1 = \mathbb{E}(|\varepsilon_{i\delta_N}|), m_4 = \mathbb{E}((\varepsilon_{i\delta_N})^4), m_8 = \mathbb{E}((\varepsilon_{i\delta_N})^8)$.

(B1) $\rho_N = \rho > 0$.

(B2) $\rho_N \to 0$ when $N \to \infty$.

Assumption **(A1)** implies that (4.1) admits a unique strong solution on $\mathbb{R}$. Under **(A1)** and **(A2)**, ν_0 is the unique invariant probability of (4.5) and (X_t) satisfies the classical ergodic theorem (see *e.g.* Rogers and Williams (2000))

$$\forall f \in L^1(d\nu_0), \quad \frac{1}{T}\int_0^T f(X_s)ds \underset{T\to\infty}{\longrightarrow} \nu_0(f) \quad a.s.$$

Moreover, under Assumption **(A1)**, for all $k \geq 1$, there exists a constant $c(k)$ such that, for all $t \geq 0$:

$$\mathbb{E}\left(\sup_{s\in[t,t+1]} |X_s|^k \middle| \mathcal{G}_t\right) \leq c(k)(1 + |X_t|^k). \tag{4.6}$$

where $\mathcal{G}_t = \sigma(B_s, s \leq t; \eta)$. (See *e.g* Gloter (2000)).

Furthermore, Assumptions **(A1)-(A3)** imply **(A4)** if η has distribution ν_0 or η is deterministic (for the latter case, see Gloter (2006), Proposition 3).

Below, we first assume that the noise level ρ_N is known and discuss later the case where ρ_N is unknown. We distinguish the two cases **(B1)-(B2)** which yield different results. Assumption **(B2)** corresponds, for example, to the case $\rho_N\varepsilon_{i\delta_N} = V_{(i+1)\delta_N} - V_{i\delta_N}$, with (V_t) a Brownian motion independent of η and (B_t). Here, $\rho_N = \sqrt{\delta_N}$.

We divide observations into k_N blocks of size p_N and set $\Delta_N = p_N\delta_N$. Since $X_{j\Delta_N}$ is unobserved, we build the local means (4.3). Notice that

$$\mathbb{E}((\varepsilon_\bullet^j)^2) = \frac{1}{p_N}, \qquad \mathbb{E}((\varepsilon_\bullet^j)^4) = \frac{3}{p_N^2} + \frac{m_4 - 3}{p_N^3}, \qquad \mathbb{E}((\varepsilon_\bullet^j)^8) \leq c\left(\frac{1}{p_N^4} + \frac{m_8}{p_N^7}\right).$$

for c a universal constant (the last inequality is obtained using the Rosenthal inequality (see Hall and Heyde (1980) p.23)). Define the σ-fields

$$\begin{gathered} \mathcal{G}_j^N = \mathcal{G}_{j\Delta_N} = \sigma(B_s, s \leq j\Delta_N; \eta), \quad \mathcal{H}_j^N = \mathcal{G}_j^N \vee \mathcal{A}_j^N, \\ \mathcal{A}_j^N = \sigma(\varepsilon_{k\Delta_N + i\delta_N}, i \leq p_N - 1, k \leq j-1) = \sigma(\varepsilon_{l\delta_N}, l \leq j\Delta_N - \delta_N) \end{gathered} \tag{4.7}$$

For $0 \leq j \leq k_N - 1$, the random variable $Y_\bullet^j$ is $\mathcal{H}_{j+1}^N$ measurable. We introduce, for further use, a condition on functions $g : \mathbb{R} \times \Theta \longrightarrow \mathbb{R}$:

(C1) The function g is the restriction of a function defined on $\mathbb{R} \times \mathcal{O}$ with $\mathcal{O}$ an open neighbourhood of Θ and

$$\exists c > 0, \forall x \in \mathbb{R} \quad \sup_{\theta \in \Theta} |g(x, \theta)| \leq c(1 + |x|).$$

We need the following statistical assumptions (**(A6)** is the usual identifiability condition for this problem) :

(A6)

$$\begin{aligned} \sigma(x, \lambda) = \sigma(x, \lambda_0) \qquad & \nu_0 \text{ almost everywhere implies } \lambda = \lambda_0, \\ b(x, \kappa) = b(x, \kappa_0) \qquad & \nu_0 \text{ almost everywhere implies } \kappa = \kappa_0. \end{aligned}$$

(A7) Functions b, σ, σ^{-1}, $\partial_x b$, $\partial_\kappa b$, $\partial_x \sigma$, $\partial_\lambda \sigma$, $\partial^2_{xx} b$, $\partial^2_{\kappa\kappa} b$, $\partial^2_{x\kappa} b$, $\partial_{xx} \sigma$, $\partial^2_{\lambda\lambda} \sigma$ and $\partial^2_{x\lambda} \sigma$ exist, are continuous and satisfy Condition **(C1)**, where ∂ denotes the partial derivative.

4.3 Small sample properties of the local means sample

The following random variables appear in the expansions below :

$$\zeta_{j+1,N} = \frac{1}{p_N} \sum_{i=0}^{p_N - 1} \int_{j\Delta_N + i\delta_N}^{(j+1)\Delta_N} dB_s, \qquad \zeta'_{j+2,N} = \frac{1}{p_N} \sum_{i=0}^{p_N - 1} \int_{(j+1)\Delta_N}^{(j+1)\Delta_N + i\delta_N} dB_s, \quad (4.8)$$

$$\xi'_{j+1,N} = \frac{1}{\Delta_N^{3/2}} \int_{(j+1)\Delta_N}^{(j+2)\Delta_N} ((j+2)\Delta_N - s) dB_s, \quad (4.9)$$

$$\xi'_{i+1,j,N} = \frac{1}{\delta_N^{3/2}} \int_{j\Delta_N + (i+1)\delta_N}^{j\Delta_N + (i+2)\delta_N} (j\Delta_N + (i+2)\delta_N - s) dB_s. \quad (4.10)$$

Some basic properties of these random variables are summarized in Lemma 4.2 and B.3 in the Appendix.

Proposition 4.1 *Let $\bar{X}_j = \Delta_N^{-1} \int_{j\Delta_N}^{(j+1)\Delta_N} X_s ds$. Under Assumption* **(A1)**, *we have*

$$\bar{X}_j - X^j_\bullet = \sqrt{\delta_N} \left(\frac{1}{p_N} \sum_{i=0}^{p_N - 1} \sigma(X_{j\Delta_N + i\delta_N}) \xi'_{i,j,N} \right) + e_{j,N}$$

with (see (4.7))

$$|\mathbb{E}(e_{j,N} | \mathcal{H}^N_j)| \leq \delta_N C(1 + |X_{j\Delta_N}|), \qquad \mathbb{E}(e^2_{j,N} | \mathcal{H}^N_j) \leq \delta^2_N C(1 + |X_{j\Delta_N}|^4).$$

The following proposition precises the local asymptotic behaviour of the observation blocks, by a first order comparison between $Y_\bullet^j$ and $X_{j\Delta_N}$. It can be compared to Proposition 2.2 in Gloter (2000).

Proposition 4.2 *Under* **(A1)**, *we have for* $j \leq k_N - 1$,

$$Y_\bullet^j - X_{j\Delta_N} = \sigma(X_{j\Delta_N})\sqrt{\Delta_N}\xi'_{j,N} + e'_{j,N} + \rho_N \varepsilon_\bullet^j \tag{4.11}$$

with $|\mathbb{E}(e'_{j,N}|\mathcal{H}_j^N)| \leq c\Delta_N(1 + |X_{j\Delta_N}|)$ *and*

$$\mathbb{E}({e'_{j,N}}^2|\mathcal{H}_j^N) \leq c\Delta_N^2(1 + |X_{j\Delta_N}|^4), \qquad \mathbb{E}({e'_{j,N}}^4|\mathcal{H}_j^N) \leq c\Delta_N^3(1 + |X_{j\Delta_N}|^4).$$

If moreover **(A5)** *holds, for* $k \leq 8$, *there exists* $c > 0$ *such that, for* $j \leq k_N - 1$ *:*

$$\mathbb{E}\left(|Y_\bullet^j - X_{j\Delta_N}|^k \middle| \mathcal{H}_j^N\right) \leq C\left(\Delta_N^{k/2}(1 + |X_{j\Delta_N}|^k) + \rho_N^k \mathbb{E}\left(|\varepsilon_\bullet^j|^k\right)\right). \tag{4.12}$$

We deduce

Corollaire 4.1 *Assume* **(A1)** *and* **(A5)**, *and consider* $f : \mathbb{R}^2 \times \Theta \to \mathbb{R}$ *such that* $f, \partial_x f, \partial_{xx}^2 f$ *satisfy* **(C1)**. *Then, there exists* $c > 0$ *such that, for all* $j \geq 0$ *and for all* $\theta \in \Theta$ *:*

$$\left|\mathbb{E}\left(f(Y_\bullet^j, \theta) - f(X_{j\Delta_N}, \theta)\middle| \mathcal{H}_j^N\right)\right| \leq c(\Delta_N(1 + |X_{j\Delta_N}|^2) + \rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^4)}) \tag{4.13}$$

and for $l = 1, 2$

$$\begin{aligned}\mathbb{E}\left((f(Y_\bullet^j, \theta) - f(X_{j\Delta_N}, \theta))^{2l}\middle| \mathcal{H}_j^N\right) \leq\ & c(1 + |X_{j\Delta_N}|^{2l} + \rho_N^{2l}\mathbb{E}((\varepsilon_\bullet^j)^{2l})) \\ & \times(\Delta_N^l(1 + |X_{j\Delta_N}|^{2l}) + \rho_N^{2l}\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^{4l})}).\end{aligned} \tag{4.14}$$

The following proposition is essential for the limit theorems of Section 4.4 and for the statistical application.

Proposition 4.3 *Under Assumptions* **(A1)** *and* **(A5)**, *we have*

$$Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^j) = \sigma(X_{j\Delta_N})(\zeta_{j+1,N} + \zeta'_{j+2,N}) + \tau_{j,N} + \rho_N(\varepsilon_\bullet^{j+1} - \varepsilon_\bullet^j)$$

where $\tau_{j,N}$ *is* $\mathcal{H}_{j+2}^N$ *mesurable, and there exists a constant* c *such that*

$$|\mathbb{E}(\tau_{j,N}|\mathcal{H}_j^N)| \leq c\Delta_N(\Delta_N(1 + |X_{j\Delta_N}|^3) + \rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^4)}),$$

$$\begin{aligned}\mathbb{E}(\tau_{j,N}^2|\mathcal{H}_j^N) + |\mathbb{E}(\tau_{j,N}\zeta_{j+1,N}|\mathcal{H}_j^N)| + |\mathbb{E}(\tau_{j,N}\zeta'_{j+2,N}|\mathcal{H}_j^N)| \leq \\ c\Delta_N(1 + |X_{j\Delta_N}|^2 + \rho_N^2\mathbb{E}((\varepsilon_\bullet^j)^2))(\Delta_N(1 + |X_{j\Delta_N}|^4) + \rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^4)}),\end{aligned}$$

$$\mathbb{E}(\tau_{j,N}^4|\mathcal{H}_j^N) \leq c(1 + |X_{j\Delta_N}|^4 + \rho_N^4\mathbb{E}((\varepsilon_\bullet^j)^4))(\Delta_N^4(1 + |X_{j\Delta_N}|^4) + \rho_N^4\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^8)}).$$

Note that, for $i = 1, 2$, $\rho_N^{2i}\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^{4i})} = O(\frac{\rho_N^{2i}}{p_N^i})$.

Remark : In Gloter (2000), Theorem 2.3., it is proved that (see Proposition 4.1)

$$\bar{X}_{j+1} - \bar{X}_j - \Delta_N b(\bar{X}_j) = \sqrt{\Delta_N}\sigma(X_{j\Delta_N})(\xi_{j,N} + \xi'_{j+1,N}) + \bar{\tau}_{j,N}$$

where $\bar{\tau}_{j,N}$ satisfies $|\mathbb{E}(\bar{\tau}_{j,N}|\mathcal{G}_j^N)| \leq c\Delta_N^2(1 + |X_{j\Delta_N}|^3)$. In Proposition 4.3, additionnal terms due to the noise appear.

4.4 Uniform convergence in probability results

In this section, $f : \mathbb{R} \times \Theta \to \mathbb{R}$ denotes a $\mathcal{C}^2$ function, such that f, $\partial_x f$, $\partial^2_{xx} f$ and $\partial_\theta f$ satisfy (**C1**). The estimation results of Section 4.5 rely on the following statements.

Proposition 4.4 *Under Assumptions* **(A1)-(A5)** *and* **(B1)** *or* **(B2)**, *we have*

$$\bar{\nu}_N(f(.,\theta)) = \frac{1}{k_N}\sum_{j=0}^{k_N-1} f(Y_\bullet^j, \theta) \longrightarrow \nu_0(f(.,\theta)) \tag{4.15}$$

uniformly in θ, *in probability, as* $N \to \infty$, *with* $\delta_N \to 0$, $p_N \to \infty$, $k_N \to \infty$, $\Delta_N = p_N\delta_N \to 0$ *and* $N\delta_N = k_N\Delta_N \to \infty$.

The next theorem precises the variation of the process $(Y_\bullet^j)$.

Théorème 4.1 *Under Assumptions* **(A1)-(A5)**, **(B1)** *or* **(B2)**, *with* $\delta_N = p_N^{-\alpha}, \alpha \in (1, 2]$, *(*$\Delta_N = p_N^{1-\alpha}$*) we have*

$$\bar{I}_N(f(.,\theta)) = \frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1}, \theta)(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1})) \xrightarrow{\mathbb{P}} 0 \tag{4.16}$$

uniformly in θ, *as* $N \to \infty$, *with* $\delta_N \to 0$, $p_N \to \infty$, $k_N \to \infty$, $\Delta_N \to 0$ *and* $N\delta_N \to \infty$.

The following result deals with the quadratic variation of $Y_\bullet^j$.

Théorème 4.2 *Assume* **(A1)-(A5)**.

1. *If* $\delta_N = p_N^{-\alpha}$ *with* $\alpha \in (1, 2)$ *(*$\Delta_N = p_N^{1-\alpha}$*) and* **(B1)** *(*$\rho_N = \rho > 0$*), then*

$$\bar{Q}_N(f(.,\theta)) = \frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1}, \theta)(Y_\bullet^{j+1} - Y_\bullet^j)^2 \xrightarrow{\mathbb{P}} \frac{2}{3}\nu_0(f(.,\theta)\sigma^2), \tag{4.17}$$

2. If $\delta_N = p_N^{-2}$ *(*$\Delta_N = \frac{1}{p_N}$*) and* **(B1)** *(*$\rho_N = \rho > 0$*), then*

$$\bar{Q}_N(f(.,\theta)) \xrightarrow{\mathbb{P}} \frac{2}{3}\nu_0(f(.,\theta)\sigma^2) + 2\rho^2\nu_0(f(.,\theta)), \tag{4.18}$$

3. If $\delta_N = p_N^{-\alpha}$*,* $\alpha \in (1,2]$ *with* **(B2)** *(*$\rho_N \to 0$*), then*

$$\bar{Q}_N(f(.,\theta)) \xrightarrow{\mathbb{P}} \frac{2}{3}\nu_0(f(.,\theta)\sigma^2), \tag{4.19}$$

where all the convergences in probability are uniform in $\theta \in \Theta$*, as* $N \to \infty$*, with* $\delta_N \to 0$*,* $p_N \to \infty$*,* $k_N \to \infty$*,* $\Delta_N \to 0$ *and* $N\delta_N \to \infty$*.*

The proofs of these two last theorems are based on the results of Proposition 4.3 and Lemma B.2 in the Appendix. Theorems 4.1 and 4.2 can be compared to the following classical results :

$$\frac{1}{k_N\Delta_N}\sum_{j=0}^{k_N-1} f(X_{j\Delta_N},\theta)(X_{(j+1)\Delta_N} - X_{j\Delta_N} - \Delta_N b(X_{j\Delta_N})) = o_P(1), \tag{4.20}$$

$$\frac{1}{k_N\Delta_N}\sum_{j=0}^{k_N-1} f(X_{j\Delta_N},\theta)(X_{(j+1)\Delta_N} - X_{j\Delta_N})^2 = \nu_0(f(.,\theta)\sigma^2) + o_P(1). \tag{4.21}$$

Theorem 4.1 gives the analogous result as (4.20), under the condition $\delta_N = p_N^{-\alpha}$, $\alpha \in (1,2]$ and provided that we introduce a lag to avoid correlation terms of order Δ_N (if no lag, the limit is not 0, see for instance Gloter (2006)). Theorem 4.2 underestimates $\nu_0(f(.,\theta)\sigma^2)$ because the variance of $\zeta_{j+1,N} + \zeta'_{j+2,N}$ (see Proposition 4.3) is equivalent to $\frac{2}{3}\Delta_N$ and not to Δ_N. Moreover, for $\rho_N = \rho$ and $\delta_N = p_N^{-2}$, an additional bias appears due to the noise.

4.5 Statistical estimation by contrast minimization

4.5.1 Definition of the contrasts

Let $c(.,\lambda) = \sigma^2(.,\lambda)$ and define

$$\mathcal{E}_N(\theta) = \sum_{j=1}^{k_N-2}\left\{\frac{3}{2\Delta_N}\frac{(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1},\kappa))^2}{c(Y_\bullet^{j-1},\lambda)} + \log(c(Y_\bullet^{j-1},\lambda))\right\}. \tag{4.22}$$

When $\rho_N = \rho$ is fixed (**(B1)**) and $\delta_N = p_N^{-\alpha}$ with $\alpha \in (1,2]$, let $c_{N,\rho}(x,\lambda) = c(x,\lambda) + 3\Delta_N^{\frac{2-\alpha}{\alpha-1}}\rho^2$ and define

$$\mathcal{E}_N^\rho(\theta) = \sum_{j=1}^{k_N-2} \left\{ \frac{3}{2\Delta_N} \frac{(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1},\kappa))^2}{c_{N,\rho}(Y_\bullet^{j-1},\lambda)} + \log(c_{N,\rho}(Y_\bullet^{j-1},\lambda)) \right\} \tag{4.23}$$

We have $\lim_{N\to\infty} c_{N,\rho}(x,\lambda) = c_\rho(x,\lambda)$ with $c_\rho(x,\lambda) = c(x,\lambda)$ if $1 < \alpha < 2$ and $c_\rho(x,\lambda) = c(x,\lambda) + 3\rho^2$ if $\alpha = 2$.

Let $\hat\theta_N$ and $\hat\theta_N^\rho$ be the associated minimum contrast estimators, defined as any solution of

$$\hat\theta_N = \operatorname*{arginf}_{\theta\in\Theta} \mathcal{E}_N(\theta) \text{ and } \hat\theta_N^\rho = \operatorname*{arginf}_{\theta\in\Theta} \mathcal{E}_N^\rho(\theta). \tag{4.24}$$

Théorème 4.3 *Assume* **(A1)-(A7)**, *and consider $\hat\theta_N$ and $\hat\theta_N^\rho$ defined by (4.24).*

1. *If* **(B1)** *or* **(B2)** *holds, with $\delta_N = p_N^{-\alpha}$, $\alpha \in (1,2)$, the estimator $\hat\theta_N$ is consistent : $\hat\theta_N \to \theta_0$ in $\mathbb{P}_{\theta_0}$ probability.*
2. *If* **(B1)** *holds, with $\delta_N = p_N^{-\alpha}$, $\alpha \in (1,2]$, the estimator $\hat\theta_N^\rho$ is consistent.*

Note that point 1 does not require the knowledge of ρ_N.

4.5.2 Estimation with unknown observation noise level under (B1)

Assuming **(B1)** with unknown ρ, we consider the estimator $\hat\rho_N^2 = \frac{1}{2N}\sum_{i=0}^{N-1}(Y_{(i+1)\delta_N} - Y_{i\delta_N})^2$, which is the half of the quadratic variation of the observations.

Lemme 4.1 *Assume* **(A1)-(A5)** *and* **(B1)**. *Then we have $\hat\rho_N^2 \xrightarrow{\mathbb{P}} \rho^2$, when $N\to\infty$, with $\delta_N \to 0$ and $N\delta_N \to \infty$. If, moreover, $N\delta_N^2 \to 0$, $\sqrt{N}(\hat\rho_N^2 - \rho^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 3\rho^4)$.*

The minimum contrast estimator $\hat\theta_N^{\hat\rho_N}$ based on the constrast $\mathcal{E}_N^{\hat\rho_N}(\theta)$ satisfies :

Corollaire 4.2 *Assume* **(A1)-(A7)**, **(B1)** *and $\delta_N = p_N^{-\alpha}$ with $\alpha \in (1,2]$. The estimator $\hat\theta_N^{\hat\rho_N}$ is consistent.*

4.5.3 Link with the case of noisy observations of integrated diffusions

Consider (V_t) a standard Brownian motion independent of (X_t) and suppose that observations are

$$Y_{i\delta_N} = X_{i\delta_N} + \rho(V_{(i+1)\delta_N} - V_{i\delta_N}).$$

Setting $\rho_N = \rho\sqrt{\delta_N}$, $\varepsilon_{i\delta_N} = (V_{j\Delta_N+(i+1)\delta_N} - V_{j\Delta_N+i\delta_N})/\sqrt{\delta_N}$, we are in case **(B2)** and

$$Y_\bullet^j = X_\bullet^j + \frac{\rho}{p_N}(V_{(j+1)\Delta_N} - V_{j\Delta_N}).$$

This kind of observations can be compared with noisy observations of integrated diffusions (see *e.g.* Baltazar-Larios and Sørensen (2009)). Indeed, consider

$$\left\{ \begin{array}{rcl} dX_t &=& b(X_t, \kappa_0)dt + \sigma(X_t, \lambda_0)dB_t \\ dZ_t &=& X_t dt + \epsilon dV_t \end{array} \right.$$

and suppose that we want to estimate θ_0 from discrete observations $(Z_{j\Delta_N})$. We have

$$\Delta_N^{-1}(Z_{(j+1)\Delta_N} - Z_{j\Delta_N}) = \bar{X}_j + \frac{\epsilon}{\Delta_N}\int_{j\Delta_N}^{(j+1)\Delta_N} dV_t.$$

This relation may be compared to $Y_\bullet^j = X_\bullet^j + \rho_N \varepsilon_\bullet^j$ where $\varepsilon_\bullet^j$ is a $\mathcal{N}(0, p_N^{-1})$ random variable if we set $\rho_N = \frac{\epsilon}{\sqrt{\delta_N}}$. As $X_\bullet^j$ and $\bar{X}_j$ have similar properties, we can use the previous contrasts with $(Y_\bullet^j)$ replaced by $\Delta_N^{-1}(Z_{(j+1)\Delta_N} - Z_{j\Delta_N})$ to estimate θ_0 provided that $\epsilon = \epsilon_N$ is such that $\frac{\epsilon_N}{\sqrt{\delta_N}} = O(1)$.

4.6 Examples

In this section, simulation results are given for several examples of diffusion models on simulated data. For the Ornstein-Uhlenbeck model, an implementation on real neuronal data is proposed.

4.6.1 The Ornstein-Uhlenbeck process (simulation)

The hidden diffusion solves

$$dX_t = \kappa X_t dt + \lambda dB_t \tag{4.25}$$

with $\kappa < 0$ and $\lambda > 0$, and X_0 is deterministic or follows the stationary distribution of X. We assume $\rho_N = \rho > 0$ and consider several distributions for the noise.

In this model, we can compute explicitly the estimator $\hat{\theta}_N$ by minimizing the contrast. With the expressions of $\partial_\kappa \mathcal{E}_N(\theta)$ and $\partial_\lambda \mathcal{E}_N(\theta)$, we find

$$\begin{aligned}\hat{\lambda}^2_N &= \frac{3}{2k_N\Delta_N}\sum_{j=1}^{k_N-2}(Y_\bullet^{j+1}-Y_\bullet^j-\Delta_N\hat{\kappa}_N Y_\bullet^{j-1})^2-3\rho^2\mathbf{1}_{\{\alpha=2\}};\\ \hat{\kappa}_N &= \frac{1}{\Delta_N}\frac{\sum_{j=1}^{k_N-2}Y_\bullet^{j-1}(Y_\bullet^{j+1}-Y_\bullet^j)}{\sum_{j=1}^{k_N-2}(Y_\bullet^{j-1})^2}.\end{aligned}$$

We can replace $\hat{\lambda^2}_N$ by

$$\tilde{\lambda}^2_N=\frac{3}{2k_N\Delta_N}\sum_{j=1}^{k_N-2}(Y_\bullet^{j+1}-Y_\bullet^j)^2,\ \text{as}\ \hat{\lambda}^2_n-\tilde{\lambda}^2_N=o_P(1).$$

In Tables 4.1-4.5, the common distribution of $\varepsilon_{i\delta}$ is $\mathcal{N}(0,1)$ and Table 4.6 presents some results with different distributions. Tables 4.1, 4.2 and 4.3 give mean and variance of $\hat{\theta}_N$ for different values of δ, α and N. The values of the parameters are $\kappa_0=-1, \lambda_0=1, \rho^2=0.5$. We have used 500 replications, and we give the empirical mean and standard deviation in parenthesis.

$N=5000, \delta=0.01$ $(N\delta=50, N\delta^2=0.5)$ $\kappa_0=-1, \lambda_0=1, \rho^2=0.5$

	$\alpha=1.17(p=50,k=100)$	$\alpha=1.5(p=22,k=227)$	$\alpha=2(p=10,k=500)$
$\hat{\kappa}_N$ (10^2 Var)	-0.58 (1.53)	-0.76 (2.75)	-0.82 (3.26)
$\hat{\lambda}^2_N$ (10^2 Var)	0.76 (1.19)	1.07 (1.25)	0.86 (2.61)

TABLE 4.1 – Influence of the size of blocks on the estimators, Ornstein-Uhlenbeck model. ($N=5000$ observations, $\delta=0.01$, 500 replications) $\kappa_0=-1, \lambda_0=1, \rho^2=0.5$

$N=20000, \delta=0.005$ $(N\delta=100, N\delta^2=0.5)$ $\kappa_0=-1, \lambda_0=1, \rho^2=0.5$

	$\alpha=1.35(p=50,k=400)$	$\alpha=1.5(p=34,k=588)$	$\alpha=2(p=14,k=1428)$
$\hat{\kappa}_N$ (10^2 Var)	-0.74 (1.08)	-0.81 (1.47)	-0.87 (1.51)
$\hat{\lambda}^2_N$ (10^3 Var)	0.95 (3.87)	1.05 (3.88)	0.92 (11.07)

TABLE 4.2 – Influence of the size of blocks on the estimators, Ornstein-Uhlenbeck model. ($N=20000$ observations, $\delta=0.005$, 500 replications) $\kappa_0=-1, \lambda_0=1, \rho^2=0.5$

First, we remark that the empirical variance is bigger in the case $\alpha=2$ than in the other cases. The parameter κ_0 is always underestimated, but the estimation

$N = 100000, \delta = 0.001$ $(N\delta = 100, N\delta^2 = 0.1)$ $\kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$

	$\alpha = 1.3 (p = 200, k = 500)$	$\alpha = 1.5 (p = 100, k = 1000)$	$\alpha = 2 (p = 32, k = 3125)$
$\hat{\kappa}_N$ (10^2 Var)	-0.81 (1.36)	-0.89 (1.49)	-0.96 (1.95)
$\hat{\lambda}^2_N$ (10^3 Var)	0.90 (2.74)	1.02 (1.99)	0.92 (3.85)

TABLE 4.3 – Influence of the size of blocks on the estimators, Ornstein-Uhlenbeck model. ($N = 100000$ observations, $\delta = 0.001$, 500 replications) $\kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$

of κ_0 is clearly improved as N grows, and δ is close to 0. The estimation of λ_0 is better in Table 4.2 than in Table 4.1, and rather close in Tables 4.2 and 4.3. The variance decreases strongly in the case $\alpha = 2$.

In Table 4.4, we study the influence of the noise on the estimators, in the case $\alpha = \frac{3}{2}$. We use 500 replications, with $\delta = 0.001$ and $N = 10^5$, and we give the empirical mean and standard deviation in parenthesis.

$N = 10^5, \delta = 10^{-3}, \alpha = 1.5, \kappa_0 = -1, \lambda_0 = 1$

	$\rho^2 = 0.1$	$\rho^2 = 1$	$\rho^2 = 2$	$\rho^2 = 5$
$\hat{\kappa}_N$ (10^2 Var)	-0.91 (1.49)	-0.89 (1.50)	-0.86 (1.75)	-0.83 (1.52)
$\hat{\lambda}^2_N$ (10^3 Var)	0.96 (1.71)	1.17 (2.92)	1.47 (4.33)	2.37 (13.42)

TABLE 4.4 – Influence of the observation noise variance on the estimators, Ornstein-Uhlenbeck model. (500 replications, $N = 10^5$, $\delta = 0.001$, $\alpha = \frac{3}{2}$)

A strong bias appears for $\hat{\lambda}_N$ when ρ^2 is bigger than 1, whereas there are no significant changes in the estimation of the drift parameter κ_0. The empirical variances for the estimation of λ_0 also increases : the presence of noise in the observations contaminates the estimation of the diffusion coefficient in this case.

In Table 4.5, we study in Table 4.5 the influence of the value of the diffusion coefficient on the estimators, in the case $\alpha = \frac{3}{2}$. We use 500 replications, with $\delta = 0.001$ and $N = 10^5$, and we give the empirical mean and standard deviation in parenthesis.

The smallest value of λ_0^2 is overestimated by $\hat{\lambda}^2_N$, and this result confirms the ones of Table 4.4 about high levels of noise. For the other values of λ_0^2, no bias is observed.

We finally study in Table 4.6 the influence of the distribution of the errors on the estimators. We choose in this case $\alpha = \frac{3}{2}$, $\kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$. We use 500 replications, with $\delta = 0.001$ and $N = 10^5$, and we give the empirical mean

$N = 10^5, \delta = 10^{-3}, \alpha = 1.5, \kappa_0 = -1, \rho^2 = 1$

	$\lambda_0^2 = 0.1$	$\lambda_0^2 = 0.5$	$\lambda_0^2 = 1$	$\lambda_0^2 = 2$
$\hat{\kappa}_N$ (10^2 Var)	-0.81 (1.48)	-0.87 (1.54)	-0.90 (1.64)	-0.89 (1.62)
$\hat{\lambda}_N^2$ (10^3 Var)	0.23 (0.12)	0.58 (0.78)	1.01 (1.95)	2.01 (6.93)

TABLE 4.5 – Influence of the diffusion coefficient on the estimators, Ornstein-Uhlenbeck model. (500 replications, $N = 10^5$, $\delta = 0.001$, $\alpha = \frac{3}{2}$)

and standard deviation in parenthesis. We make the appropriate corrections on the distributions of $\varepsilon_{i\delta}$ to have unitary variance.

$N = 10^5, \delta = 10^{-3}, \alpha = 1.5, \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$

	$\mathcal{N}(0,1)$	Laplace$(0, \frac{1}{\sqrt{2}})$	Uniform$(-\sqrt{3}, \sqrt{3})$	Logistic$(0, \frac{\sqrt{3}}{\pi})$
$\hat{\kappa}_N$ (10^2 Var)	-0.89 (1.65)	-0.90 (1.52)	-0.87 (1.53)	-0.89 (1.65)
$\hat{\lambda}_N^2$ (10^3 Var)	1.02 (2.11)	1.02 (2.18)	1.31 (3.45)	1.02 (2.10)

TABLE 4.6 – Influence of the distribution of the noise on the estimation, Ornstein-Uhlenbeck model. ($N = 10^5$ observations, $\delta = 0.0001$, $\alpha = \frac{3}{2}$)

We observe that, except in the case of a Uniform distribution, the estimators give results close to the Gaussian case. For the case $\varepsilon_{i\delta} \sim \text{Uniform}(-\sqrt{3}, \sqrt{3})$, a significant positive bias is observed, and the variance is bigger in this case than in the case of Gaussian distribution.

These simulations stress two facts : first, the value $\alpha = \frac{3}{2}$ for the local mean size parameter appears as a good compromise, with a bias in the estimation of κ lower than the bias observed for values of α close to 1, and an empirican variance on simulations lower than the variance observed for $\alpha = 2$. The second remark is about the number of observations : for $N = 5000$ observations, κ is underestimated, for all the values of α considered. Then, the context of high frequency data requires a large number of observations, with a very small discretization step, to be taken into consideration.

4.6.2 Comparison with a discretely observed Ornstein-Uhlenbeck process

In this section, we are interested in the comparison, on simulated datasets, of our method with the methods based on the direct observation of the diffusion at discrete time (see e.g. Genon-Catalot (1990) and Kessler (1997)). To compare

the quality of the noise reduction and its influence to the estimation of the parameters, we compare the results for discrete observations with noise, based on the minimization of the contrast built on the $(Y_\bullet^j)$ with those obtained for the discrete observations without noise, based on the minimization of the contrast built on the $(X_{i\delta_N})$. In both cases the same datasets of N observations with a δ_N-step of discretization are considered. The hidden diffusion (X_t) is an Ornstein-Uhlenbeck process (4.25). We compare the estimators based on the discrete observations $(X_{i\delta_N})$ with the estimator based on (Y_{t_i}) with $Y_{t_i} = X_{t_i} + \rho\varepsilon_{t_i}$, $t_i = i\delta_N$. The results based on the direct observations are given in Table 4.7, and we refer to Tables 4.1, 4.2 and 4.3 for the results based on noisy observations.

$\alpha = 1.5, \kappa_0 = -1, \lambda_0 = 1$, no noise

	$N = 5.10^3, \delta = 10^{-2}$	$N = 2.10^4, \delta = 5.10^{-3}$	$N = 10^5, \delta = 10^{-3}$
$\hat{\kappa}_N$ (Var)	-1.04 (0.21)	-1.02 (0.13)	-1.01 (0.14)
$\hat{\lambda}_N^2$ (Var)	0.99 (1.98×10^{-2})	0.99 (9.80×10^{-3})	1.00 (4.30×10^{-3})

TABLE 4.7 – Parameter estimation with direct observations of the Ornstein-Uhlenbeck model, for several numbers of observations. (500 replications, $\alpha = 1.5, \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$)

The estimation of κ_0 is better for a direct observation of the diffusion, but in this case, the whole set of N observations is taken into account, whereas the size of the set of local means is $k_N = N\delta_N^{\frac{1}{\alpha}}$.

4.6.3 The Ornstein-Uhlenbeck process (neuronal data)

Diffusion-based model has been introduced in the 90's in the field of neuronal studies. The Ornstein-Uhlenbeck diffusion is classical (see *e.g.* Ditlevsen and Lansky (2005)), and the estimation of the parameters has been studied in Ditlevsen and Ditlevsen (2007), for example, when the diffusion is assumed to be observed directly.

Let us consider the stochastic differential equation

$$dZ_t = (-\frac{Z_t}{\tau} + \kappa)dt + \lambda dB_t, \qquad X_0 = x. \tag{4.26}$$

Assume that the observations are at discrete time $t_0 < \cdots < t_N$ and that they are given by

$$Y_{t_i} = Z_{t_i} + \rho\varepsilon_{t_i}$$

where (ε_{t_i}) is a sequence of independent $\mathcal{N}(0,1)$ random variables and ρ is supposed to be known (or preliminary estimated).

Minimum contrast estimators of τ, κ, λ are given by :

$$\begin{pmatrix} -\frac{1}{k_N}\sum_{j=1}^{k_N-2} Y_\bullet^{j-1} & 1 \\ -\frac{1}{k_N}\sum_{j=1}^{k_N-2}(Y_\bullet^{j-1})^2 & \frac{1}{k_N}\sum_{j=1}^{k_N-2} Y_\bullet^{j-1} \end{pmatrix} \begin{pmatrix} \hat{\tau}_N^{-1} \\ \hat{\kappa}_N \end{pmatrix} = \begin{pmatrix} \frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2}(Y_\bullet^{j+1} - Y_\bullet^j) \\ -\frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2} Y_\bullet^{j-1}(Y_\bullet^{j+1} - Y_\bullet^j) \end{pmatrix}$$

and

$$\hat{\lambda^2}_N = \frac{3}{2k_N\Delta_N}\sum_{j=1}^{k_N-2}(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N(-\frac{Y_\bullet^{j-1}}{\hat{\tau}_N} + \hat{\kappa}_N))^2.$$

The dataset used for this study has be formerly presented in Idoux et al. (2006). An example is given in Figure 4.3.

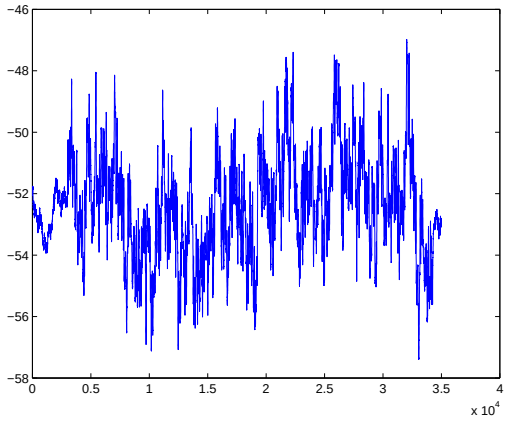

FIGURE 4.1 – The neuronal dataset ($N = 35000$ observations, $\delta = 0.02 \times 10^{-2}$ seconds).

In this dataset, we have $\delta = 0.02$ (10^{-2} seconds), $N = 35000$ observations, $t_N = N\delta = 700$ (10^{-2} seconds). We estimate ρ^2 with :

$$\hat{\rho}_N^2 = \frac{1}{2N}\sum_{i=0}^{N-1}(Y_{t_{i+1}} - Y_{t_i})^2.$$

We have for this dataset $\hat{\rho}_N^2 = 0.0014$.

Then we compute the estimators $(\hat{\tau}_N, \hat{\kappa}_N, \hat{\lambda}_N^2)$ with different choices of p. Results are presented in Table 4.8.

Due to the low level of noise, we also provide the estimators corresponding to a direct observation of the discretized diffusion in Table 4.9.

The results presented in Table 4.8 and 4.9 are rather different, but the mean of the stationary distribution $\mu = \tau\kappa$ is well estimed. Indeed, we find

	$p = 23(\alpha = 1.25)$	$p = 14(\alpha = 1.5)$	$p = 7(\alpha = 2)$
$\hat{\tau}_N$ (10^{-2} seconds)	13.52	5.85	4.67
$\hat{\kappa}_N$ (10^2 mV/ sec)	-3.87	-8.95	-11.23
$\hat{\lambda}^2_N$	1.94	1.78	1.10

TABLE 4.8 – Parameter estimation for neuronal data (with measurement error).

	$\tilde{\tau}_N$ (10^{-2} seconds)	$\tilde{\kappa}_N$ (10^2 mV/ sec)	$\tilde{\lambda}^2_N$
$N = 35000, \delta = 0.0002$	40	-1.31	0.14

TABLE 4.9 – Parameter estimation for neuronal data (without measurement error).

- $\hat{\mu}_N = \hat{\tau}_N \hat{\kappa}_N = -52.28mV$ for the estimator based on the noisy observations model,
- $\tilde{\mu}_N = \tilde{\tau}_N \tilde{\kappa}_N = -52.45mV$ for the estimator based on the direct observations.

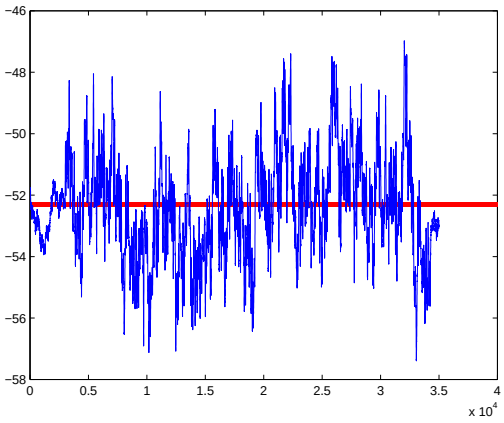

FIGURE 4.2 – The observations with the estimated mean $\hat{\mu}_N$.

However, the estimated values of τ and λ^2 are significantly different, and $\hat{\tau}$ increases with the size p of the blocks.

4.6.4 The Cox-Ingersoll-Ross process

Consider the one-dimensional diffusion process (Cox-Ingersoll-Ross process), solution of

$$dX_t = (\kappa X_t + \kappa')dt + \lambda\sqrt{X_t}dB_t, \qquad X_0 = \eta, \tag{4.27}$$

with $\kappa < 0$, $\kappa' \in \mathbb{R}$ and $\lambda > 0$, and η a positive random variable independent of $(B_t)/$

We assume that the observations at time $t_0 < \cdots < t_N$ are given by

$$Y_{t_i} = X_{t_i} \exp(\varepsilon_{t_i})$$

where (ε_{t_i}) is a sequence of independent $\mathcal{N}(0, \rho^2)$ random variables. Hence the noise is multiplicative, and the observations remain positive.

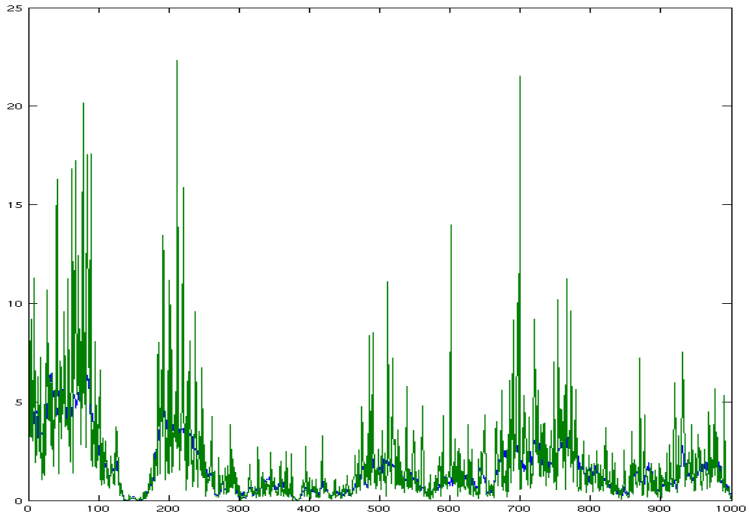

FIGURE 4.3 – An example of Cox-Ingersoll-Ross process observed with a mutliplicative noise.

We consider $U_{t_i} = \log(Y_{t_i})$ to have real valued observations.

The process $Z_t = \log(X_t)$ solves the stochastic differential equation

$$dZ_t = (\kappa + (\kappa' - \frac{\lambda^2}{2})\exp(-Z_t))dt + \lambda \exp(-\frac{Z_t}{2})dB_t.$$

We set $\kappa'' = \kappa' - \frac{\lambda^2}{2}$.

In this case, the scale density is $s(x) = \exp\left(-\frac{2\kappa}{\lambda^2}e^x - \frac{2\kappa''}{\lambda^2}x\right)$ and the speed density is $m(x) = \frac{1}{\lambda^2}\exp\left(\left(\frac{2\kappa''}{\lambda^2}+1\right)x + \frac{2\kappa}{\lambda^2}e^x\right)$. Provided $\kappa < 0$ and $\frac{2\kappa''}{\lambda^2}+1 > 0$, Assumptions **(A2)**, **(A3)** are ensured, and **(A4)** holds with $\eta \sim \nu_0$.

Explicit expressions for the estimator $\hat{\theta}_N = (\hat{\kappa}_N, \hat{\kappa''}_N, \hat{\lambda}^2_N)$ are derived : $(\hat{\kappa}_N, \hat{\kappa''}_N)$ is solution of the system

$$\begin{pmatrix} \Delta_N \sum_{j=1}^{k_N-2} \exp(Y_\bullet^{j-1}) & \Delta_N k_N \\ \Delta_N k_N & \Delta_N \sum_{j=1}^{k_N-2} \exp(-Y_\bullet^{j-1}) \end{pmatrix} \begin{pmatrix} \hat{\kappa}_N \\ \hat{\kappa''}_N \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{k_N-2} \exp(Y_\bullet^{j-1})(Y_\bullet^{j+1} - Y_\bullet^j) \\ \sum_{j=1}^{k_N-2}(Y_\bullet^{j+1} - Y_\bullet^j) \end{pmatrix}$$

and

$$\hat{\lambda}^2_N = \frac{3}{2k_N\Delta_N}\sum_{j=1}^{k_N-2} \exp(Y_\bullet^{j-1})(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N(\hat{\kappa}_N + \hat{\kappa''}_N \exp(-Y_\bullet^{j-1})))^2.$$

Recall that the following explicit expressions for the estimator $\tilde{\theta}_N = (\tilde{\kappa}_N, \tilde{\kappa'}_N, \tilde{\lambda}^2_N)$ when the diffusion (X_t) is directly observed (Kessler (1997)) :

$$\begin{pmatrix} \Delta_N \sum_{j=1}^{k_N-2} X_{j\Delta_N} & \Delta_N k_N \\ \Delta_N k_N & \Delta_N \sum_{j=1}^{k_N-2} \frac{1}{X_{j\Delta_N}} \end{pmatrix} \begin{pmatrix} \tilde{\kappa}_N \\ \tilde{\kappa'}_N \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{k_N-2}(X_{(j+1)\Delta_N} - X_{j\Delta_N}) \\ \sum_{j=1}^{k_N-2} \frac{1}{X_{j\Delta_N}}(X_{(j+1)\Delta_N} - X_{j\Delta_N}) \end{pmatrix}$$

and

$$\tilde{\lambda}^2_N = \frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2} \frac{1}{X_{j\Delta_N}}(X_{(j+1)\Delta_N} - X_{j\Delta_N} - \Delta_N(\tilde{\kappa}_N X_{j\Delta_N} + \tilde{\kappa'}_N))^2.$$

Simulation results are presented in Table 4.10 (with noise) and Table 4.11 (directly observed). For this study, $N = 500$ trajectories are simulated with parameters $\kappa_0 = -2, \kappa'_0 = 3, \lambda_0 = 4, \rho^2 = 0.5$, and then $\kappa''_0 = 1$. Due to the simulation study for the Ornstein-Uhlenbeck process, the value $\alpha = \frac{3}{2}$ as local mean size parameter.

In Table 4.10, we observe that $\kappa''_0 = 1$ is well estimated, whereas the estimation of κ_0 is negatively biased. The empirical variance, for $\hat{\kappa}_N$ and $\hat{\kappa''}_N$ decreases between $N = 5000$ and $N = 20000$ observations, but there is no significative improvement between $N = 20000$ and $N = 100000$ observations. For the diffusion parameter λ_0, the estimator $\hat{\lambda}_N$ is positively biased, with a variance decreasing as the number of observations grows.

These results can be compared with the case of direct observations, given in Table 4.11.

Notice that there is no bias in the estimation of κ_0 and κ'_0 for $N = 5000$ and $N = 20000$, contrary to the noisy case. Moreover, the estimation of λ^2_0 is more accurate, with a lower empirical variance for $\hat{\lambda}^2_N$.

$\kappa_0 = -2, \kappa_0'' = 1, \lambda_0 = 4, \rho^2 = 0.5, \alpha = 1.5$

	$N = 5.10^3, \delta = 10^{-2}$	$N = 2.10^4, \delta = 5.10^{-3}$	$N = 10^5, \delta = 10^{-3}$
$\hat{\kappa}_N$ (10^2 Var)	-1.43 (6.28)	-1.56 (3.14)	-1.78 (3.37)
$\hat{\kappa}''_N$ (10^2 Var)	0.99 (4.57)	1.03 (2.12)	1.13 (2.44)
$\hat{\lambda}^2_N$ (10^2 Var)	4.23 (37.61)	4.35 (15.15)	4.40 (8.15)

TABLE 4.10
Parameter estimation for the Cox-Ingersoll-Ross process with a multiplicative noise for different values of α. (500 replications, $\kappa_0 = -2, \kappa_0'' = 1, \lambda_0 = 4, \rho^2 = 0.5, \alpha = \frac{3}{2}$)

$\kappa_0 = -2, \kappa_0' = 3, \lambda_0 = 4, \alpha = 1.5$

	$N = 5.10^3, \delta = 10^{-2}$	$N = 2.10^4, \delta = 5.10^{-3}$	$N = 10^5, \delta = 10^{-3}$
$\hat{\kappa}_N$ (10^2 Var)	-2.04 (11.03)	-2.03 (6.65)	-2.46 (53.45)
$\hat{\kappa}'_N$ (10^2 Var)	3.02 (13.47)	3.03 (8.17)	3.45 (65.44)
$\hat{\lambda}^2_N$ (10^2 Var)	4.11 (0.95)	4.05 (0.20)	4.01 (0.36)

TABLE 4.11 – Parameter estimation for the Cox-Ingersoll-Ross process with direct observations for different values of α. (500 replications, $\kappa_0 = -2, \kappa_0' = 3, \lambda_0 = 4, \rho^2 = 0.5, \alpha = \frac{3}{2}$

4.6.5 The hyperbolic diffusion

Consider the one dimensional diffusion process solution of

$$dX_t = \kappa X_t dt + \lambda\sqrt{1 + X_t^2} dB_t, \qquad X_0 = \eta \in \mathbb{R}, \tag{4.28}$$

where η is a random variable independent of (B_t), $\kappa < 0$ and $\lambda > 0$. In this case, the model is positive recurrent if $|\kappa| + \frac{\lambda^2}{2} > 0$, and in this case, its stationary distribution has density

$$\nu(x) \propto \frac{1}{(1 + x^2)^{1+\frac{|\kappa|}{\lambda^2}}}.$$

If $X_0 = \eta$ has distribution $\nu_0(x)dx$, then, $\sqrt{1 + \frac{2\kappa}{\lambda^2}}\eta$ has Student distribution which can be easily simulated. Even if η_0 has only a finite number of finite moments, and **(A4)** does not holds, for $2(1 + \frac{|\kappa|}{\lambda^2}) > k + 1$, ν_0 has a finite moment of order k.

Now we consider

$$G(x) = \int_0^x \frac{dx}{\lambda\sqrt{1 + x^2}} = \frac{1}{\lambda} \arg\sinh(x).$$

By the Ito formula, $Z_t = G(X_t)$ satisfies

$$dZ_t = \beta(Z_t)dt + dB_t$$

with

$$\beta(z) = -\left(\frac{\kappa}{\lambda} + \frac{\lambda}{2}\right)\tanh(\lambda z).$$

Sample paths of this diffusion can be simulated exactly with the retrospective exact simulation algorithms proposed in Beskos et al. (2006).

We can compute explicitly the estimator $\hat{\theta}_N = (\hat{\kappa}_N, \hat{\lambda^2}_N)$ in this case :

$$\hat{\kappa}_N = \frac{1}{\Delta_N} \frac{\sum_{j=1}^{k_N-2} \frac{Y_\bullet^{j-1}}{1+(Y_\bullet^{j-1})^2}(Y_\bullet^{j+1} - Y_\bullet^j)}{\sum_{j=1}^{k_N-2} \frac{(Y_\bullet^{j-1})^2}{1+(Y_\bullet^{j-1})^2}}$$

and

$$\hat{\lambda^2}_N = \frac{3}{2k_N\Delta_N} \sum_{j=1}^{k_N-2} \frac{(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N\hat{\kappa}_N Y_\bullet^{j-1})^2}{1+(Y_\bullet^{j-1})^2}.$$

Some simulation results are given in Tables 4.12 and 4.13, with different distributions for the observation noise. In the different cases, $N = 500$ replications are made, and the empirical mean is given with the associated standard deviation in parenthesis. We consider for the values of the parameters : $\kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$.

$\alpha = \frac{3}{2}, \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5, \varepsilon \sim \mathcal{N}(0,1)$

	$N = 5.10^3, \delta = 10^{-2}$	$N = 2.10^4, \delta = 5.10^{-3}$	$N = 10^5, \delta = 10^{-3}$
$\hat{\kappa}_N$ (Var)	-0.75 (0.21)	-0.82 (0.16)	-0.90 (0.17)
$\hat{\lambda}^2_N$ (Var)	1.14 (0.14)	1.11 (0.08)	1.07 (0.06)

TABLE 4.12 – Parameter estimation for the hyperbolic diffusion, with a Gaussian noise, for different values of N. (500 replications, $\alpha = \frac{3}{2}, \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$)

A negative bias is observed in the estimation of κ_0, whereas a positive one appears in the estimation of λ_0^2. For the two different noise distributions, empirical means and variances are very close in this model.

4.7 Concluding remarks

The contrasts presented in this work give associated estimators for parameters involved in a non-Markovian setting : one-dimensional diffusions observed with

$\alpha = \frac{3}{2}, \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5, \varepsilon \sim$Laplace$(0, \frac{1}{\sqrt{2}})$

	$N = 5.10^3, \delta = 10^{-2}$	$N = 2.10^4, \delta = 5.10^{-3}$	$N = 10^5, \delta = 10^{-3}$
$\hat{\kappa}_N$ (Var)	-0.74 (0.21)	-0.82 (0.16)	-0.89 (0.18)
$\hat{\lambda}^2_N$ (Var)	1.15 (0.15)	1.12 (0.09)	1.08 (0.07)

TABLE 4.13 – Parameter estimation for the hyperbolic diffusion, with a Laplace noise, for different values of N. (500 replications, $\alpha = \frac{3}{2}, \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$)

a noise. The consistency of these minimum contrast estimators is illustrated on several simulations, and the estimated values are close to the values obtained for a direct observation of the diffusion. The importance of the sampling rate of local means, depending on the choice of α, appears in the case $\alpha = 2$ with $\rho_N = \rho$, where the variance of the observation noise ρ^2 appears in the limit theorem for the quadratic variation. The asymptotic normality is studied in Chapter 5.

Acknowledgements

The author is greatly indebtful to Valentine Genon-Catalot, his PhD advisor, for her help and her numerous comments on the redaction of this work, and Adeline Samson for the discussions about the numerical results and her insightful suggestions. He also thank Lee E.Moore for the dataset from his experiment on neuronal activity, and Yves Rozehnolc for his help and his comments on the exact simulation algorithm.

4.8 Proofs

It is useful to introduce the intervals $I_{j,k,N} := I_{j,k} = [j\Delta_N + k\delta_N, j\Delta_N + (k+1)\delta_N)$, for $k = 0, \ldots, p_N - 1$, $j = 0, \ldots, k_N - 1$, which satisfy for all j, if $k \neq k'$ $I_{j,k} \cap I_{j,k'} = \emptyset$ and for $j \neq j'$ and all k, k', $I_{j,k} \cap I_{j',k'} = \emptyset$.

Moreover, Assumptions **(A1)** and **(A2)** imply ergodicity for (X_t). Ergodicity is used in Lemma B.1 in the Appendix to deal with asymptotic properties when $N \to \infty$, and to derive Proposition 4.4, Theorems 4.1 and 4.2.

Lemme 4.2 *The random variables $\zeta_{j+1,N}$ and $\zeta'_{j+1,N}$ are $\mathcal{G}_{(j+1)\Delta_N}$ measurable,*

$\zeta'_{j+2,N}$ is independent of $\mathcal{G}_{(j+1)\Delta_N}$, and the following holds :

$$\zeta_{j+1,N} = \frac{1}{p_N}\sum_{k=0}^{p_N-1}(k+1)\int_{I_{j,k}} dB_s, \quad \zeta'_{j+2,N} = \frac{1}{p_N}\sum_{k=0}^{p_N-1}(p_N-1-k)\int_{I_{j+1,k}} dB_s. \tag{4.29}$$

Moreover, we have

$$\mathbb{E}(\zeta_{j,N}|\mathcal{G}_j^N) = 0, \quad \mathbb{E}(\zeta'_{j+1,N}|\mathcal{G}_j^N) = 0, \quad \mathbb{E}(\zeta_{j+1,N}\zeta'_{j+1,N}|\mathcal{G}_j^N) = \frac{\Delta_N}{6}\left(1-\frac{1}{p_N^2}\right),$$
$$\mathbb{E}((\zeta_{j+1,N})^2|\mathcal{G}_j^N) = \Delta_N\left(\frac{1}{3}+\frac{1}{2p_N}+\frac{1}{6p_N^2}\right), \quad \mathbb{E}((\zeta'_{j+1,N})^2|\mathcal{G}_j^N) = \Delta_N\left(\frac{1}{3}-\frac{1}{2p_N}+\frac{1}{6p_N^2}\right).$$

Proof of Lemma 4.2 Using (4.8), we can rearrange terms to exhibit non-overlapping intervals, hence conditionally independent variables, and obtain (4.29). Afterwards, the proof is achieved by elementary computations.

□

Proof of Proposition 4.1 First, note that, as $(X_t, t \geq 0)$ and $(\varepsilon_{k\delta_N})$ are independent, for $l = 1, 2$,

$$\mathbb{E}(e_{j,N}^l|\mathcal{H}_{j,N}) = \mathbb{E}(e_{j,N}^l|\mathcal{G}_{j,N}).$$

Thus we study the expectations given $\mathcal{G}_{j,N}$. Using $\Delta_N = p_N\delta_N$ yields

$$R_{j,N} = \bar{X}_j - X_\bullet^j = \frac{1}{p_N}\sum_{k=0}^{p_N-1}\frac{1}{\delta_N}\int_{I_{j,k}}(X_s - X_{j\Delta_N+k\delta_N})ds.$$

By the Fubini theorem, we get

$$R_{j,N} = \sqrt{\delta_N}\left(\frac{1}{p_N}\sum_{k=0}^{p_N-1}\sigma(X_{j\Delta_N+k\delta_N})\xi'_{k,j,N}\right) + e_{j,N}$$

where $e_{j,N} = \alpha_{j,N} + \beta_{j,N}$, with

$$\alpha_{j,N} = \frac{1}{p_N}\sum_{k=0}^{p_N-1}\frac{1}{\delta_N}\int_{I_{j,k}}(j\Delta_N + (k+1)\delta_N - s)(\sigma(X_s) - \sigma(X_{j\Delta_N+k\delta_N}))dB_s$$

and

$$\beta_{j,N} = \frac{1}{p_N}\sum_{k=0}^{p_N-1}\frac{1}{\delta_N}\int_{I_{j,k}}\int_{j\Delta_N+i\delta_N}^{s} b(X_u)duds.$$

Under Assumption **(A1)**, we have $|\beta_{j,N}| \leq c\delta_N(1 + \sup_{s\in[j\Delta_N,(j+1)\Delta_N]}|X_s|)$. And for all $p \geq 0$, by (4.6),

$$\mathbb{E}(|\beta_{j,N}|^p|\mathcal{G}_j^N) \leq c\delta_N^p(1 + |X_{j\Delta_N}|^p).$$

Also $\mathbb{E}(\alpha_{j,N}|\mathcal{G}_j^N)=0$, so we get $|\mathbb{E}(e_{j,N}|\mathcal{G}_j^N)|\leq \delta_N c(1+|X_{j\Delta_N}|)$. Furthermore, we get with the Jensen inequality, the Ito isometry and the Fubini theorem

$$\mathbb{E}\left((\alpha_{j,N})^2\middle|\,\mathcal{G}_j^N\right) \;\leq\; c\frac{1}{p_N}\sum_{k=0}^{p_N-1}\int_{I_{j,k}}\mathbb{E}((\sigma(X_s)-\sigma(X_{j\Delta_N+k\delta_N}))^2|\mathcal{G}_j^N)ds.$$

With Proposition B.1 in the Appendix, it comes $\mathbb{E}\left(|\alpha_{j,N}|^2|\,\mathcal{G}_j^N\right)\leq C\delta_N^2(1+|X_{j\Delta_N}|^4)$. This implies the result. □

Proof of Proposition 4.2 We have

$$Y_\bullet^j - X_{j\Delta_N} \;=\; X_\bullet^j - \bar{X}_j + \bar{X}_j - X_{j\Delta_N} + \rho_N\varepsilon_\bullet^j,$$

where $\varepsilon^{\bullet j}$ is independent of $\mathcal{H}_j^N$. Proposition 2.2 in Gloter (2000) states that, using the random variables (4.9),

$$\bar{X}_j - X_{j\Delta_N} = \sigma(X_{j\Delta_N})\sqrt{\Delta_N}\xi'_{j,N} + \bar{e}_{j,N}$$

with $|\mathbb{E}(\bar{e}_{j,N}|\mathcal{H}_j^N)| = |\mathbb{E}(\bar{e}_{j,N}|\mathcal{G}_j^N)|\leq c\Delta_N(1+|X_{j\Delta_N}|)$ and $\mathbb{E}(\bar{e}_{j,N}^2|\mathcal{H}_j^N)=\mathbb{E}(\bar{e}_{j,N}^2|\mathcal{G}_j^N)\leq c\Delta_N^2(1+|X_{j\Delta_N}^4|)$. With Proposition 4.1, setting $e'_{j,N}=e_{j,N}+\bar{e}_{j,N}$, we get the first part of Proposition 4.2. Now we need to prove that, for some $c>0$

$$\mathbb{E}\left(|r_{j,N}|^k\middle|\,\mathcal{H}_j^N\right)=\mathbb{E}\left(|r_{j,N}|^k\middle|\,\mathcal{G}_j^N\right)\leq c(1+|X_{j\Delta_N}|^k) \tag{4.30}$$

where

$$r_{j,N}=\frac{1}{p_N}\sum_{i=0}^{p_N-1}\sigma(X_{j\Delta_N+i\delta_N})\xi'_{i,j,N}$$

and $\xi'_{i,j,N}$ is defined in (4.10). With elementary computations on conditional expectation, we get (see notation (4.7))

$$\mathbb{E}\left(|r_{j,N}|^k\middle|\,\mathcal{G}_j^N\right)\leq\frac{1}{p_N}\sum_{i=0}^{p_N-1}\mathbb{E}(|\sigma(X_{j\Delta_N+i\delta_N})|^k\mathbb{E}(|\xi'_{i,j,N}|^k|\mathcal{G}_{j\Delta_N+i\delta_N})|\mathcal{G}_j^N).$$

As $\xi'_{i,j,N}$ is independent of $\mathcal{G}_{j\Delta_N+i\delta_N}$,

$$\mathbb{E}\left(|r_{j,N}|^k\middle|\,\mathcal{G}_j^N\right)\leq c\frac{1}{p_N}\sum_{i=0}^{p_N-1}\mathbb{E}(1+|X_{j\Delta_N+i\delta_N}|^k|\mathcal{G}_j^N)$$

which implies (4.30). Finally, $\mathbb{E}(|\varepsilon_\bullet^j|^k|\mathcal{H}_j^N)=\mathbb{E}(|\varepsilon_\bullet^j|^k)$ because $\varepsilon_\bullet^j$ is independent of $\mathcal{H}_j^N$. □

Proof of Corollary 4.1 We have, with Taylor's formula (order two) :

$$D_j := f(Y_\bullet^j,\theta)-f(X_{j\Delta_N},\theta)=\partial_x f(X_{j\Delta_N},\theta)(Y_\bullet^j-X_{j\Delta_N})+\frac{1}{2}\partial_{xx}^2 f(Z,\theta)(Y_\bullet^j-X_{j\Delta_N})^2$$

with $Z \in (Y_\bullet^j, X_{j\Delta_N})$. Then, with the Cauchy Schwarz inequality, using that the derivatives satisfy **(C1)**, and Proposition 4.2, there exists a constant $c > 0$ such that, for all $\theta \in \Theta$,

$$\begin{aligned}
|\mathbb{E}(D_j|\mathcal{H}_j^N)| &\leq c(1+|X_{j\Delta_N}|)|\mathbb{E}(e'_{j,N}|\mathcal{H}_j^N)| \\
&\quad +c(1+|X_{j\Delta_N}|+\rho_N\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^2)})\sqrt{\mathbb{E}((Y_\bullet^j - X_{j\Delta_N})^4|\mathcal{H}_j^N)} \\
&\leq c\Delta_N(1+|X_{j\Delta_N}|^2) \\
&\quad +c(1+|X_{j\Delta_N}|+\rho_N\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^2)}) \\
&\quad \times(\Delta_N(1+|X_{j\Delta_N}|^2)+\rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^4)}).
\end{aligned}$$

With Taylor's formula (order one), there exists a random variable $\tilde{Z} \in (Y_\bullet^j, X_{j\Delta_N})$ and a constant $c > 0$ independent of θ such that $D_j^2 = (\partial_x f(\tilde{Z},\theta))^2(Y_\bullet^j - X_{j\Delta_N})^2$ and

$$D_j^2 \leq c(1+\sup_{s\in[j\Delta_N,(j+1)\Delta_N]}[X_s|^2+\rho_N^2|\varepsilon_\bullet^j|^2)(Y_\bullet^j - X_{j\Delta_N})^2.$$

Using the Cauchy-Schwarz inequality and condition **(C1)**,

$$\mathbb{E}(D_j^2|\mathcal{H}_j^N) \leq c(1+|X_{j\Delta_N}|^2+\rho_N^2\mathbb{E}((\varepsilon_\bullet^j)^2))(\Delta_N(1+|X_{j\Delta_N}|^2)+\rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^4)}).$$

Analogously, $D_j^4 = (\partial_x f(\tilde{Z},\theta))^4(Y_\bullet^j - X_{j\Delta_N})^4$ and

$$D_j^4 \leq c(1+\sup_{s\in[j\Delta_N,(j+1)\Delta_N]}[X_s|^4+\rho_N^4|\varepsilon_\bullet^j|^4)(Y_\bullet^j - X_{j\Delta_N})^4,$$

with c independent of θ. Using the Cauchy-Schwarz inequality, it comes

$$\mathbb{E}(D_j^4|\mathcal{H}_j^N) \leq c(1+|X_{j\Delta_N}|^4+\rho_N^4\mathbb{E}((\varepsilon_\bullet^j)^4))(\Delta_N^2(1+|X_{j\Delta_N}|^4)+\rho_N^4\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^8)}).$$

□

Proof of Proposition 4.3 In this proof, we study all conditional expectation given $\mathcal{G}_j^N$ as they are identical to conditional expectations given $\mathcal{H}_j^N$ in all the terms involved below. We have

$$Y_\bullet^{j+1} - Y_\bullet^j = X_\bullet^{j+1} - X_\bullet^j + \rho_N(\varepsilon_\bullet^{j+1} - \varepsilon_\bullet^j).$$

Setting $C_j = X_\bullet^{j+1} - X_\bullet^j$ and rearranging terms yields

$$\begin{aligned}
C_j &= \frac{1}{p_N}\sum_{k=0}^{p_N-1}(X_{(j+1)\Delta_N+k\delta_N} - X_{j\Delta_N+k\delta_N}) \\
&= \frac{1}{p_N}\sum_{k=0}^{p_N-1}\sum_{l=0}^{p_N-1}\int_{I_{j,k+l}} dX_s \\
&= \frac{1}{p_N}\sum_{k=0}^{p_N-1}(k+1)\int_{I_{j,k}} dX_s + \frac{1}{p_N}\sum_{k=0}^{p_N-1}(p_N-k-1)\int_{I_{j+1,k}} dX_s.
\end{aligned}$$

We use

$$\begin{aligned}\int_{I_{j,k}} dX_s &= b(X_{j\Delta_N+k\delta_N})\delta_N + \int_{I_{j,k}} (b(X_s) - b(X_{j\Delta_N+k\delta_N}))ds \\ &\quad +\sigma(X_{j\Delta_N+k\delta_N}) \int_{I_{j,k}} dB_s + \int_{I_{j,k}} (\sigma(X_s) - \sigma(X_{j\Delta_N+k\delta_N}))dB_s.\end{aligned}$$

By splitting Δ_N into $\Delta_N = (k+1)\delta_N + (p_N - k - 1)\delta_N$ for all k, we get (see notation 4.8)

$$\begin{aligned}Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^j) &= C_j - \Delta_N b(Y_\bullet^j) + \rho_N(\varepsilon_\bullet^{j+1} - \varepsilon_\bullet^j) \\ &= \sigma(X_{j\Delta_N})(\zeta_{j+1,N} + \zeta'_{j+2,N}) + \tau_{j,N} + \rho_N(\varepsilon_\bullet^{j+1} - \varepsilon_\bullet^j)\end{aligned}$$

where $\tau_{j,N} = \sum_{\ell=1}^4 \tau_{j,N}^{(\ell)}$ and for $\ell = 1, \ldots, 4$, $\tau_{j,N}^{(\ell)} = r_{j,N}^{(\ell)} + s_{j,N}^{(\ell)}$ with

$$r_{j,N}^{(1)} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (k+1)\delta_N(b(X_{j\Delta_N+k\delta_N}) - b(Y_\bullet^j)), \tag{4.31}$$

$$s_{j,N}^{(1)} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (p_N - k - 1)\delta_N(b(X_{(j+1)\Delta_N+k\delta_N}) - b(Y_\bullet^j)), \tag{4.32}$$

$$r_{j,N}^{(2)} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (k+1)\sigma(X_{j\Delta_N+k\delta_N}) \int_{I_{j,k}} dB_s - \sigma(X_{j\Delta_N})\zeta_{j+1,N}, \tag{4.33}$$

$$s_{j,N}^{(2)} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (p_N - k - 1)\sigma(X_{(j+1)\Delta_N+k\delta_N}) \int_{I_{j+1,k}} dB_s - \sigma(X_{j\Delta_N})\zeta'_{j+2,N}, \tag{4.34}$$

$$r_{j,N}^{(3)} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (k+1) \int_{I_{j,k}} (b(X_s) - b(X_{j\Delta_N+k\delta_N}))ds, \tag{4.35}$$

$$s_{j,N}^{(3)} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (p_N - k - 1) \int_{I_{j+1,k}} (b(X_s) - b(X_{(j+1)\Delta_N+k\delta_N}))ds, \tag{4.36}$$

$$r_{j,N}^{(4)} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (k+1) \int_{I_{j,k}} (\sigma(X_s) - \sigma(X_{j\Delta_N+k\delta_N}))dB_s, \tag{4.37}$$

$$s_{j,N}^{(4)} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (p_N - k - 1) \int_{I_{j+1,k}} (\sigma(X_s) - \sigma(X_{(j+1)\Delta_N+k\delta_N}))dB_s. \tag{4.38}$$

We mainly treat the terms $r_{j,N}^{(\ell)}$ because the others are analogous. We have $\mathbb{E}(r_{j,N}^{(\ell)}|\mathcal{G}_j^N) = 0$ and $\mathbb{E}(s_{j,N}^{(\ell)}|\mathcal{G}_j^N) = 0$ for $\ell = 2, 4$. Next,

$$|\mathbb{E}(r_{j,N}^{(1)}|\mathcal{G}_j^N)| \leq \frac{1}{p_N} \sum_{k=0}^{p_N-1} (k+1)\delta_N |\mathbb{E}(b(X_{j\Delta_N+k\delta_N}) - b(Y_\bullet^j)|\mathcal{G}_j^N)|$$

We use, for $k = 0 \dots p_N - 1$ and $s \in I_{j,k}$, the inequality

$$|\mathbb{E}(b(X_s) - b(X_{j\Delta_N + k\delta_N})|\mathcal{G}_j^N)| \leq c\Delta_N(1 + |X_{j\Delta_N}|^3).$$

With (4.13), it comes $|\mathbb{E}(r_{j,N}^{(1)}|\mathcal{G}_j^N)| \leq c\Delta_N(\Delta_N(1 + |X_{j\Delta_N}|^2) + \rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^4)})$. Then, with the Fubini theorem, we derive $|\mathbb{E}(\tau_{j,N}^{(3)}|\mathcal{G}_j^N)| \leq c\Delta_N^2(1 + |X_{j\Delta_N}|^3)$. Hence

$$|\mathbb{E}(\tau_{j,N}|\mathcal{G}_j^N)| \leq c\Delta_N(\Delta_N(1 + |X_{j\Delta_N}|^3) + \rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^4)}).$$

Now we deal with $\mathbb{E}((r_{j,N}^{(1)})^2|\mathcal{G}_j^N)$. With Proposition 4.1 and the Cauchy-Schwarz inequality, it comes

$$\mathbb{E}((b(Y_\bullet^j) - b(X_{j\Delta_N}))^2|\mathcal{G}_j^N) \leq c(1+|X_{j\Delta_N}|^2+\rho_N^2\mathbb{E}((\varepsilon_\bullet^j)^2))(\Delta_N(1+|X_{j\Delta_N}|^2)+\rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^4)}).$$

Applying the Cauchy-Schwarz inequality, and after elementary computations, we obtain

$$\mathbb{E}((r_{j,N}^{(1)})^2|\mathcal{G}_j^N) \leq c\Delta_N^2(1+|X_{j\Delta_N}|^2+\rho_N^2\mathbb{E}((\varepsilon_\bullet^j)^2))(\Delta_N(1+|X_{j\Delta_N}|^2)+\rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^4)}).$$

With analogous techniques, we have

$$\begin{aligned}\mathbb{E}((\tau_{j,N}^{(3)})^2|\mathcal{G}_j^N) &\leq c\Delta_N^2 \sup_{s\in[j\Delta_N,(j+2)\Delta_N]} \mathbb{E}((b(X_s) - b(X_{j\Delta_N}))^2|\mathcal{G}_j^N) \\ &\leq c\Delta_N^3(1 + |X_{j\Delta_N}|^4).\end{aligned}$$

Using Lemma 4.2, we obtain

$$\begin{aligned}r_{j,N}^{(2)} &= \frac{1}{p_N}\sum_{k=0}^{p_N-1}(k+1)(\sigma(X_{j\Delta_N+k\delta_N}) - \sigma(X_{j\Delta_N}))\int_{I_{j,k}} dB_s, \\ s_{j,N}^{(2)} &= \frac{1}{p_N}\sum_{k=0}^{p_N-1}(p_N - k - 1)(\sigma(X_{(j+1)\Delta_N+k\delta_N}) - \sigma(X_{j\Delta_N}))\int_{I_{j+1,k}} dB_s.\end{aligned}$$

Thus $r_{j,N}^{(2)} = \int_{j\Delta_N}^{(j+1)\Delta_N} f(s)dB_s$ with

$$f(s) = \frac{1}{p_N}\sum_{k=0}^{p_N-1}(k+1)(\sigma(X_{j\Delta_N+k\delta_N}) - \sigma(X_{j\Delta_N}))\mathbf{1}_{I_{j,k}}(s) \qquad (4.39)$$

With the Ito isometry and the Fubini theorem, we have

$$\begin{aligned}\mathbb{E}((r_{j,N}^{(2)})^2|\mathcal{G}_j^N) &= \frac{1}{p_N^2}\sum_{k=0}^{p_N-1}(k+1)^2\delta_N\mathbb{E}((\sigma(X_{j\Delta_N+k\delta_N}) - \sigma(X_{j\Delta_N}))^2|\mathcal{G}_j^N) \\ &\leq c\Delta_N^2(1 + |X_{j\Delta_N}|^4)\end{aligned}$$

We use similar techniques with $r^{(4)}_{j,N}$ and $s^{(4)}_{j,N}$ to obtain

$$\mathbb{E}((\tau^{(2)}_{j,N})^2 + (\tau^{(4)}_{j,N})^2|\mathcal{G}^N_j) \quad \leq \quad c\Delta^2_N(1+|X_{j\Delta_N}|^4).$$

Collecting terms, we get the bound for $\mathbb{E}(\tau^2_{j,N}|\mathcal{G}^N_j)$.

Now, using (4.31), (4.8), Lemma 4.2 and the Cauchy Schwarz inequality we have

$$|\mathbb{E}(r^{(1)}_{j,N}\zeta_{j+1,N}|\mathcal{G}^N_j)| \leq c\Delta_N^{\frac{3}{2}}\sqrt{\mathbb{E}((b(X_{j\Delta_N+k\delta_N}) - b(Y^j_\bullet))^2|\mathcal{G}^N_j)}.$$

Corollary 4.1 implies

$$|\mathbb{E}(r^{(1)}_{j,N}\zeta_{j+1,N}|\mathcal{G}^N_j)| \leq c\Delta_N^{\frac{3}{2}}(1+|X_{j\Delta_N}|+\rho_N\sqrt{\mathbb{E}((\varepsilon^j_\bullet)^2)})(\sqrt{\Delta_N}(1+|X_{j\Delta_N}|)+\rho_N(\mathbb{E}((\varepsilon^j_\bullet)^4))^{\frac{1}{4}}).$$

The same inequality holds for $\mathbb{E}(r^{(1)}_{j,N}\zeta'_{j+2,N}|\mathcal{G}^N_j)$, $\mathbb{E}(s^{(1)}_{j,N}\zeta_{j+1,N}|\mathcal{G}^N_j)$ and $\mathbb{E}(s^{(1)}_{j,N}\zeta'_{j+2,N}|\mathcal{G}^N_j)$.

We can write $\zeta_{j+1,N} = \int_{j\Delta_N}^{(j+1)\Delta_N} g(s)dB_s$ with $g(s) = \frac{1}{p_N}\sum_{l=0}^{p_N-1}(l+1)\mathbf{1}_{I_{j,l}}(s)$.
Using (4.39) and Corollary 4.1, we obtain

$$\begin{aligned}|\mathbb{E}(r^{(2)}_{j,N}\zeta_{j+1,N}|\mathcal{G}^N_j)| \quad &\leq \quad \frac{1}{p_N^2}\sum_{k=0}^{p_N-1}(k+1)^2\delta_N|\mathbb{E}(\sigma(X_{j\Delta_N+k\delta_N}) - \sigma(X_{j\Delta_N})|\mathcal{G}^N_j)| \\ &\leq \quad c\Delta_N(1+|X_{j\Delta_N}| + \rho_N^2\mathbb{E}((\varepsilon^j_\bullet)^2))(\Delta_N(1+|X_{j\Delta_N}|^2) + \rho_N^2\sqrt{\mathbb{E}((\varepsilon^j_\bullet)^4)}).\end{aligned}$$

The same inequality holds for $|\mathbb{E}(r^{(2)}_{j,N}\zeta'_{j+2,N}|\mathcal{G}^N_j)|$.

For $r^{(3)}_{j,N}$ (see (4.35)), we use the Cauchy Schwarz inequality :

$$|\mathbb{E}(r^{(3)}_{j,N}\zeta_{j+1,N}|\mathcal{G}^N_j)| \leq \frac{1}{p_N^2}\sum_{k,l=0}^{p_N-1}(k+1)(l+1)\delta_N^{3/2}\mathbb{E}\left(\sup_{s\in I_{j,k}}(b(X_s) - b(X_{j\Delta_N+k\delta_N}))^2\middle|\mathcal{G}^N_j\right)^{\frac{1}{2}}.$$

Hence

$$|\mathbb{E}(r^{(3)}_{j,N}\zeta_{j+1,N}|\mathcal{G}^N_j)| \leq c\Delta^2_N(1+|X_{j\Delta_N}|^2).$$

Furthermore $\mathbb{E}(r^{(3)}_{j,N}\zeta'_{j+2,N}|\mathcal{G}^N_j) = 0$.

With the Fubini theorem and the Ito isometry, we have

$$\mathbb{E}(r^{(4)}_{j,N}\zeta_{j+1,N}|\mathcal{G}^N_j) \quad = \quad \frac{1}{p_N^2}\sum_{k=0}^{p_N-1}(k+1)^2\int_{I_{j,k}}\mathbb{E}(\sigma(X_s) - \sigma(X_{j\Delta_N+k\delta_N})|\mathcal{G}^N_j)ds$$

Introducing $Lf = \frac{\sigma^2}{2}f'' + bf'$ yields

$$\sigma(X_s) - \sigma(X_{j\Delta_N+k\delta_N}) = \int_{j\Delta_N+k\delta_N}^s L\sigma(X_u)du + \frac{1}{2}\int_{j\Delta_N+k\delta_N}^s \sigma(X_u)\sigma'(X_u)dB_u.$$

Therefore, $|\mathbb{E}(\sigma(X_s) - \sigma(X_{j\Delta_N+k\delta_N})|\mathcal{G}_j^N)| \leq c\Delta_N(1+|X_{j\Delta_N}|^4)$ which implies

$$|\mathbb{E}(r_{j,N}^{(4)}\zeta_{j+1,N}|\mathcal{G}_j^N)| \leq c\Delta_N^2(1+|X_{j\Delta_N}|^4).$$

Furthermore $\mathbb{E}(r_{j,N}^{(4)}\zeta'_{j+2,N}|\mathcal{G}_j^N) = 0$. The terms containing $s_{j,N}^{(3)}$ and $s_{j,N}^{(4)}$ are treated analogously. This gives the bound for $|\mathbb{E}(\tau_{j,N}\zeta_{j+1,N}|\mathcal{G}_j^N)|$ and $|\mathbb{E}(\tau_{j,N}\zeta'_{j+2,N}|\mathcal{G}_j^N)|$.

Finally, we have to bound the fourth order conditional moment of $\tau_{j,N}$. We only study the terms $r_{j,N}^{(2)}$ and $r_{j,N}^{(1)}$. Using (4.39), the Burkholder - Davies - Gundy inequality and Proposition B.1, we have

$$\begin{aligned}\mathbb{E}((r_{j,N}^{(2)})^4|\mathcal{G}_j^N) &\leq c\mathbb{E}\left(\left(\int_{j\Delta_N}^{(j+1)\Delta_N} f(s)^2ds\right)^2\middle|\mathcal{G}_j^N\right)\\ &\leq c\Delta_N^2\mathbb{E}(\sup_{s\in[j\Delta_N,(j+1)\Delta_N]}(\sigma(X_s)-\sigma(X_{j\Delta_N}))^4|\mathcal{G}_j^N) \leq c\Delta_N^4(1+|X_{j\Delta_N}|^4).\end{aligned}$$

With similar computations, we derive $\mathbb{E}((\tau_{j,N}^{(2)})^4+(\tau_{j,N}^{(4)})^4|\mathcal{G}_j^N) \leq c\Delta_N^4(1+|X_{j\Delta_N}|^4)$.

Using Proposition 4.1, we get

$$\begin{aligned}\mathbb{E}((r_{j,N}^{(1)})^4|\mathcal{G}_j^N) &\leq c\frac{\delta_N^4}{p_N}\sum_{k=0}^{p_N-1}(k+1)^4\mathbb{E}((b(Y_\bullet^j)-b(X_{j\Delta_N+k\delta_N}))^4|\mathcal{G}_j^N)\\ &\leq c(1+|X_{j\Delta_N}|^4+\rho_N^4\mathbb{E}((\varepsilon_\bullet^j)^4))(\Delta_N^6(1+|X_{j\Delta_N}|^4)+\rho_N^4\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^8)}).\end{aligned}$$

Analogously, using Proposition B.1, $\mathbb{E}((r_{j,N}^{(3)})^4|\mathcal{G}_j^N) \leq c\Delta_N^6(1+|X_{j\Delta_N}|^4)$. Finally, we get the bound for $\mathbb{E}(\tau_{j,N}^4|\mathcal{G}_j^N)$.

□

Proof of Proposition 4.4 By Lemma B.1, it is enough to prove the L^1 convergence to zero of

$$\sup_{\theta\in\Theta}\frac{1}{k_N}\sum_{j=0}^{k_n-1}|f(Y_\bullet^j,\theta)-f(X_{j\Delta_N},\theta)|.$$

By Taylor expansion and condition (**C1**) we derive the bound

$$A_j := \sup_{\theta\in\Theta}|f(Y_\bullet^j,\theta)-f(X_{j\Delta_N},\theta)| \leq c(1+|X_{j\Delta_N}|+|Y_\bullet^j|)|Y_\bullet^j - X_{j\Delta_N}|.$$

Hence, the Cauchy Schwarz inequality and Assumption **(A2)** imply

$$\mathbb{E}\left(A_j|\mathcal{H}_j^N\right) \leq c(1+|X_{j\Delta_N}|+\rho_N\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^2)})\sqrt{\mathbb{E}\left(|Y_\bullet^j - X_{j\Delta_N}|^2\middle|\,\mathcal{H}_j^N\right)}.$$

Then, with (4.12), Assumptions **(A5)** and **(B1)**, and $\mathbb{E}((\varepsilon_\bullet^j)^2) = \frac{1}{p_N}$, the result holds. □

Proof of Theorem 4.1 We have

$$\bar{I}_N(f(.,\theta)) = \tilde{I}_N(f(.,\theta)) + \frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1},\theta)\Delta_N(b(Y_\bullet^j) - b(Y_\bullet^{j-1})),$$

where $\tilde{I}_N(f(.,\theta)) = \frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2} V_j^N(\theta)$ with $V_j^N(\theta) = f(Y_\bullet^{j-1},\theta)(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^j))$. We only need to prove that $\tilde{I}_N(f(.,\theta)) \to 0$ in probability, uniformly in $\theta \in \Theta$, as the second term is $o_P(1)$, uniformly in θ. As $V_j^N(\theta)$ is $\mathcal{H}_{j+2}^N$-measurable, we split the sum into three parts

$$\sum_{j=1}^{k_N-2} V_{j,N}(\theta) = \sum_{1\le 3j\le k_N-2} V_{3j,N}(\theta) + \sum_{1\le 3j+1\le k_N-2} V_{3j+1,N}(\theta) + \sum_{1\le 3j+2\le k_N-2} V_{3j+2,N}(\theta).$$

We treat only the sum with indexes multiples of 3 and set :

$$V_{3j}^N(\theta) = v_{3j,N}^{(1)}(\theta) + v_{3j,N}^{(2)}(\theta) + v_{3j,N}^{(3)}(\theta)$$

where

$$\begin{aligned} v_{3j,N}^{(1)}(\theta) &= f(Y_\bullet^{3j-1},\theta)\sigma(X_{3j\Delta_N})(\zeta_{3j+1,N} + \zeta'_{3j+2,N}),\\ v_{3j,N}^{(2)}(\theta) &= f(Y_\bullet^{3j-1},\theta)\rho_N(\varepsilon_\bullet^{3j+1} - \varepsilon_\bullet^{3j}),\\ v_{3j,N}^{(3)}(\theta) &= f(Y_\bullet^{3j-1},\theta)\tau_{3j,N}. \end{aligned}$$

In order to prove the pointwise convergence in θ to zero, we use Lemma B.2. As $Y_\bullet^{3j-1}, X_{3j\Delta_N}$ are $\mathcal{H}_{3j}^N$-measurables and $\varepsilon_\bullet^{3j+1} - \varepsilon_\bullet^{3j}$ is independent of $\mathcal{H}_{3j}^N$, we have $\mathbb{E}(v_{3j,N}^{(1)}(\theta)|\mathcal{H}_{3j}^N) = 0$ and $\mathbb{E}(v_{3j,N}^{(2)}(\theta)|\mathcal{H}_{3j}^N) = 0$. By Proposition 4.3,

$$|\mathbb{E}(\tau_{3j,N}|\mathcal{H}_{3j}^N)| \le c\Delta_N(1+|X_{3j\Delta_N}|^2 + \rho_N^2\mathbb{E}((\varepsilon_\bullet^{3j})^2))(\Delta_N(1+|X_{3j\Delta_N}|^4) + \rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet^{3j})^4)}).$$

Using **(A4)**, this implies $\frac{1}{k_N\Delta_N}\sum_{1\le 3j\le k_N-2}\mathbb{E}(v_{3j,N}^{(3)}(\theta)|\mathcal{H}_{3j}^N) = o_P(1)$. We also have to verify for $\ell = 1,2,3$,

$$\frac{1}{(k_N\Delta_N)^2}\sum_{j=1}^{k_N-2}\mathbb{E}((v_{3j,N}^{(\ell)}(\theta))^2|\mathcal{H}_{3j}^N) = o_P(1).$$

For $\ell = 1$, we have

$$\begin{aligned} &\frac{1}{(k_N\Delta_N)^2}\sum_{1\le 3j\le k_N-2}\mathbb{E}((v_{3j,N}^{(1)})^2(\theta)|\mathcal{H}_{3j}^N)\\ &\quad= \frac{1}{(k_N\Delta_N)^2}\sum_{1\le 3j\le k_N-2} f(Y_\bullet^{3j-1},\theta)^2\sigma(X_{3j\Delta_N})^2\mathbb{E}\left(\left(\zeta_{3j+1,N} + \zeta'_{3j+2,N}\right)^2\middle|\mathcal{H}_{3j}^N\right)\\ &\quad\le \frac{1}{N\delta_N}\frac{2}{k_N}\sum_{1\le 3j\le k_N-2} f(Y_\bullet^{3j-1},\theta)^2\sigma(X_{3j\Delta_N})^2 = o_P(1). \end{aligned}$$

For $\ell = 2$,

$$\begin{aligned}\frac{1}{(k_N\Delta_N)^2}\sum_{1\leq 3j\leq k_N-2}\mathbb{E}((v^{(2)}_{3j,N})^2(\theta)|\mathcal{H}^N_{3j}) &= \frac{1}{(k_N\Delta_N)^2}\sum_{1\leq 3j\leq k_N-2} f(Y_\bullet^{3j-1},\theta)^2\rho_N^2\mathbb{E}((\varepsilon_\bullet^{3j+1}-\varepsilon_\bullet^{3j})^2)\\ &= \frac{2\rho_N^2}{N\delta_N p_N\Delta_N}\frac{1}{k_N}\sum_{1\leq 3j\leq k_N-2} f(Y_\bullet^{3j-1},\theta)^2.\end{aligned}$$

As $p_N\Delta_N = p_N^{2-\alpha}$, with $1<\alpha\leq 2$, the above term is $o_P(1)$.

For $\ell = 3$,

$$\frac{1}{(k_N\Delta_N)^2}\sum_{j=1}^{k_N-2}\mathbb{E}((v^{(3)}_{3j,N})^2(\theta)|\mathcal{H}^N_{3j}) = \frac{1}{k_N}\frac{1}{k_N}\sum_{j=1}^{k_N-2} f_\theta(Y_\bullet^{3j-1})^2\frac{1}{\Delta_N^2}\mathbb{E}(\tau^2_{j,N}|\mathcal{H}^N_{3j}) = o_P(1),$$

using that, by Proposition 4.3, $\Delta_N^{-2}\mathbb{E}(\tau^2_{j,N}|\mathcal{H}^N_j)$ is $O_P(1)$.

To obtain uniformity in θ, we shall use Proposition B.2 and evaluate $\sup_{N\in\mathbb{N}}\mathbb{E}(\sup_{\theta\in\Theta}|\partial_\theta\tilde{I}_N(f_\theta)|)$. To study

$$\partial_\theta\tilde{I}_N(f_\theta) = \frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2}\partial_\theta V_j^N(\theta),$$

we use the same method, split the sum in three parts, and define :

$$S_N^{(\ell)}(\theta) = \frac{1}{k_N\Delta_N}\sum_{1\leq 3j\leq k_N-2} v^{(\ell)}_{3j,N}(\theta).$$

The sum for $\ell = 3$ is the simplest. With assumption (**C1**) for $\partial_\theta f$, we deduce

$$\mathbb{E}(\sup_{\theta\in\Theta}|\partial_\theta v^{(3)}_{3j,N}(\theta)||\mathcal{H}^N_{3j}) \leq c(1+|Y_\bullet^{3j-1}|)\sqrt{\mathbb{E}(\tau^2_{3j,N}|\mathcal{H}^N_{3j})}.$$

With the Cauchy Schwarz inequality, we have

$$\begin{aligned}\mathbb{E}(\sup_{\theta\in\Theta}|\partial_\theta v^{(3)}_{3j,N}(\theta)||\mathcal{H}^N_{3j}) \leq\ & c\sqrt{\Delta_N}(1+|Y_\bullet^{3j-1}|)(1+|X_{3j\Delta_N}|+\rho_N\sqrt{\mathbb{E}((\varepsilon_\bullet^{3j})^2)})\\ &\times(\sqrt{\Delta_N}(1+|X_{3j\Delta_N}|^2)+\rho_N\left(\mathbb{E}\left(\left(\varepsilon_\bullet^{3j}\right)^4\right)\right)^{\frac{1}{4}})\end{aligned}$$

and with Lemma B.1 and **(A4)-(A5)**, this implies $\sup_{N\in\mathbb{N}}\mathbb{E}(\sup_{\theta\in\Theta}|\partial_\theta S_N^{(3)}(\theta)|) < \infty$.

We cannot use the same method to study $S_N^{(\ell)}(\theta), \ell = 1,2$. Instead, we use Theorem 20 in Appendix 1 of Ibragimov and Has′minskiĭ (1981) : it is enough to show that, for $\ell = 1,2$, there exists two constants $M\geq 0$ and $\epsilon > 0$ such that :

$$\begin{aligned}&\forall\theta\in\Theta,\forall N\in\mathbb{N},\quad \mathbb{E}(|S_N^{(\ell)}|^{2+\epsilon})\leq M\\ \text{and}\quad &\forall\theta,\theta'\in\Theta,\forall N\in\mathbb{N},\quad D_N(\theta,\theta')\leq M|\theta-\theta'|^{2+\epsilon}\end{aligned}\tag{4.40}$$

where $D_N(\theta,\theta') = \mathbb{E}(|S_N^{(\ell)}(\theta) - S_N^{(\ell)}(\theta')|^{2+\epsilon})$.

For $\ell = 1$, using the Rosenthal inequality for martingales, we get, for any $\epsilon > 0$:

$$\begin{aligned}\mathbb{E}(|S_N^{(1)}(\theta)|^{2+\epsilon}) &\leq \frac{1}{(k_N\Delta_N)^{2+\epsilon}}\mathbb{E}\left(\left|\sum_{1\leq 3j\leq k_N-2}\mathbb{E}\left((v_{3j,N}^{(1)}(\theta))^2\Big|\mathcal{H}_{3j}^N\right)\right|^{1+\frac{\epsilon}{2}}\right)\\ &\quad+\frac{1}{(k_N\Delta_N)^{2+\epsilon}}\sum_{1\leq 3j\leq k_N-2}\mathbb{E}(|v_{3j,N}^{(1)}(\theta)|^{2+\epsilon})\end{aligned}$$

Then it comes :

$$\mathbb{E}\left(\left|\sum_{1\leq 3j\leq k_N-2}\mathbb{E}\left((v_{3j,N}^{(1)}(\theta))^2\Big|\mathcal{H}_{3j}^N\right)\right|^{1+\frac{\epsilon}{2}}\right) \leq k_N^{\frac{\epsilon}{2}}\sum_{1\leq 3j\leq k_N-2}\mathbb{E}\left(\left|\mathbb{E}\left((v_{3j,N}^{(1)}(\theta))^2\Big|\mathcal{H}_{3j}^N\right)\right|^{1+\frac{\epsilon}{2}}\right)$$

With $\mathbb{E}((\zeta_{3j+1,N}+\zeta'_{3j+2,N})^2|\mathcal{H}_{3j}^N) = \Delta_N\left(1-\frac{1}{3}\left(\frac{p_N^2-1}{p_N^2}\right)\right)$, Assumption **(A5)** and **(C1)**, we derive

$$\sup_{j,N}\mathbb{E}\left(\left|\mathbb{E}\left((v_{3j,N}^{(1)}(\theta))^2\Big|\mathcal{H}_{3j}^N\right)\right|^{1+\frac{\epsilon}{2}}\right) \leq c\Delta_N^{1+\frac{\epsilon}{2}} \text{ and } \sup_{j,N}\mathbb{E}\left(\left|v_{3j,N}^{(1)}(\theta)\right|^{2+\epsilon}\right) \leq c\Delta_N^{1+\frac{\epsilon}{2}}.$$

Hence

$$\mathbb{E}\left(\left|S_N^{(1)}(\theta)\right|^{2+\epsilon}\right) \leq c\left(\frac{1}{(k_N\Delta_N)^{1+\frac{\epsilon}{2}}}+\frac{1}{(k_N\Delta_N)^{1+\frac{\epsilon}{2}}}\frac{1}{k_N^{\frac{\epsilon}{2}}}\right).$$

The study of $D_N(\theta,\theta')$ is analogous, so (4.40) holds. This implies $S_N^{(1)}(\theta) = o_P(1)$ uniformly in θ.

We use similar tools for $S_N^{(2)}$. With the Rosenthal inequality, we have

$$\begin{aligned}\mathbb{E}(|S_N^{(2)}(\theta)|^{2+\epsilon}) &\leq \frac{1}{(k_N\Delta_N)^{2+\epsilon}}\mathbb{E}\left(\left|\sum_{1\leq 3j\leq k_N-2}\mathbb{E}\left((v_{3j,N}^{(2)}(\theta))^2\Big|\mathcal{H}_{3j}^N\right)\right|^{1+\frac{\epsilon}{2}}\right)\\ &\quad+\frac{1}{(k_N\Delta_N)^{2+\epsilon}}\sum_{1\leq 3j\leq k_N-2}\mathbb{E}(|v_{3j,N}^{(2)}(\theta)|^{2+\epsilon}).\end{aligned}$$

Hence, with $\mathbb{E}\left((v_{3j,N}^{(2)}(\theta))^2\Big|\mathcal{H}_{3j}^N\right) = 2\rho_N^2 f(Y_\bullet^{3j-1},\theta)^2\sigma(X_{3j\Delta_N})^2\mathbb{E}((\varepsilon_\bullet^{3j})^2)$ and $\mathbb{E}((\varepsilon_\bullet^{3j})^2) = \frac{1}{p_N}$, and $\Delta_N = p_N^{1-\alpha}$, we obtain (4.40). Finally $\tilde{I}_N(f_\theta) = o_P(1)$, uniformly in θ. □

Proof of Theorem 4.2 Let $W_{j,N}(\theta) = f(Y_\bullet^{j-1},\theta)(Y_\bullet^{j+1}-Y_\bullet^j)^2$. By Proposition

4.3, we have $W_{j,N}(\theta) = \sum_{i=1}^{6} w_{j,N}^{(i)}(\theta)$ with

$$\begin{aligned}
w_{j,N}^{(1)}(\theta) &= f(Y_\bullet^{j-1},\theta)\sigma(X_{j\Delta_N})^2(\zeta_{j+1,N}+\zeta'_{j+2,N})^2\\
w_{j,N}^{(2)}(\theta) &= f(Y_\bullet^{j-1},\theta)\rho_N^2(\varepsilon_\bullet^{j+1}-\varepsilon_\bullet^{j})^2\\
w_{j,N}^{(3)}(\theta) &= f(Y_\bullet^{j-1},\theta)(\Delta_N b(Y_\bullet^j)+\tau_{j,N})^2\\
w_{j,N}^{(4)}(\theta) &= f(Y_\bullet^{j-1},\theta)2\sigma(X_{j\Delta_N})(\zeta_{j+1,N}+\zeta'_{j+2,N})\rho_N(\varepsilon_\bullet^{j+1}-\varepsilon_\bullet^{j})\\
w_{j,N}^{(5)}(\theta) &= f(Y_\bullet^{j-1},\theta)2\sigma(X_{j\Delta_N})(\zeta_{j+1,N}+\zeta'_{j+2,N})(\Delta_N b(Y_\bullet^j)+\tau_{j,N})\\
w_{j,N}^{(6)}(\theta) &= f(Y_\bullet^{j-1},\theta)2\rho_N(\varepsilon_\bullet^{j+1}-\varepsilon_\bullet^{j})(\Delta_N b(Y_\bullet^j)+\tau_{j,N}),
\end{aligned}$$

where we recall that $Y_\bullet^{j-1}, X_{j\Delta_N}$ are $\mathcal{H}_j^N$-measurable and $\varepsilon_\bullet^{j+1}-\varepsilon_\bullet^{j}$ is independent of $\mathcal{H}_j^N$. Therefore, splitting again into three parts, we consider, for $\ell = 0, 1, 2$,

$$T_{\ell,N}^{(i)}(\theta) = \frac{1}{k_N\Delta_N}\sum_{1\le 3j+\ell\le k_N-2} w_{3j+\ell,N}^{(i)}(\theta) \qquad \text{for } i = 1,\dots,6.$$

We start by studying $T_{0,N}^{(1)}(\theta)$:

$$\mathbb{E}(w_{3j,N}^{(1)}(\theta)|\mathcal{H}_{3j}^N) = f(Y_\bullet^{3j-1},\theta)\sigma(X_{3j\Delta_N})^2\Delta_N\left(1-\frac{1}{3}\left(\frac{p_N^2-1}{p_N^2}\right)\right)$$

and

$$\mathbb{E}((w_{3j,N}^{(1)}(\theta))^2|\mathcal{H}_{3j}^N) = 3f(Y_\bullet^{3j-1},\theta)^2\sigma(X_{3j\Delta_N})^4\Delta_N^2\left(\frac{2}{3}+\frac{1}{3p_N^2}\right)^2.$$

Applying Lemma B.2 with Lemma B.1, we get, for all θ, $T_{0,N}^{(1)}(\theta) = \frac{1}{3}\times\frac{2}{3}\nu_0(f(.,\theta)\sigma^2)+o_P(1)$. Thus

$$T_{0,N}^{(1)}(\theta)+T_{1,N}^{(1)}(\theta)+T_{2,N}^{(1)}(\theta) = \frac{2}{3}\nu_0(f(.,\theta)\sigma^2)+o_P(1).$$

Then, we study $T_{0,N}^{(2)}(\theta)$:

$$\begin{aligned}
\mathbb{E}(w_{3j,N}^{(2)}(\theta)|\mathcal{H}_{3j}^N) &= f(Y_\bullet^{3j-1},\theta)\rho_N^2\mathbb{E}((\varepsilon_\bullet^{3j+1}-\varepsilon_\bullet^{3j})^2)\\
&= 2f(Y_\bullet^{3j-1},\theta)\rho_N^2 p_N^{-1}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}((w_{3j,N}^{(2)}(\theta))^4|\mathcal{H}_{3j}^N) &= f(Y_\bullet^{3j-1},\theta)^2\rho_N^4\mathbb{E}((\varepsilon_\bullet^{3j+1}-\varepsilon_\bullet^{3j})^4)\\
&= f(Y_\bullet^{3j-1},\theta)^2\rho_N^4(12p_N^{-2}(1+o(1)))
\end{aligned}$$

Recall that $\Delta_N = p_N^{1-\alpha}$, $1<\alpha\le 2$. If $\alpha<2$, with Lemma B.2, $T_{0,N}^{(2)} = o_P(1)$. But if $\alpha = 2$, *i.e.* $\Delta_N = \frac{1}{p_N}$, and $\rho_N = \rho$, we have $T_{0,N}^{(2)}(\theta) = \frac{1}{3}\times 2\rho^2\nu_0(f(.,\theta)^2)+o_P(1)$. and

$$T_{0,N}^{(2)}(\theta)+T_{1,N}^{(2)}(\theta)+T_{2,N}^{(2)}(\theta) = 2\rho^2\nu_0(f(.,\theta)^2)+o_P(1).$$

We easily deduce from Proposition 4.3, Lemma B.2 and Lemma B.1 that $T^{(3)}_{0,N}(\theta) = o_P(1)$. For $T^{(4)}_{0,N}(\theta)$, we have

$$\mathbb{E}(w^{(4)}_{3j,N}(\theta)|\mathcal{H}^N_{3j}) = 2f(Y^{3j-1}_\bullet,\theta)\sigma(X_{3j\Delta_N})\rho_N\mathbb{E}((\zeta_{3j+1,N}+\zeta'_{3j+2,N})(\varepsilon^{3j+1}_\bullet-\varepsilon^{3j}_\bullet)|\mathcal{H}^N_{3j})$$

Given $\mathcal{H}^N_{3j}$, the random variables $(\zeta_{3j+1,N}+\zeta'_{3j+2,N})$ and $(\varepsilon^{3j+1}_\bullet-\varepsilon^{3j}_\bullet)$ are independent, so $\mathbb{E}(w^{(4)}_{3j,N}(\theta)|\mathcal{H}^N_{3j})=0$. Furthermore

$$\begin{aligned}\mathbb{E}((w^{(4)}_{3j,N}(\theta))^2|\mathcal{H}^N_{3j}) &= 4f(Y^{3j-1}_\bullet,\theta)^2\sigma(X_{3j\Delta_N})^2\rho_N^2\mathbb{E}((\zeta_{3j+1,N}+\zeta'_{3j+2,N})^2(\varepsilon^{3j+1}_\bullet-\varepsilon^{3j}_\bullet)^2|\mathcal{H}^N_{3j})\\ &= 8f(Y^{3j-1}_\bullet,\theta)^2\sigma(X_{3j\Delta_N})^2\rho_N^2\Delta_N\left(\frac{2}{3}+\frac{1}{3p_N^2}\right)\frac{1}{p_N}.\end{aligned}$$

Then, with Proposition 4.3, Lemma B.2 and Lemma B.1, $T^{(4)}_{0,N}(\theta)=o_P(1)$.

We have

$$\mathbb{E}(w^{(5)}_{3j,N}(\theta)|\mathcal{H}^N_{3j}) = 2f(Y^{3j-1}_\bullet,\theta)\sigma(X_{3j\Delta_N})\mathbb{E}((\zeta_{3j+1,N}+\zeta'_{3j+2,N})(\Delta_N b(Y^{3j}_\bullet)+\tau_{3j,N})|\mathcal{H}^N_{3j}).$$

With the Cauchy Schwarz inequality,

$$\begin{aligned}|\mathbb{E}(w^{(5)}_{3j,N}(\theta)|\mathcal{H}^N_{3j})| &\le c|f(Y^{3j-1}_\bullet,\theta)|\sigma(X_{3j\Delta_N})\sqrt{\Delta_N}\sqrt{\mathbb{E}((\Delta_N b(Y^{3j}_\bullet)+\tau_{3j,N})^2|\mathcal{H}^N_{3j})}\\ &\le c|f(Y^{3j-1}_\bullet,\theta)|\sigma(X_{3j\Delta_N})\sqrt{\Delta_N}\sqrt{\Delta_N^2\mathbb{E}(b(Y^{3j}_\bullet)^2|\mathcal{H}^N_j)+\mathbb{E}(\tau^2_{3j,N}|\mathcal{H}^N_{3j})}.\end{aligned}$$

Moreover, with the Cauchy Schwarz inequality,

$$\begin{aligned}\mathbb{E}((w^{(5)}_{3j,N}(\theta))^2|\mathcal{H}^N_{3j}) &= 4f(Y^{3j-1}_\bullet,\theta)^2\sigma(X_{3j\Delta_N})^2\mathbb{E}((\zeta_{3j+1,N}+\zeta'_{3j+2,N})^2(\Delta_N b(Y^{3j}_\bullet)+\tau_{3j,N})^2|\mathcal{H}^N_{3j})\\ &\le cf(Y^{3j-1}_\bullet,\theta)^2\sigma(X_{3j\Delta_N})^2\Delta_N^2\sqrt{\mathbb{E}((\Delta_N b(Y^{3j}_\bullet)+\tau_{3j,N})^4|\mathcal{H}^N_{3j})}.\end{aligned}$$

Then, with Proposition 4.3, Lemma B.2 and Lemma B.1 $T^{(5)}_{0,N}=o_P(1)$.

With some straightforward computations, $T^{(6)}_{3j,N}=o_P(1)$.

We prove now uniformity in θ in these convergences, using Proposition B.2. For $w^{(1)}_{j,N}(\theta)$, we get

$$\mathbb{E}\left(\sup_{\theta\in\Theta}\left|\frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2}\partial_\theta w^{(1)}_{j,N}(\theta)\right|\right)<\infty$$

with

$$\mathbb{E}\left(\sigma(X_{j\Delta_N})^2(\zeta_{j+1,N}+\zeta'_{j+2,N})^2\middle|\,\mathcal{H}^N_j\right)\le c\Delta_N\sigma(X_{j\Delta_N})^2.$$

With similar arguments for $w^{(i)}_{j,N}(\theta), i=2\dots6$, we derive uniformity in θ.

□

Proof of Lemma 4.1 We have $\hat{\rho^2}_N - \rho^2 = a_{1,N} + a_{2,N} + a_{3,N}$ where

$$a_{1,N} = \frac{\rho^2}{2N}\sum_{i=0}^{N-1}\{(\varepsilon_{(i+1)\delta_N} - \varepsilon_{i\delta_N})^2 - 2\}, \qquad a_{2,N} = \frac{1}{2N}\sum_{i=0}^{N-1}(X_{(i+1)\delta_N} - X_{i\delta_N})^2,$$
$$a_{3,N} = \frac{\rho}{N}\sum_{i=0}^{N-1}(X_{(i+1)\delta_N} - X_{i\delta_N})(\varepsilon_{(i+1)\delta_N} - \varepsilon_{i\delta_N}).$$

With the usual law of large numbers, $a_{1,N} = o_P(1)$. With Proposition B.1,

$$\mathbb{E}(a_{2,N}) \leq c\delta_N(1 + \sup_{t\geq 0}\mathbb{E}(X_t^2)) = \delta_N O(1), \quad \mathbb{E}((a_{2,N})^2) \leq \frac{c\delta_N}{N}.$$

Hence $\hat{\rho}_N - \rho^2 = o_P(1)$. Moreover, $\sqrt{N}a_{2,N} = \sqrt{N}\delta_N O_P(1)$ and $\sqrt{N}a_{3,N} = \sqrt{\delta_N}O_P(1)$ tend to 0 as $N \to \infty$ for $N\delta_N^2 = o(1)$. To study the main term, let us set $u_i = \frac{\rho^2}{\sqrt{N}}(\varepsilon_{i\delta_N}^2 - 1 - \varepsilon_{(i-1)\delta_N}\varepsilon_{i\delta_N})$ so that $\sqrt{N}a_{1,N} = \sum_{i=1}^{N-1} u_i + o_P(1)$. With

$$\mathbb{E}(u_i|\varepsilon_{\ell\delta_N}, \ell \leq i-1) = 0, \quad \sum_{i=1}^{N-1}\mathbb{E}(u_i^2|\varepsilon_{\ell\delta_N}, \ell \leq i-1) = 3\rho^4 + o_P(1),$$
$$\mathbb{E}(u_i^4|\varepsilon_{\ell\delta_N}, \ell \leq i-1) = o_P(1),$$

we conclude by the Central Limit Theorem for martingale arrays. □

Proof of Theorem 4.3. For this proof, recall that $b(.) = b(., \kappa_0)$, $c(.) = c(., \lambda_0)$ denote the drift and diffusion coefficients at the true value θ_0. Developping $\mathcal{E}_N(\theta)$ (see (4.22)) yields :

$$\begin{aligned}\mathcal{E}_N(\theta) &= k_N\left\{\frac{3}{2}\bar{Q}_N(\frac{1}{c(.,\lambda)}) + \bar{\nu}_N(\log(c(.,\lambda)))\right\}\\ &\quad +3k_N\Delta_N\left\{\frac{1}{2}\bar{\nu}_N\left(\frac{b(.,\kappa)^2 - 2b(.,\kappa)b(.,\kappa_0)}{c(.,\lambda)}\right) - \bar{I}_N\left(\frac{b(.,\kappa)}{c(.,\lambda)}\right)\right\}.\end{aligned}$$

Proposition 4.4, Theorem 4.1 and Theorem 4.2 imply that $\mathcal{E}_N(\theta)$ is the sum of two terms with different rates of convergence. Therefore, to prove consistency of $\hat{\theta}_N$, we must proceed in two steps as in Kessler (1997) and Gloter (2006). It is enough to prove that, first,

$$\frac{1}{k_N}\mathcal{E}_N(\theta) \underset{N\to\infty}{\longrightarrow} \nu_0\left(\frac{c(.,\lambda_0)}{c(.,\lambda)} + \log(c(.,\lambda))\right) \tag{4.41}$$

in probability, uniformly in θ. This will ensure the convergence of $\hat{\lambda}_N$ to λ_0. Second, we prove that

$$\frac{1}{k_N\Delta_N}(\mathcal{E}_N(\kappa,\lambda) - \mathcal{E}_N(\kappa_0,\lambda)) \underset{N\to\infty}{\longrightarrow} \frac{3}{2}\nu_0\left(\frac{(b(.,\kappa) - b(.,\kappa_0)^2}{c(.,\lambda)}\right) \tag{4.42}$$

in probability, uniformly in θ.

Using Theorem 4.1, Theorem 4.2 and Proposition 4.4, with $\Delta_N \to 0$ we obtain (4.41) and (4.42).

For the second case, we have $\|c_{N,\rho}(.,\lambda) - c_\rho(.,\lambda)\|_\infty = 0$ if $\alpha = 2$, and

$$\|c_{N,\rho}(.,\lambda) - c_\rho(.,\lambda)\|_\infty \leq 3\Delta_N^{\frac{2-\alpha}{\alpha-1}}\rho^2 \text{ if } \alpha \in (1,2).$$

Then, $c_{N,\rho}$ converges uniformly (in (x,λ)) to c_ρ. Moreover, by Assumption **(A7)**, c^{-1} satisfies **(C1)**. Thus

$$|c_{N,\rho}(x,\lambda)^{-1} - c_\rho(x,\lambda)^{-1}| \leq c\|c_{N,\rho}(.,\lambda) - c_\rho(.,\lambda)\|_\infty(1+|x|^4)$$

and

$$|\log(c_{N,\rho}(x,\lambda)) - \log(c_\rho(x,\lambda))| \leq c\|c_{N,\rho}(.,\lambda) - c_\rho(.,\lambda)\|_\infty(1+|x|^2).$$

The end of the proof is identical, replacing $\mathcal{E}_N$ by $\mathcal{E}_N^\rho$ and c by c_ρ in the limits (4.41)-(4.42).□

Proof of Corollary 4.2. As formerly, we evaluate

$$\|c_{N,\hat{\rho}_N}(.,\lambda) - c_\rho(.,\lambda)\|_\infty \leq 3\Delta_N^{\frac{2-\alpha}{\alpha-1}}|\rho^2 - \hat{\rho}_N^2|.$$

We conclude using Lemma 4.1. □

Appendices

Annexe B

appendix

Appendix

The following lemma can be found in Gloter (2006), and precises a result from Kessler (1997) :

Lemme B.1 *Assume* **(A1)-(A3)**. *Let $f \in \mathcal{C}^1(\mathbb{R}\times O)$, where O is an open neighbourhood of Θ, satisfy*

$$\sup_{\theta\in\Theta}\{|f(x,\theta)| + |\partial_x f(x,\theta)| + |\partial_\theta f(x,\theta)|\} \leq C(1+|x|)$$

then :

$$\frac{1}{k_N}\sum_{j=0}^{k_N-1} f(X_{j\Delta_N},\theta) \underset{k_N\to\infty}{\longrightarrow} \nu_0(f(.,\theta)) \tag{B.1}$$

uniformly in θ, in probability.

The following proposition can be found in Gloter (2000) and Gloter (2006), and the numerical constant c may varies.

Proposition B.1 *Assume* **(A1)** *and let $f \in \mathcal{C}^1(\mathbb{R})$ satisfy :*

$$\exists\gamma \geq 0, \exists c > 0, \forall x \in \mathbb{R} |f'(x)| \leq c(1+|x|).$$

Then for all integer $k \geq 1$, there exists $c > 0$ such that, for all $j \geq 0$:

$$\mathbb{E}\left(\sup_{s\in[j\Delta_N,(j+1)\Delta_N]} |f(X_s) - f(X_{j\Delta_N})|^k |\mathcal{G}_j^N\right) \leq c\Delta_N^{\frac{k}{2}}(1+|X_{j\Delta_N}|^{1+k})$$

In particular, with $f(x) = x$, we have :

$$\mathbb{E}\left(\sup_{s\in[j\Delta_N,(j+1)\Delta_N]}|X_s - X_{j\Delta_N}|^k \middle| \mathcal{G}_j^N\right) \leq c\Delta_N^{k/2}(1+|X_{j\Delta_N}|^k).$$

We also recall the following lemma which is given in Genon-Catalot and Jacod (1993), setting $\mathcal{G}_j^N = \mathcal{G}_{j\Delta_N}$

Lemme B.2 *Let χ_j^N, U be random variables, with χ_j^N being $\mathcal{G}_j^N$-measurable. The following two conditions :*

$$\begin{array}{c}\sum_{j=0}^{k_N-1}\mathbb{E}(\chi_j^N|\mathcal{G}_{j-1}^N) \xrightarrow{\mathbb{P}} U,\\ \sum_{j=0}^{k_N-1}\mathbb{E}((\chi_j^N)^2|\mathcal{G}_{j-1}^N) \xrightarrow{\mathbb{P}} 0\end{array}$$

imply $\sum_{j=0}^{k_N-1}\chi_j^N \xrightarrow{\mathbb{P}} U$.

The following proposition is given in Gloter (2006), to obtain convergences in probability uniformly in θ.

Proposition B.2 *Let $S_n(\omega,\theta)$ be a sequence of measurable real valued functions defined on $\Omega\times\Theta$ where $(\Omega,\mathcal{F},\mathbb{P})$ is a probability space, and Θ is product of compact intervals of $\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$. We assume that $S_n(.,\theta)$ converges to zero in probability for all $\theta\in\Theta$ and that there exists an open neigbourhood of Θ on which $S_n(\omega,.)$ is continuously differentiable for all $\omega\in\Omega$. Furthermore, we suppose that $\sup_{n\in\mathbb{N}}\mathbb{E}(\sup_{\theta\in\Theta}|\nabla_\theta S_n(\theta)|) < \infty$. Then*

$$S_n(\theta) \to 0$$

uniformly in θ, in probability.

Lemme B.3 *The random variables $\xi_{j,N}$ and $\xi'_{j+1,N}$ are independent and gaussian ; $\xi_{j,N}$ is $\mathcal{G}_{j+1}^N$ measurable and independent of $\mathcal{G}_j^N$; $\xi'_{j+1,N}$ is $\mathcal{G}_{j+2}^N$ measurable and independent of $\mathcal{G}_{j+1}^N$. We will use the following expectations :*

$$\begin{array}{c}\mathbb{E}(\xi_{j,N}|\mathcal{G}_j^N) = \mathbb{E}(\xi'_{j+1,N}|\mathcal{G}_j^N) = 0,\\ \mathbb{E}(\xi_{j,N}^2|\mathcal{G}_j^N) = \mathbb{E}({\xi'}_{j+1,N}^2|\mathcal{G}_j^N) = \frac{1}{3},\\ \mathbb{E}((\xi_{j,N}^2-\frac{1}{3})^2|\mathcal{G}_j^N) = \mathbb{E}(({\xi'}_{j+1,N}^2-\frac{1}{3})^2|\mathcal{G}_j^N) = \frac{2}{9},\\ \mathbb{E}((\xi_{j,N}^2-\frac{1}{3})\xi'_{j,N}|\mathcal{G}_j^N) = \mathbb{E}(({\xi'}_{j+1,N}^2-\frac{1}{3})\xi'_{j,N}|\mathcal{G}_j^N) = 0,\\ \mathbb{E}(\xi_{j,N}\xi'_{j,N}|\mathcal{G}_j^N) = \frac{1}{6}.\end{array}$$

This lemma, based on elementary computations, is mentioned in Gloter (2000).

Chapitre 5

Estimating equations for noisy observations of ergodic diffusions

Abstract

In this chapter, general estimating functions for ergodic diffusions sampled at high frequency with noisy observations are presented. The theory is formulated in term of approximate martingale estimating functions based on local means of the observations, and simple conditions are given for rate optimality. The estimation of diffusion parameter is faster that the estimation of drift parameter, and the rate of convergence in the Central Limit Theorem is classical for the drift parameter but not classical for the diffusion parameter. The link with specific minimum contrast estimators is established, as an example.

5.1 Introduction

The aim of this chapter is to study estimating functions based on the observations of a noisy discretely observed one-dimensional diffusion inspired by Sørensen (2009). The notations are the same as in the previous chapter. Recall that we consider the one-dimensional stochastic differential equation

$$dX_t = b(X_t, \kappa)dt + \sigma(X_t, \lambda)dB_t, \qquad X_0 = \eta \tag{5.1}$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion and η is a real valued random variable independent of B. The functions $b(x, \kappa)$ and $\sigma(x, \lambda)$ are respectively defined on $\mathbb{R} \times \Theta_1$ and $\mathbb{R} \times \Theta_2$ where Θ_1) (resp. Θ_2) is a compact convex subset of $\mathbb{R}^{d_1}$ (resp. $\mathbb{R}^{d_2}$). For simplicity of notations, in the proofs, we assume that $d_1 = d_2 = 1$. We denote by $\theta_0 = (\kappa_0, \lambda_0)$ the true value of the parameter and assume that $\theta_0 \in \overset{\circ}{\Theta}$ where $\Theta = \Theta_1 \times \Theta_2$. We set $\mathbb{E}_\theta$ the expectation under the probability distribution $\mathbb{P}_\theta$, for $\theta \in \Theta$, and $\mathbb{E}_{\theta_0}$ the expectation under the probability distribution $\mathbb{P}_{\theta_0}$.

At time $t_i = i\delta_N, i = 0, \ldots, N$, with δ_N the sampling time, the observation $Y_{i\delta_N}$ is given by

$$Y_{i\delta_N} = X_{i\delta_N} + \rho_N \varepsilon_{i\delta_N}$$

where ρ_N is the standard deviation of the observation noise, and $(\varepsilon_{i\delta_N})$ is a sequence of independent and identically distributed centered random variables, independent of the diffusion (X_t), with unitary variance. Two cases are considered for ρ_N :

– in the case **(B1)**, $\rho_N = \rho > 0$,

– whereas in the case **(B2)**, $\rho_N \to 0$.

In this chapter, ρ_N is assumed to be known.

In the previous chapter, we have studied minimum contrast estimators based on the local means of the $(Y_{i\delta_N})$: with p_N and k_N such that $N = p_N k_N$, we set $\Delta_N = p_N \delta_N$ and

$$Y_\bullet^j = \frac{1}{p_N} \sum_{i=0}^{p_N - 1} Y_{j\Delta_N + i\delta_N}$$

with the asymptotic framework : $N \to \infty$, $\delta_N \to 0$, $t_N = N\delta_N = k_N \Delta_N \to \infty$ and $\delta_N = p_N^{-\alpha}$, $\alpha \in (1, 2]$. The parameter α is the *local mean size parameter*. The case $\alpha = 2$ is specific because the variance ρ_N^2 has to be taken into account in the results.

Our focus is on approximate martingale estimating functions based on the $(Y_\bullet^j)$. An exhaustive review on estimating functions for diffusion processes can

be found in Sørensen (2010). We consider estimating functions depending on $\alpha \in (1, 2]$

$$G_{N,\alpha}(\theta) = \sum_{j=1}^{k_N-2} g_\alpha(\delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta, \rho_N) \tag{5.2}$$

where the function $g_\alpha(\delta, y, x; \theta, \rho)$ with values in $\mathbb{R}^2$ is such that $G_{N,\alpha}$ is approximately a martingale estimating function. Under simple assumptions, we show that the estimator $\hat{\theta}_N$ given as the solution of $G_{N,\alpha}(\hat{\theta}_N) = 0$ is consistent and asymptotically Gaussian. We denote by $(g_{1,\alpha}, g_{2,\alpha})^T$ the two components of g_α. Notice the lag in formula (5.2) that must be introduced as $(Y_\bullet^j)$ is not Markov.

We assume that $g_\alpha(\delta, y, x; \theta, \rho)$ satisfies the conditions of Sørensen (2009) which are the following ones. First, the condition for *rate optimality* is

$$\partial_y g_{2,\alpha}(0, 0, x; \theta, \rho) = 0 \tag{5.3}$$

for all $x \in \mathbb{R}$, all $\rho > 0$ and all $\theta \in \Theta$. This condition is also called *Jacobsen's condition* in Sørensen (2009), and corresponds to one of the conditions obtained in Jacobsen (2002).

By $\partial_y g_{2,\alpha}(0, 0, x; \theta, \rho) = 0$, we mean $\partial_y g_{2,\alpha}(0, y, x; \theta, \rho) = 0$ evaluated at $y = 0$. For directly observed diffusion models, rate optimality is important because the diffusion coefficient parameter can be estimated at a higher rate than the drift parameter, and we show that the result remains true for a diffusion observed with a noise.

The second condition, called the *second Jacobsen's condition*, is

$$\partial_y g_{1,\alpha}(0, 0, x; \theta, \rho) = \frac{\partial_\kappa b(x, \kappa)}{c(x, \lambda)} \quad \text{and} \quad \partial_{y^2}^2 g_{2,\alpha}(0, 0, x; \theta, \rho) = \frac{\partial_\lambda c(x, \lambda)}{c(x, \lambda)^2} \tag{5.4}$$

for all $x \in \mathbb{R}$, all $\rho > 0$ and all $\theta \in \Theta$, where $c(x, \lambda) = \sigma(x, \lambda)^2$. In Sørensen (2009), this condition ensures the efficiency of the estimators, in the case of direct observations of an ergodic diffusion.

The case of martingale estimating functions for discrete observations of diffusion processes has been treated in Bibby and Sørensen (1995), and the case of estimating functions that do not have the martingale property has been treated in Kessler (2000), with a fixed sampling time Δ. Kessler and Sørensen (1999) introduce martingale estimating functions based on the eigenfunctions of the infinitesimal generator of the diffusion. Sørensen (2009) focuses on the high-frequency asymptotics for an ergodic diffusion.

This chapter is organized as follows : Section 5.2 is devoted to the presentation of the model and the assuptions, which are similar to those presented in Chapter 4, and Section 5.3 contains the main results for the convergence in distribution

of the variation and the quadratic variation of the local means, and the result of asymptotic normality for the estimators associated to the estimating functions. Section 5.4 is devoted to the examples, and the proofs are gathered in Section 5.7.

5.2 Model and Assumptions

In this section, assumptions on the model and the estimating functions are precised. Assumptions on the diffusion (X_t) are identical to those in the previous chapter (see Section 4.2 in Chapter 4, Assumptions **(A1)-(A7)**).

We define the class $\mathcal{C}_{p_1,p_2,p_3}(\mathbb{R}_+ \times \mathbb{R}^2 \times \Theta \times \mathbb{R}_+)$ of real functions $f(t, y, x; \theta, \rho)$ satisfying that

1. f is p_1 times continuously differentiable in t, p_2 times continuously differentiable in y and p_3 continuously differentiable in κ and λ;
2. f and all partial derivatives of f are of polynomial growth uniformly for $\theta \in \Theta$.

The class $\mathcal{C}_{p_1,p_2,p_3}(\mathbb{R}^2 \times \Theta \times \mathbb{R}_+)$ is defined in an similar way for functions $f(y, x; \theta, \rho)$. A function $f(y, x; \theta, \rho)$ is said to be of polynomial growth in y and x uniformly for $\theta \in \Theta$ (recall that Θ is assumed to be a compact subset of $\mathbb{R}^2$) if there exists a constant $c > 0$ such that, for all $x, y \in \mathbb{R}$ and all $\rho > 0$,

$$\sup_{\theta \in \Theta} |f(y, x, ; \theta, \rho)| \leq c(1 + |x|^c + |y|^c).$$

In the sequel of this chapter, $R(\delta, y, x; \theta, \rho)$ denotes a generic function such that $|R(\delta, y, x; \theta, \rho)| \leq F(y, x; \theta, \rho)$ where F is of polynomial growth in y,x and ρ, uniformly in θ. We assume that the function g_α belongs to the class $\mathcal{C}_{1,3,2}$ in the sequel of the chapter.

Moreover, we stress the fact that $\delta_N = p_N^{-\alpha}$, hence $\Delta_N = p_N \delta_N = p_N^{1-\alpha} = \delta_N^{1-\frac{1}{\alpha}}$, for a given $\alpha \in (1, 2]$. We will discuss the choice of α in the examples.

We have to consider conditional expectations given $\mathcal{H}_j^N = \sigma(X_s, 0 \leq s \leq j\Delta_N) \vee \sigma(\varepsilon_{i\delta_N}, i\delta_N \leq (j-1)\Delta_N + (p_N - 1)\delta_N)$, such that $Y_\bullet^j$ is $\mathcal{H}_{j+1}^N$ measurable.

For sake of simplicity in the notations, we write $g_\alpha = g$, $g_{1,\alpha} = g_1$ and $g_{2,\alpha} = g_2$ in the sequel. We consider estimating functions where $g(\delta, y, x; \theta, \rho)$ satisfies the following condition **(D)** :

1. For all $\theta \in \Theta$ we have

$$\begin{aligned} \mathbb{E}_\theta(g(\delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta, \rho_N)|\mathcal{H}_j^N) &= \delta_N^{2-\frac{2}{\alpha}} R(\delta_N, Y_\bullet^{j-1}, X_{j\Delta_N}; \theta, \rho_N) \\ &= \Delta_N^2 R(\delta_N, Y_\bullet^{j-1}, X_{j\Delta_N}; \theta, \rho_N), \end{aligned}$$

2. The function $g(\delta, y, x; \theta, \rho)$ has an expansion in power of δ : there exist functions $g^{(1)}$ and $g^{(\alpha)}$ such that

   (a) if $\alpha \in (1, \frac{3}{2}]$ or $\alpha = 2$, the expansion is

   $$g(\delta, y, x; \theta, \rho) = g(0, y, x; \theta, \rho) + \delta^{1-\frac{1}{\alpha}} g^{(1)}(y, x; \theta, \rho) + \delta^{2-\frac{2}{\alpha}} R(\delta, y, x; \theta, \rho);$$

   (b) if $\alpha \in (\frac{3}{2}, 2)$, the expansion is

   $$\begin{aligned} g(\delta, y, x; \theta, \rho) &= g(0, y, x; \theta, \rho) + \delta^{1-\frac{1}{\alpha}} g^{(1)}(y, x; \theta, \rho) \\ &\quad + \delta^{\frac{1}{\alpha}} g^{(\alpha)}(y, x; \theta, \rho) + \delta^{2-\frac{2}{\alpha}} R(\delta, y, x; \theta, \rho) \end{aligned}$$

   with $R(\Delta, y, x; \theta, \rho)$ a generic function such that $|R(\Delta, y, x; \theta, \rho)| \leq F(y, x; \theta, \rho)$ where F is of polynomial growth in y and x uniformly in θ.

The condition on the expansion has to be discussed : for $\alpha \in (1, \frac{3}{2}]$, we have $\frac{1}{\alpha} \geq 2 - \frac{2}{\alpha}$, and for $\alpha = 2$, we have $1 - \frac{1}{\alpha} = \frac{1}{\alpha}$, whereas for $\alpha \in (\frac{3}{2}, 2)$, we have $2 - \frac{2}{\alpha} < \frac{1}{\alpha} < 1 - \frac{1}{\alpha}$. Notice that $\delta_N^{1-\frac{1}{\alpha}} = \Delta_N$ and $\delta_N^{\frac{1}{\alpha}} = \Delta_N^{\frac{1}{\alpha-1}}$, this term is needed to take into account the noise.

Finally, for any non-singular matrix M_N, the estimating functions $G_{N,\alpha}$ and $M_N G_{N,\alpha}$ give the same estimator, and the matrix M_N may depend on δ_N. Therefore, a given version of an estimating function may not satisfy the above condition, but the point is that one version, up to a matrix M_N must exist and satisfy this condition. For example, it may be necessary to multiply one of the coordinates by $\Delta_N = \delta_N^{1-\frac{1}{\alpha}}$.

5.3 Rate-optimal estimating functions for local means

In this section, we give asymptotic results for approximate martingale estimating functions. We prove that the estimator of the parameter in the diffusion coefficient converges faster than the estimator of the parameter in the drift coefficient.

The infinitesimal generator of the diffusion (X_t) is defined by

$$L_\theta = b(x, \kappa)\frac{d}{dx} + \frac{1}{2}c(x, \lambda)\frac{d^2}{dx^2}.$$

For a function $h(y, x)$ of two variables, we use the notation

$$L_\theta(h(\delta, \theta', \rho))(y, x) = b(x, \kappa)\partial_y h(\delta, y, x; \theta', \rho) + \frac{1}{2}c(x, \lambda)\partial_{y^2}^2 h(\delta, y, x; \theta', \rho).$$

We also introduce the *modified* generator

$$\begin{aligned}\bar{L}_\theta &= b(x,\kappa)\frac{d}{dx} + \frac{1}{3}c(x,\lambda)\frac{d^2}{dx^2} \\ &= L_\theta - C_\theta\end{aligned}$$

with $C_\theta = \frac{1}{6}c(x,\lambda)\frac{d^2}{dx^2}$.

Indeed, we derive from Proposition 5.1 (proved in the previous chapter and recalled below in 5.6) that, for f a twice continuously differentiable real function with bounded second derivative, with $\Delta_N = p_N\delta_N = \delta_N^{1-\frac{1}{\alpha}}$,

$$\begin{aligned}\mathbb{E}_{\theta_0}(f(Y_\bullet^{j+1} - Y_\bullet^j)|\mathcal{H}_j^N) &= f(0) + f'(0)\mathbb{E}_{\theta_0}(Y_\bullet^{j+1} - Y_\bullet^j|\mathcal{H}_j^N) \\ &\quad + \frac{1}{2}f''(0)\mathbb{E}_{\theta_0}((Y_\bullet^{j+1} - Y_\bullet^j)^2|\mathcal{H}_j^N) + \Delta_N o_P(1) \\ &= f(0) + \Delta_N(f'(0)b(X_{j\Delta_N},\kappa_0) + \frac{1}{3}f''(0)c(X_{j\Delta_N},\lambda_0)) \\ &\quad + \Delta_N^{\frac{1}{\alpha-1}}f''(0)\rho_N^2 + \Delta_N o_P(1)\end{aligned}$$

by Taylor's formula.

The following lemma provides an essential identity for the sequel of this chapter.

Lemme 5.1 *Under Condition* **(D)**, *we have, for all* $x \in \mathbb{R}$, *all* $\theta \in \Theta$ *and all* $\rho > 0$

$$g(0,0,x;\theta,\rho) = 0, \tag{5.5}$$

$$g^{(1)}(0,x;\theta,\rho) = -\bar{L}_\theta(g(0,\theta,\rho)(0,x) \text{ if } \alpha \in (1,2), \tag{5.6}$$

$$g^{(\alpha)}(0,x;\theta,\rho) = -\rho^2\partial_{y^2}^2 g(0,0,x;\theta,\rho) \text{ if } \frac{3}{2} < \alpha < 2, \tag{5.7}$$

$$g^{(1)}(0,x;\theta,\rho) = -\bar{L}_\theta(g(0,\theta,\rho)(0,x) - \rho^2\partial_{y^2}^2 g(0,0,x,\theta,\rho) \text{ if } \alpha = 2. \tag{5.8}$$

Remark. This lemma has to be compared with the result in Sørensen (2009), when the diffusion is directly observed. For $\alpha \in (1,2)$, the result on $g^{(1)}$ is the same except that it involves $\bar{L}_\theta$ in place of L_θ. For $\alpha \in (\frac{3}{2},2)$, the rate of sampling has to be taken into account, by the additional term $g^{(\alpha)}$ which involves the variance ρ^2 of the observation noise. In the case $\alpha = 2$, we have $\Delta_N = p_N^{-1}$ and an additional term is needed in $g^{(1)}$.

Setting $\rho_\infty = \lim_{N\to\infty}\rho_N$, two cases have to be distinguished for the limit theorems :

- Assume **(B1)** or**(B2)** when $\alpha \in (1,2)$, or $\alpha = 2$ and $\rho_\infty = 0$,
- Assume **(B1)** when $\alpha = 2$.

We define

$$\begin{aligned}\gamma(\theta,\theta_0,\rho_\infty) &= \nu_0((b(.,\kappa_0)-b(.,\kappa))\partial_y g(0,0,.;\theta,\rho_\infty)) \\ &\quad +\frac{1}{3}\nu_0((c(.,\lambda_0)-c(.,\lambda))\partial^2_{y^2} g(0,0,.;\theta,\rho_\infty)),\end{aligned} \tag{5.9}$$

and the matrix

$$\bar{A}_\theta(.,\rho)=\begin{pmatrix} \partial_\kappa b(.,\kappa)\partial_y g_1(0,0,.;\theta,\rho) & \frac{1}{3}\partial_\lambda c(.,\lambda)\partial^2_{y^2} g_1(0,0,.;\theta,\rho) \\ \partial_\kappa b(.,\kappa)\partial_y g_2(0,0,.;\theta,\rho) & \frac{1}{3}\partial_\lambda c(.,\lambda)\partial^2_{y^2} g_2(0,0,.;\theta,\rho)\end{pmatrix}.$$

Notice that $\bar{A}_\theta(x,\rho)=\partial_\theta \bar{L}_\theta(g(0,\theta,\rho))(0,x)=\partial_\theta L_\theta(g(0,\theta,\rho))(0,x)-\partial_\theta C_\theta(g(0,\theta,\rho))(0,x)$. We set $S=\nu_0(A_{\theta_0}(.,\rho_\infty))$. We derive the identity

$$\begin{aligned}\partial_\theta g^{(1)}(0,x;\theta,\rho) &= -\partial_\theta L_\theta(g(0,\theta,\rho))(0,x)-L_\theta(\partial_\theta g(0,\theta,\rho))(0,x) \\ &\quad +\partial_\theta C_\theta(g(0,\theta,\rho))(0,x)+C_\theta(\partial_\theta g(0,\theta,\rho))(0,x).\end{aligned}$$

Setting

$$\phi(\theta,\theta_0,\rho_\infty)=\nu_0(\bar{L}_{\theta_0}(\partial_\theta g(0;\theta,\rho_\infty))(0,.)-\bar{L}_\theta(\partial_\theta g(0;\theta,\rho_\infty))(0,.))-\nu_0(\bar{A}_\theta(.,\rho_\infty)),$$

the following lemma contains the main statements for convergence in probability results (∂_θ is the matrix of partial derivatives with respect to κ and λ).

Lemme 5.2 *We have*

$$\frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2} g(\Delta_N, Y_\bullet^{j+1}-Y_\bullet^j, Y_\bullet^{j-1};\theta,\rho_N) \xrightarrow{\mathbb{P}_{\theta_0}} \gamma(\theta,\theta_0,\rho_\infty), \tag{5.10}$$

$$\frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2} \partial_\theta g(\Delta_N, Y_\bullet^{j+1}-Y_\bullet^j, Y_\bullet^{j-1};\theta,\rho_N) \xrightarrow{\mathbb{P}_{\theta_0}} \phi(\theta,\theta_0,\rho_\infty) \tag{5.11}$$

uniformly in $\theta\in\Theta$ compact.

Recall that, for $f=f(.,\theta_0)$ a function satisfying Condition **(C1)** (see Chapter 4)

$$\bar{\nu}_N(f)=\frac{1}{k_N}\sum_{j=0}^{k_N-1} f(Y_\bullet^j),$$

$$\bar{I}_N(f)=\frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1})(Y_\bullet^{j+1}-Y_\bullet^j-\Delta_N b(Y_\bullet^{j-1})),$$

$$\bar{Q}_N(f)=\frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1})(Y_\bullet^{j+1}-Y_\bullet^j)^2,$$

where $b = b(., \kappa_0)$.

The following theorems precise the asymptotic behaviour of the functionals $\bar{I}_N$ and $\bar{Q}_N$ of the $(Y_\bullet^j)$'s, and are useful to study the asymptotic normality of the estimator $\hat{\theta}_N$.

Théorème 5.1 *Assume* **(A1)-(A7)** *and* **(B1)** *or* **(B2)**. *Then, if* $\alpha \in (1, 2]$ *and* $N\delta_N^{3-\frac{2}{\alpha}} \to 0$, *we have*

$$\sqrt{N\delta_N}\bar{I}_N(f) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu_0(f^2\sigma^2)).$$

A comment has to be done about the assumption $N\delta_N^{3-\frac{2}{\alpha}} \to 0$: when the diffusion is directly observed at time $i\delta_N$, a similar result holds under the condition $N\delta_N^2 \to 0$. Here, we need that $k_N\Delta_N^3 \to 0$, and $k_N\Delta_N^3 = N\delta_N^{3-\frac{2}{\alpha}}$. When $\alpha = 2$, the usual condition $N\delta_N^2 \to 0$ is obtained, but when $\alpha \in (1, 2)$, a stronger assumption is needed on the sampling rate than if the discretized diffusion process were observed directly.

Now we precise the convergence in distribution for the quadratic variation.

Théorème 5.2 *Assume* **(A1)-(A7)**, *and* $N\delta_N^{2-\frac{1}{\alpha}} \to 0$. *Then*

– *if* **(B1)** *or* **(B2)** *with* $\alpha \in (1, 2)$ *or* $\alpha = 2$ *and* **(B2)** *(*$\rho_N \to 0$*), we have*

$$\sqrt{N\delta_N^{\frac{1}{\alpha}}}(\bar{Q}_N(f) - \bar{\nu}_N(f(\frac{2}{3}\sigma^2 + 2\rho_N^2\Delta_N^{\frac{2-\alpha}{\alpha-1}}))) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu_0(f^2\sigma^4));$$

– *if* $\alpha = 2$ *and* **(B1)** *(*$\rho_N = \rho$*) , we have*

$$\sqrt{N\delta_N^{\frac{1}{\alpha}}}\left(\bar{Q}_N(f) - \bar{\nu}_N\left(f\left(\frac{2}{3}\sigma^2 + 2\rho^2\right)\right)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu_0(f^2(\sigma^4+4\sigma^2\rho^2+12\rho^4)).$$

In this second theorem, we need that $N\delta_N^{2-\frac{1}{\alpha}} = k_N\Delta_N^2 \to 0$. This condition is more stringent that $N\delta_N^{2-\frac{2}{\alpha}} \to 0$: for $\alpha = 2$, we may assume that $N\delta_N^{\frac{3}{2}} \to 0$. Moreover, a distinction has to be done in the case $\alpha = 2$, $\rho_N = \rho$, because the variance of the observation noise ρ^2 appears in the asymptotic variance, which is increased significantly. Notice also that the rate $k_N = N\delta_N^{\frac{1}{\alpha}}$ is not classical : when the diffusion is directly observed, the usual rate of convergence for a similar central limit theorem is N, whereas here it is slower. That explains why the condition $N\delta_N^{2-\frac{1}{\alpha}}$ is more stringent in the case of diffusions observed with noise.

And there is the multivariate version of the two last theorems.

Théorème 5.3 *Assume* **(A1)-(A7)**, *and* $N\delta_N^{2-\frac{1}{\alpha}} \to 0$. *Then, for* f *a function which satisfies* **(C1)**, *and* $\alpha \in (1,2]$,

$$\begin{pmatrix} \sqrt{N\delta_N}\bar{I}_N(f) \\ \sqrt{N\delta_N^{\frac{1}{\alpha}}}\left(\bar{Q}_N(f) - \bar{\nu}_N\left(f\left(\frac{2}{3}\sigma^2 + 2\Delta_N^{\frac{2-\alpha}{\alpha-1}}\rho_N^2\right)\right)\right) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} W_1(f) & 0 \\ 0 & W_2(f) \end{pmatrix}\right) \tag{5.12}$$

with, in the case $\alpha \in (1,2)$ *and* **(B1)** *or* **(B2)**, *or in the case* $\alpha = 2$ *with* **(B2)**,

$$W_1(f) = \nu_0(f^2\sigma^2) \text{ and } W_2(f) = \nu_0(f^2\sigma^4)$$

and in the case $\alpha = 2$, *with* **(B1)**,

$$W_1(f) = \nu_0(f^2\sigma^2) \text{ and } W_2(f) = \nu_0(f^2(\sigma^4 + 4\rho^2\sigma^2 + 12\rho^4)).$$

For the applications, the following corollary is needed.

Corollaire 5.1 *Assume* **(A1)-(A7)**, $N\delta_N^{2-\frac{1}{\alpha}} \to 0$ *for an* $\alpha \in (1,2]$. *Consider a sequence of functions* $f_N(x,\theta)$ *and a function* $f(x,\theta)$ *satisfying* **(C1)**, *a sequence* $v_N \to 0$ *such that, for all* x,

$$\sup_{\theta\in\Theta} |f_N(x,\theta) - f(x,\theta)| \leq v_N(1+|x|),$$

then

$$\begin{pmatrix} \sqrt{N\delta_N}\bar{I}_N(f_N) \\ \sqrt{N\delta_N^{\frac{1}{\alpha}}}\left(\bar{Q}_N(f_N) - \bar{\nu}_N\left(f_N\left(\frac{2}{3}\sigma^2 + 2\Delta_N^{\frac{2-\alpha}{\alpha-1}}\rho_N^2\right)\right)\right) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} W_1(f) & 0 \\ 0 & W_2(f) \end{pmatrix}\right). \tag{5.13}$$

Now we can state the main theorem about the estimation of $\theta_0 = (\kappa_0, \lambda_0)$ and the asymptotic behaviour of the estimator $\hat{\theta}_N = (\hat{\kappa}_N, \hat{\lambda}_N)$.

With the *rate optimality* condition and the *second Jacobsen's condition*, the Central Limit Theorem gives the optimal rate of convergence for the estimator of the parameter involved in the diffusion coefficent.

Théorème 5.4 *Assume* **A1)-(A7)**, **(D)**, *and that the following identifiability condition*

$$\begin{cases} \nu_0((b(.,\kappa_0) - b(.,\kappa))\partial_y g_1(0,0,.;\theta,\rho_\infty)) \neq 0 & \text{for } \kappa \neq \kappa_0 \\ \nu_0((c(.,\lambda_0) - c(.,\lambda_0))\partial_{y^2}^2 g_2(0,0,.;\theta,\rho_\infty)) \neq 0 & \text{for } \lambda \neq \lambda_0 \end{cases}$$

is satisfied. Define $S = \nu_0(A_{\theta_0}(.,\rho_\infty))$. *Assume that* S *is invertible,* $S_{1,1} \neq 0, S_{2,2} \neq 0$, *and* $\partial_\kappa \partial_{y^2}^2 g_2(0,0,x;\theta,\rho) = 0$ *for all* θ, x *and* ρ, *then there exists a consistent*

estimator $\hat{\theta}_N$ *solution of* $G_{N,\alpha}(\hat{\theta}_N) = 0$*, is unique with a probabily that goes to one as* $N \to \infty$*, and asymptotically Gaussian :*

$$\begin{pmatrix} \sqrt{N\delta_N}(\hat{\kappa}_N - \kappa_0) \\ \sqrt{N\delta_N^{\frac{1}{\alpha}}}(\hat{\lambda}_N - \lambda_0) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}_2, \begin{pmatrix} \frac{W_1(\partial_y g_1(0,0,.;\theta_0,\rho_\infty))}{S_{1,1}^2} & 0 \\ 0 & \frac{W_2(\partial_{y^2}^2 g_2(0,0,.;\theta_0,\rho_\infty))}{4S_{2,2}^2} \end{pmatrix}\right) \tag{5.14}$$

as $N \to \infty$*, with* $N\delta_N \to \infty$ *and* $N\delta_N^{2-\frac{1}{\alpha}} \to 0$*.*

The assumption $N\delta_N^{2-\frac{1}{\alpha}}$ in this theorem corresponds to $k_N\Delta_N^2 \to 0$. In the case of direct observations, the usual condition is $N\delta_N^2 \to 0$ (see *e.g.* Sørensen (2009)).

One of the main differences between the estimating functions built on the $(Y_\bullet^j)$ and the estimating functions built on the $(X_{j\Delta_N})$ is that the approximation $Y_\bullet^{j+1} - Y_\bullet^j$ is $\mathcal{H}_{j+2}^N$ measurable and involves the random variables $\zeta_{j+1,N}$ and $\zeta'_{j+2,N}$ (defined in (4.8), Chapter 4) which have an MA(1) structure, whereas the approximation $X_{(j+1)\Delta_N} - X_{j\Delta_N}$ involves the random variable $B_{(j+1)\Delta_N} - B_{j\Delta_N}$ and is $\mathcal{G}_{j+1}^N$ measurable. To avoid cumbersome correlations between $Y_\bullet^j$ and $Y_\bullet^{j+1}$, we consider the estimating function for λ_0

$$H_{N,\alpha}(\theta) = \frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2} g_{2,\alpha}(\delta_N, Y_\bullet^{3j+1} - Y_\bullet^{3j}, Y_\bullet^{3j-1}; \theta, \rho_N), \tag{5.15}$$

for a sample $X_{i\delta_N}, i = 0, \ldots, 3N$ of size $3N+1$ in place of $N+1$, with $N = p_N k_N$. The results of uniform convergence in probability under $\mathbb{P}_{\theta_0}$ remain valid :

$$\begin{aligned} H_{N,\alpha}(\theta) &= \frac{1}{3}\nu_0((c(.,\lambda_0) - c(.,\lambda))\partial_{y^2}^2 g_2(0,0,.;\theta,\rho_\infty)) + o_P(1), \\ \partial_\theta H_{N,\alpha}(\theta) &= \nu_0(\bar{L}_{\theta_0}(g_2(0,\theta,\rho_\infty)) - \bar{L}_\theta(g_2(0,\theta,\rho_\infty))) - \frac{1}{3}\nu_0(\partial_\lambda c(.,\lambda)\partial_{y^2}^2 g_2(0,0,.;\theta,\rho_\infty)) + o_P(1) \end{aligned}$$

uniformly in $\theta \in \Theta$. But for the convergence in distribution, we have

Théorème 5.5 *Assume that* $N\delta_N^{2-\frac{1}{\alpha}} \to 0$*. Under* **(A1)-(A7)***, for a function* f *satisfying Condition* **(C1)***,*

$$\sqrt{N\delta_N^{\frac{1}{\alpha}}}\left(\frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2} f(Y_\bullet^{3j-1})(Y_\bullet^{3j+1} - Y_\bullet^{3j})^2 - \frac{1}{k_N}\sum_{j=0}^{k_N-2} f(Y_\bullet^{3j})(\frac{2}{3}\sigma(Y_\bullet^{3j})^2 + 2\rho_N^2\Delta_N^{\frac{2-\alpha}{\alpha-1}})\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, W_3(f))$$

where $W_3(f) = \frac{8}{9}\nu_0(f^2\sigma^4)$ *if* $\alpha \in (1,2)$ *and* **(B1)** *or* **(B2)***, or* $\alpha = 2$ *and* **(B2)***, and* $W_3(f) = \nu_0(f^2(\frac{8}{9}\sigma^4 + 8\rho^4))$ *if* $\alpha = 2$ *and* **(B1)***.*

Then, with an estimator $\hat{\lambda}_N$ of λ_0 based on the estimating function $H_{N,\alpha}$ and a sample of size $3N+1$, we deduce that $\hat{\lambda}_N$ is consistent and asymptotically Gaussian :

$$\sqrt{N\delta_N^{\frac{1}{\alpha}}}(\hat{\lambda}_N - \lambda_0) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{W_3(\partial^2_{y^2} g_2(0,0,.;\theta_0,\rho_\infty))}{4S^2_{2,2}}\right) \tag{5.16}$$

as $N \to \infty$, with $N\delta_N \to \infty$ and $N\delta_N^{2-\frac{1}{\alpha}} \to 0$.

Then, for $\alpha \in (1,2)$, the estimator of λ_0 based on the estimating function $H_{N,\alpha}$ has a better asymptotic variance. If the *second Jacobsen's condition* holds, the asymptotic variance is

$$\frac{W_3(\partial^2_{y^2} g_2(0,0,.;\theta_0,\rho_\infty))}{4S^2_{2,2}} = 2\left\{\nu_0\left(\frac{(\partial_\lambda c(.,\lambda_0))^2}{c(.,\lambda_0)^2}\right)\right\}^{-1}.$$

This is the same asymptotic variance as in the case of direct observations of the diffusion, and the estimator is efficient.

5.4 Applications and examples

The main application of these results about estimating functions is the asymptotic normality of the minimum contrast estimators built in Chapter 4.

Consider the contrast

$$\mathcal{E}_N(\theta) = \sum_{j=1}^{k_N-2}\left\{\frac{3}{2\Delta_N}\frac{(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1},\kappa))^2}{c_N(Y_\bullet^{j-1},\lambda)} + \log(c_N(Y_\bullet^{j-1},\lambda))\right\} \tag{5.17}$$

where $c_N(.,\lambda) = c(.,\lambda) + 3\rho_N^2\Delta_N^{\frac{2-\alpha}{\alpha-1}}$. Then

$$\begin{aligned}
\frac{\partial}{\partial\kappa}\mathcal{E}_N(\theta_0) &= -3\sum_{j=1}^{k_N-2}\left(\frac{\partial_\kappa b(Y_\bullet^{j-1},\kappa_0)}{c_N(Y_\bullet^{j-1},\lambda_0)}\left(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1},\kappa_0)\right)\right),\\
\frac{\partial}{\partial\lambda}\mathcal{E}_N(\theta_0) &= -\sum_{j=1}^{k_N-2}\left(\frac{\partial_\lambda c(Y_\bullet^{j-1},\lambda_0)}{c_N(Y_\bullet^{j-1},\lambda_0)^2}\right)\frac{3}{2\Delta_N}(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1},\lambda_0))^2\\
&\quad + \sum_{j=1}^{k_N-2}\left(\frac{\partial_\lambda c(Y_\bullet^{j-1},\lambda_0)}{c_N(Y_\bullet^{j-1},\lambda_0)^2}\right)(c(Y_\bullet^{j-1},\lambda_0) + 3\Delta_N^{\frac{2-\alpha}{\alpha-1}}).
\end{aligned}$$

Hence, for $\alpha \in (1,2]$, let $G_{N,\alpha}(\theta)$ the estimating function defined by

$$G_{N,\alpha}(\theta) = \frac{1}{k_N\Delta_N}\begin{pmatrix}
\sum_{j=1}^{k_N-2}\frac{\partial_\kappa b(Y_\bullet^{j-1},\kappa)}{c_N(Y_\bullet^{j-1},\lambda)}\left(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1},\kappa)\right)\\
\sum_{j=1}^{k_N-2}\frac{\partial_\lambda c(Y_\bullet^{j-1},\lambda)}{c_N(Y_\bullet^{j-1},\lambda)^2}\left(\frac{1}{2}(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1},\kappa))^2 - \frac{\Delta_N}{3}c(Y_\bullet^{j-1},\lambda) - \Delta_N^{\frac{1}{\alpha-1}}\rho_N^2\right)
\end{pmatrix}$$

with $\Delta_N = \delta_N^{1-\frac{1}{\alpha}}$.

We set $c_\rho(x,\lambda) = c(x,\lambda)$ if $1 < \alpha < 2$ and $c_\rho(x,\lambda) = c(x,\lambda) + 3\rho^2$ if $\alpha = 2$. By Corollary 5.1, as $|c_N(x,\lambda) - c_\rho(x,\lambda)| \leq 3\Delta_N^{\frac{2-\alpha}{\alpha-1}}\rho_N$ if $\alpha \in (1,2)$ or $\alpha = 2$ and $\rho_N \to 0$, and vanishes if $\alpha = 2$ and $\rho_N = \rho$, let $G^\star_{N,\alpha}(\theta)$ be the estimating function defined by

$$G^\star_{N,\alpha}(\theta) = \frac{1}{k_N\Delta_N}\begin{pmatrix} \displaystyle\sum_{j=1}^{k_N-2} \frac{\partial_\kappa b(Y_\bullet^{j-1},\kappa)}{c_\rho(Y_\bullet^{j-1},\lambda)}\left(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1},\kappa)\right) \\ \displaystyle\sum_{j=1}^{k_N-2} \frac{\partial_\lambda c(Y_\bullet^{j-1},\lambda)}{c_\rho(Y_\bullet^{j-1},\lambda)^2}\left(\frac{1}{2}(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1},\kappa))^2 - \frac{\Delta_N}{3}c(Y_\bullet^{j-1},\lambda) - \Delta_N^{\frac{1}{\alpha-1}}\rho_N^2\right) \end{pmatrix}$$

Let $\hat\theta_N$ be defined by

$$G^\star_{N,\alpha}(\hat\theta_N) = 0.$$

Setting

$$\begin{aligned} g_{1,\alpha}(\delta,y,x;\theta,\rho) &= \frac{\partial_\kappa b(x,\kappa)}{c_\rho(x,\lambda)}(y - \delta^{1-\frac{1}{\alpha}}b(x,\kappa)), \\ g_{2,\alpha}(\delta,y,x;\theta,\rho) &= \frac{\partial_\lambda c(x,\kappa)}{c_\rho(x,\lambda)^2}\left\{\frac{1}{2}(y - \delta^{1-\frac{1}{\alpha}}b(x,\kappa))^2 - \frac{\delta^{1-\frac{1}{3}}}{3}c(x,\lambda) - \delta^{\frac{1}{\alpha}}\rho^2\right\}, \end{aligned}$$

we have

$$\begin{aligned} g^{(1)}_{1,\alpha}(y,x;\theta,\rho) &= -\frac{\partial_\kappa b(x,\kappa)}{c_\rho(x,\lambda)}b(x,\kappa), \\ g^{(1)}_{2,\alpha}(y,x;\theta,\rho) &= -\frac{1}{3}c(x,\lambda)\frac{\partial_\lambda c(x,\lambda)}{c_\rho(x,\lambda)^2} - yb(x,\kappa) \quad \text{if } 1<\alpha<2, \\ &= -\frac{1}{3}c_\rho(x,\lambda)\frac{\partial_\lambda c(x,\lambda)}{c_\rho(x,\lambda)^2} - yb(x,\kappa) \quad \text{if } \alpha = 2, \\ g^{(\alpha)}_{1,\alpha}(y,x;\theta,\rho) &= 0, \\ g^{(\alpha)}_{2,\alpha}(y,x;\theta,\rho) &= -\rho^2\frac{\partial_\lambda c(x,\lambda)}{c_\rho(x,\lambda)^2} \end{aligned}$$

and $g^{(\alpha)}$ is only needed if $\alpha \in (\frac{3}{2}, 2]$. Otherwise, $\delta_N^{\frac{1}{\alpha}} = O(\delta_N^{2-\frac{2}{\alpha}})$.

The estimating function $G^\star_{N,\alpha}$ satisfies the *rate optimality* condition :

$$\partial_y g_{2,\alpha}(0,0,x;\theta,\rho) = 0$$

and the *Jacobsen condition* if $c_\rho = c$, *i.e.* if $\alpha \in (1,2)$ or if $\alpha = 2$ and **(B2)**. Hence, in this case, the following result holds.

Théorème 5.6 *Assume that $\alpha \in (1,2)$ and* **(B1)** *or* **(B2)**, *or $\alpha = 2$ and* **(B2)**. *Then the estimator $\hat\theta_N = (\kappa_N, \lambda_N)$ defined by $G^\star_{N,\alpha}(\hat\theta_N) = 0$ is consistent, and*

asymptotically Gaussian :

$$\begin{pmatrix} \sqrt{N\delta_N}(\hat{\kappa}_N - \kappa_0) \\ \sqrt{N\delta_N^{\frac{1}{\alpha}}}(\hat{\lambda}_N - \lambda_0) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \left\{\nu_0\left(\frac{(\partial_\kappa b(.,\kappa_0))^2}{c(.,\lambda_0)}\right)\right\}^{-1} & 0 \\ 0 & \frac{9}{4}\left\{\nu_0\left(\frac{(\partial_\lambda c(.,\lambda_0))^2}{c(.,\lambda_0)^2}\right)\right\}^{-1} \end{pmatrix}\right)$$

as $N \to \infty$*, with* $N\delta_N \to \infty$ *and* $N\delta_N^{2-\frac{1}{\alpha}} \to 0$*.*

In the case $\alpha = 2$ *and* $\rho_N = \rho$*, with* $c_\rho(x,\lambda) = c(x,\lambda) + 3\rho^2$*, the* second Jacobsen's condition *is not satisfied. The estimator* $\hat{\theta}_N$ *is still consistent, but we have*

$$\begin{pmatrix} \sqrt{N\delta_N}(\hat{\kappa}_N - \kappa_0) \\ \sqrt{N\delta_N^{\frac{1}{\alpha}}}(\hat{\lambda}_N - \lambda_0) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \frac{\nu_0\left(\frac{(\partial_\kappa b(.,\kappa_0))^2 c(.,\lambda_0)}{c_\rho(.,\lambda_0)^2}\right)}{\nu_0\left(\frac{(\partial_\kappa b(.,\kappa_0))^2}{c_\rho(.,\lambda_0)}\right)^2} & 0 \\ 0 & \frac{9}{4}\frac{\nu_0\left(\frac{(\partial_\lambda c(.,\lambda_0))^2(c^2+4c\rho^2+12\rho^4)}{c_\rho(.,\lambda_0)^4}\right)}{\nu_0\left(\frac{(\partial_\lambda c(.,\lambda_0))^2}{c_\rho(.,\lambda_0)^2}\right)^2} \end{pmatrix}\right)$$

as $N \to \infty$*, with* $N\delta_N \to \infty$ *and* $N\delta_N^{2-\frac{1}{\alpha}} \to 0$*.*

An other estimating function can be considered :

$$G^{\diamondsuit}_{N,\alpha}(\theta) = \frac{1}{k_N\Delta_N}\begin{pmatrix} \sum_{j=1}^{k_N-2} \frac{\partial_\kappa b(Y_\bullet^{j-1},\kappa)}{c(Y_\bullet^{j-1},\lambda)}\left(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1},\kappa)\right) \\ \sum_{j=1}^{k_N-2} \frac{\partial_\lambda c(Y_\bullet^{j-1},\lambda)}{c(Y_\bullet^{j-1},\lambda)^2}\left(\frac{1}{2}(Y_\bullet^{j+1} - Y_\bullet^j)^2 - \frac{\Delta_N}{3}c(Y_\bullet^{j-1},\lambda) - \Delta_N^{\frac{1}{\alpha-1}}\rho_N^2\right) \end{pmatrix}$$

but $G^{\diamondsuit}_{N,\alpha}$ is not a gradient : there is no function V such that $G^{\diamondsuit}_{N,\alpha} = \text{grad}\,(V)$. Indeed, setting

$$\begin{aligned} g^{\diamondsuit}_{1,\alpha}(\delta,y,x;\theta,\rho) &= \frac{\partial_\kappa b(x,\kappa)}{c(x,\lambda)}(y - \delta^{1-\frac{1}{\alpha}}b(x,\kappa)), \\ g^{\diamondsuit}_{2,\alpha}(\delta,y,x;\theta,\rho) &= \frac{\partial_\lambda c(x,\lambda)}{c(x,\lambda)^2}\left(\frac{y^2}{2} - \delta^{1-\frac{1}{\alpha}}\frac{c(x,\lambda)}{3}c(x,\lambda) - \delta^{\frac{1}{\alpha}}\rho^2\mathbf{1}_{\alpha\in(\frac{3}{2},2]}\right) \end{aligned}$$

we have $\partial_\lambda g^{\diamondsuit}_{1,\alpha} \neq \partial_\kappa g^{\diamondsuit}_{2,\alpha}$.

However, $g^{\diamondsuit}_{\alpha}$ satisfies the *second Jacobsen's condition.* Setting $\hat{\theta}^{\diamondsuit}_N$ as the solution of $G^{\diamondsuit}_{N,\alpha}(\theta) = 0$,

$$\begin{pmatrix} \sqrt{N\delta_N}(\hat{\kappa}_N - \kappa_0) \\ \sqrt{N\delta_N^{\frac{1}{\alpha}}}(\hat{\lambda}_N - \lambda_0) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \left\{\nu_0\left(\frac{(\partial_\kappa b(.,\kappa_0))^2}{c(.,\lambda_0)}\right)\right\}^{-1} & 0 \\ 0 & \frac{9}{4}\frac{\nu_0\left(\frac{(\partial_\lambda c(.,\lambda_0))^2(c^2+4c\rho^2+12\rho^4)}{c(.,\lambda_0)^4}\right)}{\nu_0\left(\frac{(\partial_\lambda c(.,\lambda_0))^2}{c(.,\lambda_0)^2}\right)^2} \end{pmatrix}\right)$$

as $N \to \infty$, with $N\delta_N \to \infty$ and $N\delta_N^{2-\frac{1}{\alpha}} \to 0$.

Hence, the asymptotic variance obtained for the estimation of κ_0 is the same that in the case of direct observations of the diffusion process.

5.4.1 The Ornstein-Uhlenbeck diffusion

Let (X_t) be the solution of

$$dX_t = \kappa X_t dt + \lambda dB_t,$$

with $\kappa < 0$, $\lambda > 0$ and X_0 deterministic or distributed with the stationary distribution of (X_t). Explicit estimators $\hat{\kappa}_N$ and $\hat{\lambda}_N$ are given in Chapter 4, and for $\alpha \in (1,2)$ or $\alpha = 2$ and $\rho_N \to 0$,

$$\begin{pmatrix} \sqrt{N\delta_N}(\hat{\kappa}_N - \kappa_0) \\ \sqrt{N\delta_N^{\frac{1}{\alpha}}}(\hat{\lambda}_N - \lambda_0) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 2|\kappa| & 0 \\ 0 & \frac{9}{4}\lambda^4 \end{pmatrix}\right)$$

as $N \to \infty$, with $N\delta_N \to \infty$ and $N\delta_N^{2-\frac{1}{\alpha}} \to 0$.

5.4.2 The hyperbolic diffusion

Let (X_t) be the solution of

$$dX_t = \kappa X_t dt + \lambda\sqrt{1+X_t^2}dB_t, \qquad X_0 = \eta \in \mathbb{R}, \tag{5.18}$$

where η is a random variable independent of (B_t), $\kappa < 0$ and $\lambda > 0$. In this case, the model is positive recurrent if $|\kappa| + \frac{\lambda^2}{2} > 0$, and in this case, its stationary distribution has density

$$\nu_0(x) \propto \frac{1}{(1+x^2)^{1+\frac{|\kappa|}{\lambda^2}}}.$$

If $X_0 = \eta$ has distribution $\nu_0(x)dx$, then, $\sqrt{1+\frac{2|\kappa|}{\lambda^2}}\eta$ has Student distribution. In this case,

$$\mathbb{E}_\theta(X_0) = 0 \text{ and } \mathbb{E}_\theta\left(\frac{X_0^2}{1+X_0^2}\right) = \frac{3+\frac{2|\kappa|}{\lambda^2}}{\left(\frac{2|\kappa|}{\lambda^2}+1\right)^{3/2}}.$$

Then

$$\begin{pmatrix} \sqrt{N\delta_N}(\hat{\kappa}_N - \kappa_0) \\ \sqrt{N\delta_N^{\frac{1}{\alpha}}}(\hat{\lambda}_N - \lambda_0) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \frac{\lambda^2\left(\frac{2|\kappa|}{\lambda^2}+1\right)^{3/2}}{3+\frac{2|\kappa|}{\lambda^2}} & 0 \\ 0 & \frac{9}{4}\lambda^4 \end{pmatrix}\right)$$

as $N \to \infty$, with $N\delta_N \to \infty$, $\delta_N \to 0$ and $N\delta_N^{2-\frac{1}{\alpha}} \to 0$, in the case $\alpha \in (1,2)$ or $\alpha = 2$ and $\rho_N \to 0$.

5.5 Concluding remarks

In this chapter, the asymptotic normality of the minimum constrast estimators of Chapter 4 is proved in the context of estimating functions, which correspond to the gradient of the contrast function. Two different rates of convergence in distribution are obtained for the drift parameter and the diffusion coefficient parameter. The drift parameter estimator $\hat{\kappa}_N$ is asymptotically Gaussian at rate $\sqrt{N\delta_N}$, which is the usual rate for directly observed diffusions, whereas the diffusion coefficient parameter estimator $\hat{\lambda}_N$ is asymptotically Gaussian at rate $\sqrt{N\delta_N^{\frac{1}{\alpha}}}$: this is slower than the rate of convergence $\sqrt{N}$ for directly observed diffusions, and depends on the local mean parameter α. Finally, the results in this Chapter show that even for non-Markovian observations, it is possible to introduce an estimating function, and obtain an estimator with good asymptical properties.

5.6 Auxiliary tools

In Chapter 4, several properties of $Y_\bullet^j$ and some results of convergence in probability for functionals of the blocks have been established. Recall that the random variables $\zeta_{j,N}$ and $\zeta'_{j+1,N}$ are defined in (4.8), and Lemma B.2 is used to establish convergence in probability results.

Proposition 5.1 *Under Assumptions* **(A1)-(A2)** *and* **(A5)**, *we have*

$$Y_\bullet^{j+1} - Y_\bullet^j = \Delta_N b(X_{j\Delta_N}) + \sigma(X_{j\Delta_N})(\zeta_{j+1,N} + \zeta'_{j+2,N}) + \tau_{j,N} + \rho_N(\varepsilon_\bullet^{j+1} - \varepsilon_\bullet^j)$$

where $\tau_{j,N}$ *is* $\mathcal{H}_{j+2}^N$ *mesurable, and there exists a constant* c *such that*

$$\begin{aligned}
|\mathbb{E}(\tau_{j,N}|\mathcal{G}_j^N)| &\leq c\Delta_N(\Delta_N(1+|X_{j\Delta_N}|^3) + \rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^4)}),\\
\mathbb{E}(\tau_{j,N}^2|\mathcal{G}_j^N) &\leq c\Delta_N(1+|X_{j\Delta_N}|^2 + \rho_N^2\mathbb{E}((\varepsilon_\bullet^j)^2))(\Delta_N(1+|X_{j\Delta_N}|^4) + \rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^4)}),\\
\mathbb{E}(\tau_{j,N}^4|\mathcal{G}_j^N) &\leq c(1+|X_{j\Delta_N}|^4 + \rho_N^4\mathbb{E}((\varepsilon_\bullet^j)^4))(\Delta_N^4(1+|X_{j\Delta_N}|^4) + \rho_N^4\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^8)}),\\
|\mathbb{E}(\tau_{j,N}\zeta_{j+1,N}|\mathcal{G}_j^N)| &\leq c\Delta_N(1+|X_{j\Delta_N}|^2 + \rho_N^2\mathbb{E}((\varepsilon_\bullet^j)^2))(\Delta_N(1+|X_{j\Delta_N}|^4) + \rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^4)}),\\
|\mathbb{E}(\tau_{j,N}\zeta'_{j+2,N}|\mathcal{G}_j^N)| &\leq c\Delta_N(1+|X_{j\Delta_N}|^2 + \rho_N^2\mathbb{E}((\varepsilon_\bullet^j)^2))(\Delta_N(1+|X_{j\Delta_N}|^4) + \rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^4)}).
\end{aligned}$$

5.7 Proofs

In the proofs, we use for sake of simplicity $\Delta_N = \delta_N^{1-\frac{1}{\alpha}}$.

Proof of Lemma 5.1. We have, with Taylor's formula,

$$\begin{aligned}
\mathbb{E}_\theta(g(\Delta_N, Y_\bullet^{j+1}-Y_\bullet^j, Y_\bullet^{j-1};\theta,\rho_N)|\mathcal{H}_j^N) &= g(0,0,Y_\bullet^{j-1};\theta,\rho_N)+\partial_y g(0,0,Y_\bullet^{j-1};\theta,\rho_N)\mathbb{E}_\theta(Y_\bullet^{j+1}-Y_\bullet^j|\mathcal{H}_j^N)\\
&= +\frac{1}{2}\partial^2_{y^2} g(0,0,Y_\bullet^{j-1},\theta,\rho_N)\mathbb{E}_\theta((Y_\bullet^{j+1}-Y_\bullet^j)^2|\mathcal{H}_j^N)\\
&\quad +\Delta_N g^{(1)}(0,Y_\bullet^{j+1};\theta,\rho_N)+\Delta_N^{\frac{1}{\alpha-1}} g^{(\alpha)}(0,Y_\bullet^{j-1};\theta,\rho_N)\\
&\quad +\Delta_N^2 R(\Delta_N, Y_\bullet^{j-1}, X_{j\Delta_N};\theta,\rho_N)\\
&= g(0,0,Y_\bullet^{j-1};\theta,\rho_N)+\partial_y g(0,0,Y_\bullet^{j-1};\theta,\rho_N)\Delta_N b(X_{j\Delta_N},\kappa)\\
&\quad +\frac{1}{2}\partial^2_{y^2} g(0,0,Y_\bullet^{j-1},\theta,\rho_N)(\Delta_N c(X_{j\Delta_N},\lambda)+2\Delta_N^{\frac{1}{\alpha-1}}\rho_N)\\
&\quad +\Delta_N g^{(1)}(0,Y_\bullet^{j-1};\theta,\rho_N)+\Delta_N^{\frac{1}{\alpha-1}} g^{(\frac{1}{\alpha-1})}(0,Y_\bullet^{j-1};\theta,\rho_N)\\
&\quad +\Delta_N^2 R(\Delta_N, Y_\bullet^{j-1}, X_{j\Delta_N};\theta,\rho_N).
\end{aligned}$$

Then, we replace $X_{j\Delta_N}$ by $Y_\bullet^{j-1}$ and obtain

$$\begin{aligned}
\mathbb{E}_\theta(g(\Delta_N, Y_\bullet^{j+1}-Y_\bullet^j, Y_\bullet^{j-1};\theta,\rho_N)|\mathcal{H}_j^N) &= g(0,0,Y_\bullet^{j-1};\theta,\rho_N)\\
&\quad +\Delta_N(\bar{L}_\theta(g(0,\theta,\rho_N))(0,Y_\bullet^{j-1})+g^{(1)}(0,Y_\bullet^{j-1};\theta,\rho_N))\\
&\quad +\Delta_N^{\frac{1}{\alpha-1}}(g^{(\frac{1}{\alpha-1})}(0,Y_\bullet^{j-1};\theta,\rho_N)+\rho_N^2\partial^2_{y^2} g(0,0,Y_\bullet^{j-1},\theta,\rho_N))\\
&\quad +\Delta_N^2 R(\Delta_N, Y_\bullet^{j-1}, X_{j\Delta_N};\theta,\rho_N)
\end{aligned}$$

where the terms

$$\Delta_N\partial_y g(0,0,Y_\bullet^{j-1};\theta,\rho_N)(b(X_{j\Delta_N},\kappa)-b(Y_\bullet^{j-1},\kappa))$$

and

$$\frac{\Delta_N}{2}\partial^2_{y^2} g(0,0,Y_\bullet^{j-1},\theta,\rho_N)(c(X_{j\Delta_N},\lambda)-c(Y_\bullet^{j-1},\lambda))$$

are $\mathcal{H}_j^N$ measurable and negligible : their conditional expectation given $\mathcal{H}_{j-1}^N$ is of order Δ_N^2. The proof ends with

$$\mathbb{E}_\theta(g(\Delta_N, Y_\bullet^{j+1}-Y_\bullet^j, Y_\bullet^{j-1};\theta,\rho_N)|\mathcal{H}_j^N)=O(\Delta_N^2)$$

and the identification of the terms. □

Proof of Lemma 5.2. Tools for the proof are closed to those of Chapter 4. Set $A_{j,N}(\theta,\theta_0)=\mathbb{E}_{\theta_0}(g(\Delta_N, Y_\bullet^{j+1}-Y_\bullet^j, Y_\bullet^{j-1};\theta,\rho_N)|\mathcal{H}_j^N)$. We have

$$\begin{aligned}
A_{j,N}(\theta,\theta_0) &= \Delta_N(g^{(1)}(0,Y_\bullet^{j-1};\theta,\rho_N)+\bar{L}_{\theta_0}(g(0,\theta))(0,Y_\bullet^{j-1},\rho_N)\\
&\quad +\Delta_N^{\frac{1}{\alpha-1}}(g^{(\alpha)}(0,Y_\bullet^{j-1};\theta,\rho_N)+\rho_N^2\partial^2_{y^2} g(0,0,Y_\bullet^{j-1},\theta,\rho_N))\\
&\quad +\Delta_N^2 R(\Delta_N, Y_\bullet^{j-1}, X_{j\Delta_N};\theta,\rho_N)\\
&= \Delta_N(\bar{L}_{\theta_0}((g(0,\theta,\rho_N))(0,Y_\bullet^{j-1})-\bar{L}_\theta(g(0,\theta,\rho_N))(0,Y_\bullet^{j-1})+\Delta_N^2 R(\Delta_N, Y_\bullet^{j-1};\theta,\rho_N).
\end{aligned}$$

Splitting the sum $\frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2} A_{j,N}(\theta,\theta_0)$ in three parts, we conclude for the pointwise convergence with Lemma B.2 in the appendix of Chapter 4. For the uniform convergence in θ, recall that Θ is compact and the functionals $\bar{I}_N(f(.,\theta))$ and $\bar{Q}_N(f(.\theta))$ studied in Chapter 4 converge in probability uniformly in θ.

Thus, with Taylor's formula, as g admits a developpment in powers of Δ,

$$\begin{aligned} g(\Delta, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta, \rho_N) &= g(0,0,Y_\bullet^{j-1};\theta,\rho_N) \\ &\quad +(Y_\bullet^{j+1} - Y_\bullet^j)\partial_y g(0,0,Y_\bullet^{j-1},\theta,\rho_N) \\ &\quad +\frac{1}{2}(Y_\bullet^{j+1} - Y_\bullet^j)^2 \partial^2_{y^2} g(0,0,Y_\bullet^{j-1};\theta,\rho_N) \\ &\quad +\Delta_N g^{(1)}(0,Y_\bullet^{j-1};\theta,\rho_N) \\ &\quad +\Delta_N^{\frac{1}{\alpha-1}} g^{(\frac{1}{\alpha-1})}(0,Y_\bullet^{j-1};\theta,\rho_N) \\ &\quad +\Delta_N^2 R(\Delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1};\theta,\rho_N). \end{aligned}$$

Hence, we have, with $g^{(1)}(0,Y_\bullet^{j-1};\theta,\rho_N) = -\bar{L}_\theta(g(0,\theta,\rho_N))$, the convergence of

$$\frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2}(\partial_y g(0,0,Y_\bullet^{j-1};\theta,\rho_N)(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1},\kappa_0))$$
$$+\partial^2_{y^2} g(0,0,Y_\bullet^{j-1};\theta,\rho_N)(\frac{1}{2}(Y_\bullet^{j+1} - Y_\bullet^j)^2 - \frac{\Delta_N}{3}c(Y_\bullet^{j-1},\lambda_0) - \Delta_N^{\frac{1}{\alpha-1}} g^{(\frac{1}{\alpha-1})}(0,Y_\bullet^{j-1};\theta)))$$

in probability uniformly in $\theta \in \Theta$.

□

Proof of Theorem 5.1. In this proof, we set $b(.) = b(.,\kappa_0)$ and $\sigma(.) = \sigma(.,\lambda_0)$. Considering a function f satisfying Condition **(C1)** and the random variables defined in Lemma 4.2, we have, for $1 \le j \le k_N - 2$,

$$\begin{aligned} f(Y_\bullet^{j-1})(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1})) &= f(Y_\bullet^{j-1})(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(X_{j\Delta_N})) \\ &\quad +\Delta_N(b(X_{j\Delta_N}) - b(Y_\bullet^{j-1})). \end{aligned}$$

With Proposition 5.1,

$$\begin{aligned} f(Y_\bullet^{j-1})(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(X_{j\Delta_N})) &= f(Y_\bullet^{j-1})\sigma(X_{j\Delta_N})(\zeta_{j+1,N} + \zeta'_{j+2,N}) \\ &\quad +f(Y_\bullet^{j-1})\rho_N(\varepsilon_\bullet^{j+1} - \varepsilon_\bullet^j) \\ &\quad +f(Y_\bullet^{j-1})\tau_{j,N}. \end{aligned}$$

We study

$$\sqrt{k_N\Delta_N}\bar{I}_N(f) = \sum_{\ell=1}^4 \bar{R}_N^{(\ell)}$$

with

$$\begin{aligned} \bar{R}_N^{(1)} &= \frac{1}{\sqrt{k_N\Delta_N}}\sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1})\sigma(X_{j\Delta_N})(\zeta_{j+1,N} + \zeta'_{j+2,N}), \\ \bar{R}_N^{(2)} &= \frac{1}{\sqrt{k_N\Delta_N}}\sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1})\rho_N(\varepsilon_\bullet^{j+1} - \varepsilon_\bullet^j), \\ \bar{R}_N^{(3)} &= \frac{1}{\sqrt{k_N\Delta_N}}\sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1})\tau_{j,N}, \\ \bar{R}_n^{(4)} &= \frac{1}{\sqrt{k_N\Delta_N}}\sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1})\Delta_N(b(X_{j\Delta_N}) - b(Y_\bullet^{j-1})). \end{aligned}$$

Let $r_{j,N}^{(1)}$ be defined by

$$r_{j,N}^{(1)} = f(Y_\bullet^{j-1})\sigma(X_{j\Delta_N})\zeta_{j+1,N} + f(Y_\bullet^{j-2})\sigma(X_{(j-1)\Delta_N})\zeta'_{j+1,N},$$

where the terms have been rearranged, to have $r_{j,N}^{(1)}$ be $\mathcal{H}_{j+1}^N$ measurable.

First, we study

$$\bar{R}_N^{(1)} = \frac{1}{\sqrt{k_N\Delta_N}} \sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1})\sigma(X_{j\Delta_N})(\zeta_{j+1,N} + \zeta'_{j+2,N})$$

and we have

$$\bar{R}_N^{(1)} = \frac{1}{\sqrt{k_N\Delta_N}} \sum_{j=2}^{k_N-2} r_{j,N}^{(1)} + o_P(1).$$

To prove the convergence in distribution of $\bar{R}_N^{(1)}$, we use a Central Limit Theorem for martingale arrays (Theorem 3.2 in Hall and Heyde (1980)). We have the conditional centering $\mathbb{E}(r_{j,N}^{(1)}|\mathcal{H}_j^N) = 0$. We compute the conditional variance :

$$\begin{aligned}\mathbb{E}((r_{j,N}^{(1)})^2|\mathcal{H}_j^N) &= \Delta_N \left(f(Y_\bullet^{j-1})^2\sigma(X_{j\Delta_N})^2\Delta_N \left(\frac{1}{3} + \frac{1}{2p_N} + \frac{1}{6p_N^2}\right)\right. \\ &\quad + f(Y_\bullet^{j-2})^2\sigma(X_{(j-1)\Delta_N})^2 \left(\frac{1}{3} - \frac{1}{2p_N} + \frac{1}{6p_N^2}\right) \\ &\quad \left. + f(Y_\bullet^{j-2})\sigma(X_{(j-1)\Delta_N}) f(Y_\bullet^{j-1})\sigma(X_{j\Delta_N})\frac{1}{3}\left(1 - \frac{1}{p_N^2}\right)\right).\end{aligned}$$

Hence

$$\frac{1}{k_N\Delta_N} \sum_{j=2}^{k_N-1} \mathbb{E}((r_{j,N}^{(1)})^2|\mathcal{H}_j^N) \xrightarrow{\mathbb{P}} \nu_0(f^2\sigma^2).$$

We easily bound $\mathbb{E}((r_{j,N}^{(1)})^4|\mathcal{H}_j^N)$ with Proposition 5.1 and derive

$$\frac{1}{k_N^2\Delta_N^2} \sum_{j=2}^{k_N-1} \mathbb{E}((r_{j,N}^{(1)})^4|\mathcal{H}_j^N) \xrightarrow{\mathbb{P}} 0.$$

Then,

$$\bar{R}_N^{(1)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu_0(f^2\sigma^2)).$$

We now prove that $\bar{R}_N^{(\ell)} = o_P(1)$ for $\ell = 2, 3$ and apply Slutsky's lemma to conclude. We first deal with $\bar{R}_N^{(3)}$ and we have

$$\left|\mathbb{E}(f(Y_\bullet^{j-1})\tau_{j,N}|\mathcal{H}_j^N)\right| \leq c(1 + |Y_\bullet^{j-1}|)\Delta_N(\Delta_N(1 + |X_{j\Delta_N}|^3) + \rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^4)}).$$

Then we use the same arguments as in Lemma 5.2 for the pointwise convergence : as $\tau_{j,N}$ is $\mathcal{H}^N_{j+2}$ measurable, we split $\bar{R}^{(3)}_N$ in three sums, in order to prove that it converges in probability to 0 with Lemma B.2. With the condition $k_N\Delta_N^3 \to 0$,

$$\frac{1}{k_N\Delta_N}\sum_{2\leq 3j\leq k_N-2}\sqrt{k_N\Delta_N}\mathbb{E}(f(Y_\bullet^{3j-1})\tau_{3j,N}|\mathcal{H}^N_{3j}) \xrightarrow{\mathbb{P}} 0.$$

We get

$$\begin{aligned}\mathbb{E}(f(Y_\bullet^{j-1})^2\tau^2_{j,N}|\mathcal{H}^N_j) &\leq c(1+|Y^j_\bullet|^2)\Delta_N(1+|X_{j\Delta_N}|^2+\rho_N^2\mathbb{E}((\varepsilon^j_\bullet)^2))\\ &\quad\times(\Delta_N(1+|X_{j\Delta_N}|^4)+\rho_N^2\sqrt{\mathbb{E}((\varepsilon^j_\bullet)^4)}).\end{aligned}$$

Hence we get

$$\frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2}\mathbb{E}(f(Y_\bullet^{3j-1})^2\tau^2_{3j,N}|\mathcal{H}^N_{3j}) \longrightarrow 0$$

in probability.

We consider now $\bar{R}^{(2)}_N$. We define the $\mathcal{H}^N_{j+1}$ measurable random variable

$$r^{(2)}_{j,N} = (f(Y_\bullet^{j-2}) - f(Y_\bullet^{j-1}))\rho\varepsilon^j_\bullet$$

and we have

$$\bar{R}^{(2)}_N = \frac{1}{\sqrt{k_N\Delta_N}}\sum_{j=2}^{k_N-2} r^{(2)}_{j,N} + o_P(1).$$

To use Lemma B.2, we compute

$$\mathbb{E}(r^{(2)}_{j,N}|\mathcal{H}^N_j) = 0$$

and

$$\mathbb{E}((r^{(2)}_{j,N})^2|\mathcal{H}^N_j) = (f(Y_\bullet^{j-2}) - f(Y_\bullet^{j-1}))^2\rho_N^2\mathbb{E}((\varepsilon^j_\bullet)^2).$$

As $\Delta_N = p_N^{1-\alpha}$ it comes

$$\frac{1}{k_N}\sum_{j=2}^{k_N-2}\frac{\rho_N^2}{p_N^{2-\alpha}}(f(Y_\bullet^{j-2}) - f(Y_\bullet^{j-1}))^2$$

and this quantity vanishes in probability, for $\alpha \in (1,2]$, with Assumption **(B1)** or **(B2)**.

Finally,

$$\bar{R}^{(4)}_n = \frac{1}{\sqrt{k_N\Delta_N}}\sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1})\Delta_N(b(X_{j\Delta_N}) - b(Y_\bullet^{j-1})).$$

We have, for $2 \leq j \leq k_N - 2$,

$$\begin{aligned} f(Y_\bullet^{j-1})(b(X_{j\Delta_N}) - b(Y_\bullet^{j-1})) &= f(Y_\bullet^{j-2})(b(X_{j\Delta_N}) - b(Y_\bullet^{j-1})) \\ &\quad + f(Y_\bullet^{j-2})(b(X_{(j-1)\Delta_N}) - b(Y_\bullet^{j-1})) \\ &\quad + (f(Y_\bullet^{j-1}) - f(Y_\bullet^{j-2})(b(X_{j\Delta_N}) - b(Y_\bullet^{j-1}). \end{aligned}$$

With Proposition 4.2 of Chapter 4 and the Cauchy Schwarz inequality, it comes

$$\begin{aligned} &|\mathbb{E}(f(Y_\bullet^{j-2})(b(X_{j\Delta_N}) - b(Y_\bullet^{j-1}))|\mathcal{H}_{j-1}^N)| \leq \\ &\quad cf(Y_\bullet^{j-2})(\Delta_N(1 + X_{(j-1)\Delta_N}^2) + \rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet^{j-1})^4)}), \\ &|\mathbb{E}(f(Y_\bullet^{j-2})(b(X_{(j-1)\Delta_N}) - b(Y_\bullet^{j-1}))|\mathcal{H}_{j-1}^N)| \leq \\ &\quad cf(Y_\bullet^{j-2})(\Delta_N(1 + X_{(j-1)\Delta_N}^2) + \rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet^{j-1})^4)}), \\ &|\mathbb{E}((f(Y_\bullet^{j-1}) - f(Y_\bullet^{j-2})(b(X_{j\Delta_N}) - b(Y_\bullet^{j-1})|\mathcal{H}_{j-1}^N)| \leq \\ &\quad c(1 + X_{j\Delta_N}^2 + \rho_N^2\mathbb{E}((\varepsilon_\bullet^{j-1})^2))(\Delta_N(1 + X_{j\Delta_N}^4) + \rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet^{j-1})^4)}). \end{aligned}$$

With $k_N\Delta_N^3 = N\delta_N^{3-\frac{2}{\alpha}} \to 0$, $\bar{R}_N^{(4)} = o_P(1)$. □

Proof of Theorem 5.2. We use the same notations as in Theorem 4.2.We have, with the notations of Proposition 5.1,

$$\begin{aligned} (Y_\bullet^{j+1} - Y_\bullet^j)^2 &= (\Delta_N b(X_{j\Delta_N}) + \tau_{j,N})^2 + \sigma(X_{j\Delta_N})^2(\zeta_{j+1,N} + \zeta'_{j+2,N})^2 \\ &\quad + \rho_N^2(\varepsilon_\bullet^{j+1} - \varepsilon_\bullet^j)^2 + 2\rho_N(\Delta_N b(X_{j\Delta_N}) + \tau_{j,N})(\varepsilon_\bullet^{j+1} - \varepsilon_\bullet^j) \\ &\quad + 2(\Delta_N b(X_{j\Delta_N}) + \tau_{j,N})\sigma(X_{j\Delta_N})(\zeta_{j+1,N} + \zeta'_{j+2,N}) \\ &\quad + 2\rho_N\sigma(X_{j\Delta_N})(\zeta_{j+1,N} + \zeta'_{j+2,N})(\varepsilon_\bullet^{j+1} - \varepsilon_\bullet^j). \end{aligned}$$

We set $U_N(f) = \sqrt{k_N}(\bar{Q}_N(f) - \bar{\nu}_N(f(\frac{2}{3}\sigma^2 + 2\rho_N^2\Delta_N^{\frac{2-\alpha}{\alpha-1}})))$ and we define

$$\begin{aligned} u_{j,N}^{(1)} &= f(Y_\bullet^{j-1})\sigma(X_{j\Delta_N})^2(\frac{1}{\Delta_N}(\zeta_{j+1,N} + \zeta'_{j+2,N})^2 - \frac{2}{3}) \\ u_{j,N}^{(2)} &= \frac{2}{3}f(Y_\bullet^{j-1})(\sigma(X_{j\Delta_N})^2 - \sigma(Y_\bullet^j)^2) \\ u_{j,N}^{(3)} &= f(Y_\bullet^{j-1})\rho_N^2(\frac{1}{\Delta_N}(\varepsilon_\bullet^{j+1} - \varepsilon_\bullet^j)^2 - 2\Delta_N^{\frac{2-\alpha}{\alpha-1}}) \\ u_{j,N}^{(4)} &= 2f(Y_\bullet^{j-1})\sigma(X_{j\Delta_N})\rho_N\frac{1}{\Delta_N}(\zeta_{j+1,N} + \zeta'_{j+2,N})(\varepsilon_\bullet^{j+1} - \varepsilon_\bullet^j) \\ u_{j,N}^{(5)} &= f(Y_\bullet^{j-1})(\Delta_N b(X_{j\Delta_N}) + \tau_{j,N})^2 \\ u_{j,N}^{(6)} &= 2f(Y_\bullet^{j-1})(\Delta_N b(X_{j\Delta_N}) + \tau_{j,N})\sigma(X_{j\Delta_N})(\zeta_{j+1,N} + \zeta'_{j+2,N}) \\ u_{j,N}^{(7)} &= 2f(Y_\bullet^{j-1})\rho_N(\Delta_N b(X_{j\Delta_N}) + \tau_{j,N})(\varepsilon_\bullet^{j+1} - \varepsilon_\bullet^j) \end{aligned}$$

so that

$$U_N(f) = \frac{1}{\sqrt{k_N}}\sum_{j=1}^{k_N-2}\sum_{\ell=1}^{7} u_{j,N}^{(\ell)}.$$

Recall that $\zeta_{j+1,N}$ is $\mathcal{H}^N_{j+1}$ measurable and $\zeta'_{j+2,N}$ is $\mathcal{H}^N_{j+2,N}$ measurable, so we have to reorder some terms to use a Central Limit Theorem for martingale array. Then,

$$U_N^{(1)} = \frac{1}{\sqrt{k_N}} \sum_{j=1}^{k_N-2} u_{j,N}^{(1)} = \frac{1}{\sqrt{k_N}} \sum_{j=2}^{k_N-2} s_{j,N}^{(1)} + \frac{1}{\sqrt{k_N}} \sum_{j=2}^{k_N-2} \tilde{s}_{j,N}^{(1)} + o_P(1)$$

with

$$\begin{aligned}
s_{j,N}^{(1)} &= f(Y_\bullet^{j-1})\sigma(X_{j\Delta_N})^2 \left(\frac{\zeta_{j+1,N}^2}{\Delta_N} - m_N \right) \\
&\quad + f(Y_\bullet^{j-2})\sigma(X_{(j-1)\Delta_N})^2 \left(\frac{{\zeta'}_{j+1,N}^2}{\Delta_N} - {m'}_N \right) \\
&\quad + 2f(Y_\bullet^{j-2})\sigma(X_{(j-1)\Delta_N})^2 \frac{\zeta_{j,N}{\zeta'}_{j+1,N}}{\Delta_N}, \\
\tilde{s}_{j,N}^{(1)} &= f(Y_\bullet^{j-1})\sigma(X_{j\Delta_N})^2 \left(\frac{1}{2p_N} + \frac{1}{6p_N^2} \right) \\
&\quad + f(Y_\bullet^{j-2})\sigma(X_{(j-1)\Delta_N})^2 \left(-\frac{1}{2p_N} + \frac{1}{6p_N^2} \right)
\end{aligned}$$

where we set $m_N = \frac{1}{3} + \frac{1}{2p_N} + \frac{1}{6p_N^2}$, $m'_N = \frac{1}{3} - \frac{1}{2p_N} + \frac{1}{6p_N^2}$ and $\chi_N = \frac{1}{6}(1 - \frac{1}{p_N^2})$ such that $\zeta_{j+1,N} \sim \mathcal{N}(0, m_N\Delta_N)$, $\zeta'_{j+1,N} \sim \mathcal{N}(0, {m'}_N\Delta_N)$, and Cov $(\zeta_{j+1,N}, \zeta'_{j+1,N}) = \chi_N\Delta_N$.

As $p_N^{-1} = \Delta_N^{\frac{1}{\alpha-1}} \leq \Delta_N$, with $k_N\Delta_N^2 = N\delta_N^{2-\frac{1}{\alpha}} \to 0$, we deduce

$$\frac{1}{\sqrt{k_N}} \sum_{j=2}^{k_N-2} \tilde{s}_{j,N}^{(1)} = o_P(1).$$

We have the conditional centering condition : $\mathbb{E}(s_{j,N}^{(1)} | \mathcal{H}_j^N) = 0$. We compute

$$\begin{aligned}
(s_{j,N}^{(1)})^2 &= f(Y_\bullet^{j-1})^2\sigma(X_{j\Delta_N})^4 \left(\frac{\zeta_{j+1,N}^2}{\Delta_N} - m_N \right)^2 \\
&\quad + f(Y_\bullet^{j-2})^2\sigma(X_{(j-1)\Delta_N})^4 \left(\frac{{\zeta'}_{j+1,N}^2}{\Delta_N} - m'_N \right)^2 \\
&\quad + 4f(Y_\bullet^{j-2})^2\sigma(X_{(j-1)\Delta_N})^4 \frac{\zeta_{j,N}^2 {\zeta'}_{j+1,N}^2}{\Delta_N^2} \\
&\quad + 2f(Y_\bullet^{j-2})\sigma(X_{(j-1)\Delta_N})^2 f(Y_\bullet^{j-1})\sigma(X_{j\Delta_N})^2 \left(\frac{\zeta_{j+1,N}^2}{\Delta_N} - m_N \right) \left(\frac{{\zeta'}_{j+1,N}^2}{\Delta_N} - m'_N \right) \\
&\quad + 4f(Y_\bullet^{j-2})\sigma(X_{(j-1)\Delta_N})^2 f(Y_\bullet^{j-1})\sigma(X_{j\Delta_N})^2 \left(\frac{\zeta_{j+1,N}^2}{\Delta_N} - m_N \right) \frac{\zeta_{j,N}{\zeta'}_{j+1,N}}{\Delta_N} \\
&\quad + 4f(Y_\bullet^{j-2})^2\sigma(X_{(j-1)\Delta_N})^4 \left(\frac{{\zeta'}_{j+1,N}^2}{\Delta_N} - m'_N \right) \frac{\zeta_{j,N}{\zeta'}_{j+1,N}}{\Delta_N}
\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}((s_{j,N}^{(1)})^2|\mathcal{H}_j^N) &= 2f(Y_\bullet^{j-1})^2\sigma(X_{j\Delta_N})^4 m_N^2 \\ &\quad +2f(Y_\bullet^{j-2})^2\sigma(X_{(j-1)\Delta_N})^4 {m'_N}^2 \\ &\quad +4f(Y_\bullet^{j-2})^2\sigma(X_{(j-1)\Delta_N})^4\frac{\zeta_{j,N}^2}{\Delta_N}m'_N \\ &\quad +4f(Y_\bullet^{j-2})\sigma(X_{(j-1)\Delta_N})^2 f(Y_\bullet^{j-1})\sigma(X_{j\Delta_N})^2\chi_N^2.\end{aligned}$$

With Lemma B.2 and (B.1) we conclude :

$$\frac{1}{k_N}\sum_{j=2}^{k_N-2}\mathbb{E}((s_{j,N}^{(1)})^2|\mathcal{H}_j^N) \longrightarrow \nu_0(f^2\sigma^4)$$

in probability. We easily bound $\mathbb{E}((s_{j,N}^{(1)})^4|\mathcal{H}_j^N)$ and we derive

$$\frac{1}{k_N^2}\sum_{j=2}^{k_N-2}\mathbb{E}((s_{j,N}^{(1)})^4|\mathcal{H}_j^N) \longrightarrow 0$$

in probability. Then, we can apply Theorem 3.2 in Hall and Heyde (1980) and obtain

$$U_N^{(1)} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \nu_0(f^2\sigma^4)\right).$$

We consider $U_N^{(2)} = \frac{1}{\sqrt{k_N}}\sum_{j=1}^{k_N-2} u_{j,N}^{(2)}$. With Proposition 4.2, we have

$$|\mathbb{E}(u_{j,N}^{(2)}|\mathcal{H}_j^N)| \leq c|f(Y_\bullet^{j-1})|(\Delta_N(1+X_{j\Delta_N}^2) + \rho_N^2\mathbb{E}((\varepsilon_\bullet^j)^2))$$

and with $k_N\Delta_N^2 = N\delta_N^{2-\frac{1}{\alpha}}$, $U_N^{(2)} = o_P(1)$.

Now let us consider

$$U_N^{(3)} = \frac{1}{\sqrt{k_N}}\sum_{j=1}^{k_N-2} u_{j,N}^{(3)} = \frac{1}{\sqrt{k_N}}\sum_{j=2}^{k_N-2} s_{j,N}^{(3)} + o_P(1)$$

where

$$s_{j,N}^{(3)} = (f(Y_\bullet^{j-2}) + f(Y_\bullet^{j-1}))\rho_N^2\left(\frac{(\varepsilon_\bullet^j)^2}{\Delta_N} - \Delta_N^{\frac{2-\alpha}{\alpha-1}}\right) - 2f(Y_\bullet^{j-2})\rho_N^2\frac{\varepsilon_\bullet^{j-1}\varepsilon_\bullet^j}{\Delta_N}$$

is $\mathcal{H}_{j+1}^N$ measurable.

We have $\mathbb{E}(s_{j,N}^{(3)}|\mathcal{H}_j^N) = 0$ and

$$\mathbb{E}((s_{j,N}^{(3)})^2|\mathcal{H}_j^N) = 2(f(Y_\bullet^{j-2}) + f(Y_\bullet^{j-1}))^2\frac{\rho_N^4}{\Delta_N^2 p_N^2} + 4f(Y_\bullet^{j-1})^2\rho_N^4\frac{(\varepsilon_\bullet^{j-1})^2}{\Delta_N^2 p_N}.$$

Then, we have to examine two cases :

1. If $\alpha \in (1,2)$ or $\rho_N \to 0$ and $\alpha = 2$, we derive from the previous computation

$$\frac{1}{k_N}\sum_{j=2}^{k_N-2} \mathbb{E}((s_{j,N}^{(3)})^2|\mathcal{H}_j^N) = o_P(1),$$

2. If $\alpha = 2$ and $\rho_N = \rho$,

$$\frac{1}{k_N}\sum_{j=2}^{k_N-2} \mathbb{E}((s_{j,N}^{(3)})^2|\mathcal{H}_j^N) = 12\nu_0(f^2\rho^4) + o_P(1).$$

In the case 1, we conclude with Lemma B.2 that $U_N^{(3)} = o_P(1)$, but in the case 2, we use a Central Limit Theorem for martingale array : using that

$$\frac{1}{k_N^2}\sum_{j=2}^{k_N-2} \mathbb{E}((s_{j,N}^{(3)})^4|\mathcal{H}_j^N) = o_P(1),$$

we derive

$$U_N^{(3)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 12\nu_0(f^2\rho^4)).$$

We also notice that, in this case, $\mathbb{E}(s_{j,N}^{(1)}s_{j,N}^{(3)}|\mathcal{H}_j^N) = 0$, and then $U_N^{(1)}$ and $U_N^{(3)}$ are asymptotically uncorrelated.

We deal with

$$U_N^{(4)} = \frac{1}{\sqrt{k_N}}\sum_{j=1}^{k_N-2} u_{j,N}^{(4)} = \frac{1}{\sqrt{k_N}}\sum_{j=2}^{k_N-2} s_{j,N}^{(4)} + o_P(1)$$

where

$$\begin{aligned} s_{j,N}^{(4)} &= 2f(Y_\bullet^{j-2})\rho_N\sigma(X_{(j-1)\Delta_N})\frac{1}{\Delta_N}\left(\zeta_{j,N}\varepsilon_\bullet^j + \zeta'_{j+1,N}\varepsilon_\bullet^j - \zeta'_{j+1,N}\varepsilon_\bullet^{j-1}\right) \\ &\quad -2f(Y_\bullet^{j-1})\rho_N\sigma(X_{j\Delta_N})\frac{\varepsilon_\bullet^j\zeta_{j+1,N}}{\Delta_N}. \end{aligned}$$

and $r_{j,N}^{(4)}$ is $\mathcal{H}_{j+1}^N$ measurable, with $\mathbb{E}(r_{j,N}^{(4)}|\mathcal{H}_j^N) = 0$.

Let us compute

$$\begin{aligned} \mathbb{E}\left((s_{j,N}^{(4)})^2\Big|\mathcal{H}_j^N\right) &= 4f(Y_\bullet^{j-2})^2\rho_N^2\sigma(X_{(j-1)\Delta_N})^2\left(\frac{\zeta_{j,N}^2}{\Delta_N} + m'_N + \frac{(\varepsilon_\bullet^{j-1})^2}{\Delta_N}m'_N\right) \\ &\quad +4f(Y_\bullet^{j-1})^2\rho_N^2\sigma(X_{j\Delta_N})^2 m_N \\ &\quad -8f(Y_\bullet^{j-1})\sigma(X_{j\Delta_N})f(Y_\bullet^{j-2})\sigma(X_{(j-1)\Delta_N})\rho_N^2\chi_N. \end{aligned}$$

Hence we have to consider the same two situations :

1. If $\alpha \in (1,2)$ or $\rho_N \to 0$ and $\alpha = 2$, it comes

$$\frac{1}{k_N}\sum_{j=2}^{k_N-2} \mathbb{E}\left((s_{j,N}^{(4)})^2 \Big| \mathcal{H}_j^N \right) = o_P(1).$$

2. If $\alpha = 2$ and $\rho_N = \rho$, it comes

$$\frac{1}{k_N}\sum_{j=2}^{k_N-2} \mathbb{E}\left((s_{j,N}^{(4)})^2 \Big| \mathcal{H}_j^N \right) = 4\nu_0(f^2\rho^2\sigma^2) + o_P(1).$$

Besides, in case 1, using Lemma B.2, $U_N^{(4)} = o_P(1)$, whereas in case 2, we verify that

$$\frac{1}{k_N^2}\sum_{j=2}^{k_N-2} \mathbb{E}\left((s_{j,N}^{(4)})^4 \Big| \mathcal{H}_j^N \right) = o_P(1)$$

and deduce from a Central Limit Theorem for martingale array

$$U_N^{(4)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\nu_0(f^2\rho^2\sigma^2).$$

Finally, we have

$$\mathbb{E}(s_{j,N}^{(1)} s_{j,N}^{(4)} | \mathcal{H}_j^N) = 0, \quad \mathbb{E}(s_{j,N}^{(2)} s_{j,N}^{(4)} | \mathcal{H}_j^N) = 0$$

and $U_N^{(1)}, U_N^{(3)}, U_N^{(4)}$ are asymptotically uncorrelated.

To finish the proof, let $U_N^{(\ell)} = \frac{1}{\sqrt{k_N}}\sum_{j=1}^{k_N-2} u_{j,N}^{(\ell)}$ for $\ell = 5,6,7$. With Propositions 4.2 and 4.3, we have, with the Cauchy Schwarz inequality,

$$\begin{aligned}
\mathbb{E}(|u_{j,N}^{(5)}||\mathcal{H}_j^N) &\leq c|f(Y_\bullet^{j-1})|\Delta_N(1 + X_{j\Delta_N}^2 + \rho_N^2\mathbb{E}((\varepsilon_\bullet^j)^2)) \\
&\quad \times(\Delta_N(1 + X_{j\Delta_N}^4 + \rho_N^2\sqrt{\mathbb{E}((\varepsilon_\bullet)^4)}), \\
\mathbb{E}(|u_{j,N}^{(6)}| + |u_{j,N}^{(7)}||\mathcal{H}_j^N) &\leq c|f(Y_\bullet^{j-1})|\Delta_N(1 + |X_{j\Delta_N}|^2 + \rho_N\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^2)}) \\
&\quad \times(\sqrt{\Delta_N}(1 + X_{j\Delta_N}^2 + \rho_N\mathbb{E}((\varepsilon_\bullet)^4)^{\frac{1}{4}}).
\end{aligned}$$

With $k_N\Delta_N^2 \to 0$, we have $U_N^{(\ell)} = o_P(1)$ for $\ell = 5,6,7$.

□

Proof of Theorem 5.3. As we deal with martingale arrays, with the notations of Theorems 5.1 and 5.2, it is sufficient to remark that

$$\frac{1}{k_N}\sum_{j=1}^{k_N-2} \mathbb{E}(r_{j,N}^{(1)} s_{j,N}^{(1)} | \mathcal{H}_j^N) = o_P(1).$$

□

Proof of Corollary 5.1. As $f_N(x) - f(x) = v_N(1 + |x|)$, we have

$$\begin{aligned} |\bar{\nu}_N(f_N) - \bar{\nu}_N(f)| &\leq \frac{1}{k_N} \sum_{j=0}^{k_N-1} |f_N(Y_\bullet^j) - f(Y_\bullet^j)| \\ &\leq v_N \times \frac{1}{k_N} \sum_{j=0}^{k_N-1} (1 + |Y_\bullet^j|) \end{aligned}$$

and the last quantity tends to 0 in probability. It is also sufficient to consider the case $f = 0$.

Then, considering the proof of Theorem 5.1, and keeping the same notations for $\bar{R}_N^{(\ell)}, \ell = 1 \dots 4$, we consider

$$\bar{R}_N^{(1)} = \frac{1}{\sqrt{k_N \Delta_N}} \sum_{j=1}^{k_N-2} f_N(Y_\bullet^{j-1}) \sigma(X_{j\Delta_N})(\zeta_{j+1,N} + \zeta'_{j+2,N}).$$

Splitting the sum into three parts, using Lemma B.2 and

$$\mathbb{E}(f(Y_\bullet^{3j-1})^2 \sigma(X_{j\Delta_N})^2 (\zeta_{j+1,N} + \zeta'_{j+2,N})^2 | \mathcal{H}_{3j}^N) \leq c v_N (1 + |Y_\bullet^{j-1}|^2) \sigma(X_{j\Delta_N})^2 \Delta_N,$$

$\bar{R}_N^{(1)}$ tends to 0 in probability. The same argument is used for the other terms in $\sqrt{k_N \Delta_N} \bar{I}_N(f_N)$ and $\sqrt{k_N} \bar{Q}_N(f_N)$.

□

Proof of Theorem 5.4. Consider the estimating function

$$G_N(\theta) = \frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} g(\delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta, \rho_N).$$

By Lemma 5.2, we have $G_N(\theta_0) \to 0$ and $\partial_\theta G_N(\theta) \to \phi(\theta, \theta_0, \rho_\infty)$ in probability, under $\mathbb{P}_{\theta_0}$, uniformly in $\theta \in \Theta$. As $\phi(\theta_0, \theta_0, \rho_\infty) = -S$ is invertible, this implies the existence and the consistency of $\hat{\theta}_N$, with standard arguments (see Bibby and Sørensen (1995), Theorem 3.2).

A Taylor's formula around $\hat{\theta}_N$ shows :

$$G_N(\hat{\theta}_N) - G_N(\theta_0) = \left(\int_0^1 \partial_\theta G_N(\theta_0 + u(\hat{\theta}_N - \theta_0)) du \right) (\hat{\theta}_N - \theta_0).$$

As $\partial_\theta G_N(\theta)$ converges uniformly in probability to $\phi(\theta, \theta_0, \rho_\infty)$ and $\hat{\theta}_N$ converges in probability to θ_0,

$$\int_0^1 \partial_\theta G_N(\theta_0 + u(\hat{\theta}_N - \theta_0)) du \xrightarrow{\mathbb{P}_{\theta_0}} \phi(\theta_0, \theta_0, \rho_\infty) = -S.$$

With a second Taylor's formula around $Y_\bullet^{j+1} - Y_\bullet^j$:

$$\begin{aligned} g(\delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta_0, \rho_N) &= (Y_\bullet^{j+1} - Y_\bullet^j)\partial_y g(0,0,Y_\bullet^{j-1};\theta_0,\rho_N) \\ &+\frac{1}{2}(Y_\bullet^{j+1} - Y_\bullet^j)^2 \partial^2_{y^2} g(0,0,Y_\bullet^{j-1},\theta_0,\rho_N) \\ &+\Delta_N g^{(1)}(0, Y_\bullet^{j-1}, \theta_0, \rho_N) \\ &+\Delta_N^{\frac{1}{\alpha-1}} g^{(\alpha)}(0, Y_\bullet^{j-1}; \theta_0, \rho_N) \\ &+\Delta_N^2 R(\Delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta_0, \rho_N). \end{aligned}$$

Then, with Lemma 5.1,

$$\begin{aligned} g(\delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta_0, \rho_N) &= (Y_\bullet^{j+1} - Y_\bullet^j)\partial_y g(0,0,Y_\bullet^{j-1};\theta_0,\rho_N) \\ &-\Delta_N b(Y_\bullet^{j-1},\kappa_0)\partial_y g(0,0,Y_\bullet^{j-1};\theta_0,\rho_N) \\ &+\frac{1}{2}(Y_\bullet^{j+1} - Y_\bullet^j)^2 \partial^2_{y^2} g(0,0,Y_\bullet^{j-1},\theta_0,\rho_N) \\ &-\frac{\Delta_N}{3}\partial^2_{y^2} g(0,0,Y_\bullet^{j-1},\theta_0,\rho_N) \\ &-\Delta_N^{\frac{1}{\alpha-1}} \rho_N^2 \partial^2_{y^2} g(0,0,Y_\bullet^{j-1},\theta_0,\rho_N) \\ &+\Delta_N^2 R(\Delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta_0, \rho_N). \end{aligned}$$

Thus,

$$\begin{aligned} &g(\delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta_0, \rho_N) = \\ &\begin{pmatrix} \partial_y g_1(0,0,Y_\bullet^{j-1};\theta_0,\rho_N)(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1},\kappa_0)) \\ 0 \end{pmatrix} \\ +&\begin{pmatrix} \partial^2_{y^2} g_1(0,0,Y_\bullet^{j-1};\theta_0,\rho_N)(\frac{1}{2}(Y_\bullet^{j+1} - Y_\bullet^j)^2 - \frac{\Delta_N}{3}c(Y_\bullet^{j-1},\lambda_0) - \Delta_N^{\frac{1}{\alpha-1}}\rho_N^2) \\ \partial^2_{y^2} g_2(0,0,Y_\bullet^{j-1};\theta_0,\rho_N)(\frac{1}{2}(Y_\bullet^{j+1} - Y_\bullet^j)^2 - \frac{\Delta_N}{3}c(Y_\bullet^{j-1},\lambda_0) - \Delta_N^{\frac{1}{\alpha-1}}\rho_N^2) \end{pmatrix} \\ &+\Delta_N^2 R(\Delta_N, Y_\bullet^{j+1} - Y_\bullet^j, Y_\bullet^{j-1}; \theta_0, \rho_N). \end{aligned}$$

The results comes from Theorem 5.3. □

Proof of Theorem 5.5. With the notations of the proof of Theorem 5.2, it is sufficient to remark that $u_{3j,N}^{(1)}$ is $\mathcal{H}_{3j+2}^N$ measurable, conditionally centered, and

$$\mathbb{E}((u_{3j,N}^{(1)})^2 | \mathcal{H}_{3j}^N) = \frac{8}{9}\nu_0(f^2\sigma^4) + o_P(1).$$

Then, the correlation between $\zeta_{j+1,N}$ and $\zeta'_{j+1,N}$ is avoided, and the asymptotic variance is reduced in the Central Limit Theorem. □

Troisième partie

Etude asymptotique du filtre particulaire

Chapitre 6

On the asymptotic variance in the Central Limit Theorem for particle filters

Abstract

Particle filter algorithms approximate a sequence of distributions by a sequence of empirical measures generated by a population of simulated particles. In the context of Hidden Markov Models (HMM), they provide approximations of the distribution of optimal filters associated to these models. Given a set of observations, the asymptotic behaviour of particle filters, as the number of particles tends to infinity, has been studied : a central limit theorem holds with an asymptotic variance depending on the fixed set of observations. In this paper we establish, under general assumptions on the hidden Markov model, the tightness of the sequence of asymptotic variances when considered as functions of the random observations as the number of observations tends to infinity. We discuss our assumptions on examples and provide numerical simulations. The case of the Kalman filter is treated separately.

6.1 Introduction

Hidden Markov models (or state-space models) form a class of stochastic models which are used in numerous fields of applications. In these models, a discrete time process $(Y_n, n \geq 0)$ – the signal – is observed while the process of interest $(X_n, n \geq 0)$ – the state process – is not observed. The standard assumptions for the joint-process $(X_n, Y_n)_{n\geq 0}$ are that (X_n) is a Markov chain, that, given $(X_n, n \geq 0)$ the random variables (Y_n) are conditionally independent and the conditional distribution of Y_n only depends on the corresponding state variable X_n. (For general references, see *e.g.* Künsch (2001) or Cappé et al. (2005)).
Nonlinear filtering is concerned with the estimation of X_k or the prediction of X_{k+1} given the observations $(Y_0, \ldots, Y_k) := Y_{0:k}$. For this, one has to compute the conditional distributions $\pi_{k|k:0} = \mathcal{L}(X_k|Y_k, \ldots, Y_0)$ or $\eta_{k+1|k:0} = \mathcal{L}(X_{k+1}|Y_k, \ldots, Y_0)$ which are derived recursively by a sequence of measure-valued operators depending on the observations

$$\pi_{k|k:0} = \Psi_{Y_k}(\pi_{k-1|k-1:0}) \text{ and } \eta_{k+1|k:0} = \Phi_{Y_k}(\eta_{k|k-1:0})$$

(for more details, see *e.g.* Del Moral (2004) or Del Moral and Jacod (2001a)).
Unfortunately, except for very few models, such as the Kalman filter or some other models (for instance, those presented in Chaleyat-Maurel and Genon-Catalot (2006)), these recursions rapidly lead to intractable computations and exact formulae are out of reach. Moreover, the standard Monte-Carlo methods fail to provide good approximations of these distributions (see *e.g.* the introduction in Van Handel (2009)). This justifies the huge popularity of sequential Monte-Carlo methods which are generally the only possible computing approach to solve these problems (see Doucet et al. (2001) or Robert and Casella (2004)). Sequential Monte-Carlo methods (or particle filters, or Interacting Particle Systems) are iterative algorithms based on simulated "particles" which provide approximations of the conditional distributions involved in prediction and filtering.
Denoting by $\pi^N_{k|k:0}$ (resp. $\eta^N_{k+1|k:0}$) the particle filter approximations of $\pi_{k|k:0}$ (resp. $\eta_{k+1|k:0}$) based on N particles, several recent contributions have been concerned with the evaluation of errors between the approximate and the exact filter as N grows to infinity, for a given (fixed) set of data $(Y_k, \ldots, Y_0)$ (see *e.g.* Douc et al. (2005)). In particular, for the bootstrap particle filter, Del Moral and Jacod (2001a) prove that, for a wide class of real-valued functions f, $\sqrt{N}(\pi^N_{k|k:0}(f) - \pi_{k|k:0}(f))$ (resp. $\sqrt{N}(\eta^N_{k+1|k:0}(f) - \eta_{k+1|k:0}(f))$) converges in distribution to $\mathcal{N}(0, \Gamma_{k|k:0}(f))$ (resp. $\mathcal{N}(0, \Delta_{k+1|k:0}(f))$). Central limit theorems for an exhaustive class of sequential Monte-Carlo methods are also proved in Chopin (2004) and Künsch (2005).

To our knowledge, still little attention has been paid to the time behaviour (with respect to k) of the approximations. Recently, Van Handel (2009) has studied a uniform time average consistency of Monte-Carlo particle filters (see also Oudjane and Rubenthaler (2005) for uniform approximation).
In this paper, we are concerned with the tightness of the asymptotic variances $\Gamma_{k|k:0}(f)$, $\Delta_{k+1|k:0}(f)$ in the central limit theorem for the bootstrap particle filter, when considered as random variables functions of $Y_0, \ldots, Y_k$ as $k \to \infty$. This is an important issue since these asymptotic variances measure the accuracy of the numerical method and provide confidence intervals. In Chopin (2004), for the case of the bootstrap filter, the asymptotic variance $\Gamma_{k|k:0}(f)$ is proved to be bounded from above by a constant, under stringent assumptions on the conditional distribution of Y_i given X_i and on the transition densities of the unobserved Markov chain. In Del Moral and Jacod (2001b) the asymptotic variance $\Gamma_{k|k:0}(f)$ is proved to be tight (in k) in the case of the Kalman filter. The proof is based on explicit computations which are possible in this model. Below, we consider a general model and prove the tightness of both $\Gamma_{k|k:0}(f)$ and $\Delta_{k+1|k:0}(f)$ for f a bounded function under a set of assumptions which are milder than those in Chopin (2004) but which do not include the Kalman filter. In general, authors concentrate on filtering rather than on prediction as filtering in more important for applications. However, from the theoretical point of view, we focus on the prediction because computations are a litte simpler. First we prove the tightness of the asymptotic variances $\Delta_{k+1|k:0}(f)$ obtained in the central limit theorem for prediction, and then we deduce the analogous result for $\Gamma_{k|k:0}(f)$. For the transition kernel of the Markov chain, we rely on a strong assumption, which mainly holds when the state space of the hidden chain is compact (Assumption **(A)**). Nevertheless, such an assumption is of common use in this kind of studies (see *e.g.* Atar and Zeitouni (1997), Douc et al. (2005)). In the sense of Douc et al. (2005), it means that the whole state space of the hidden chain is "small" (see Douc et al. (2005)). On the other hand, our assumptions on the conditional distributions of Y_i given X_i are weak (**(B1)-(B2)**). Assumption **(B3)** involves the distribution of the observations, and is easy to check on several classical models.
The paper is organized as follows. In Section 6.2, we present our notations and assumptions, and give the formulae for $\Gamma_{k|k:0}(f)$ and $\Delta_{k+1|k:0}(f)$ and some preliminary propositions in order to obtain formulae as simple as possible for the asymptotic variances. Section 6.3 is devoted to the proof of the tightness of $\Delta_{k+1|k:0}(f)$ from which we deduce the tightness of $\Gamma_{k|k:0}(f)$. Moreover, we illustrate our assumptions on examples and provide some numerical simulation results.

6.2 Notations, assumptions and preliminary results

Let (X_k) be the time-homogeneous hidden Markov chain, with state space $\mathcal{X}$ and transition kernel $Q(x, dx')$. The observed random variables (Y_k) take values in another space $\mathcal{Y}$ and are conditionally independent given $(X_k)_{k\geq 0}$ with $\mathcal{L}(Y_i|(X_k)_{k\geq 0}) = F(X_i, dy)$. For $0 \leq i \leq k$, denote by $Y_{i:k}$ the vector $(Y_i, Y_{i+1} \dots Y_k)$. Denote also, for f a bounded measurable function, $Qf(x) = \int f(x')Q(x, dx')$. Denote by $\pi^{\eta_0}_{k|k:0} = \mathcal{L}_{\eta_0}(X_k|Y_{k:0})$ (resp. $\eta^{\eta_0}_{k|k-1:0} = \mathcal{L}_{\eta_0}(X_k|Y_{k-1:0})$) the filtering distribution (resp. the predictive distribution) at step k when η_0 is the initial distribution of the chain (distribution of X_0). By convention, $\eta^{\eta_0}_{0|-1:0} = \eta_0$.

Let us now introduce our assumptions. Assumptions **A** concern the hidden chain, Assumptions **B** concern the conditional distribution of Y_i given X_i.

(A0) $\mathcal{X}$ is a convex subset of $\mathbb{R}^d$. The transition operator Q admits transition densities with respect to the Lebesgue measure on $\mathcal{X}$ denoted by $dx' : Q(x, dx') = p(x, x')dx'$. The transition densities are positive and continuous on $\mathcal{X} \times \mathcal{X}$. For φ bounded and continuous on $\mathcal{X}$, $Q\varphi$ is bounded and continuous on $\mathcal{X}$ (Q is Feller).

(A1) The transition operator Q admits a stationary distribution $\pi(dx)$ having a density h with respect to dx which is continuous and positive on $\mathcal{X}$.

(A2) There exists a probability measure μ and two positive numbers $\epsilon_- \leq \epsilon_+$ such that

$$\forall x \in \mathcal{X}, \forall B \in \mathcal{B}(\mathcal{X}) \quad \epsilon_-\mu(B) \leq Q(x, B) \leq \epsilon_+\mu(B).$$

Moreover, for all f continuous and positive on $\mathcal{X}$, $\mu(f) > 0$.

(B1) The conditional distribution of Y_k given X_k has density $f(y|x)$ with respect to a dominating measure $\kappa(dy)$, and $(x, y) \longmapsto f(y|x)$ is measurable and positive.

(B2) $x \longmapsto f(y|x)$ is continuous and bounded from above for all y κ a.e.

Under **(B2)**, $q(y) = \sup_{x\in\mathcal{X}} f(y|x)$ is well defined and positive. Up to changing $\kappa(dy)$ into $\frac{1}{q(y)}\kappa(dy)$, we can assume without loss of generality that

$$\forall x \in \mathcal{X}, \quad f(y|x) \leq 1. \tag{6.1}$$

Finally, we introduce an assumption involving the distribution of the observations. Define $g_k(x) := f(Y_k|x)$ for $k \geq 0$.

(B3) For some $\delta > 0$

$$\sup_{k\geq 0} \mathbf{E} \left| \log \left(\eta^{\eta_0}_{k|k-1:0} (g_k) \right) \right|^{1+\delta} < \infty, \tag{6.2}$$

where $\mathbf{E}$ denotes the expectation with respect to the distribution of $(Y_k)_{k\geq 0}$.

Except **(A2)** these assumptions are weak and standard. For instance, **(A0)-(A1)** easily hold for discretized diffusion processes with constant discretization step. Assumption **(A2)**, which is the most stringent, is nevertheless classical and is verified when $\mathcal{X}$ is compact. (see Atar and Zeitouni (1997) and the chronological discussion in Douc et al. (2009)). Assumptions **(B1)-(B2)** are mild. Note that they are weaker than the corresponding ones in Chopin (2004) and the same as in Van Handel (2009). By **(A0)**, for φ non null, non negative and continuous on $\mathcal{X}$, $Q\varphi > 0$. With **(B2)**, for all y κ a.e., $Q(f(y|.))$ is positive, continuous and bounded (by 1).

Note that in **(A0)** we could replace $\mathcal{X}$ by a subset of a Polish space. Choosing $\mathcal{X} \subset \mathbb{R}^d$ is a simplification to check easily Assumptions **(A0)-(A1)** on the examples, especially when the hidden Markov chain comes from the discretization of a diffusion process.

We shall discuss Assumption **(B3)**, because it is not classical. This new assumption is discussed in Section 4 and clarified on examples. It holds whenever $f(y|x)$ is uniformly lower bounded (as in Chopin (2004)) but it is strictly weaker.

Some more notations are needed for the sequel. Define the family of operators : for $f : \mathcal{X} \longrightarrow \mathbb{R}$ measurable and bounded, and $k \geq 0$,

$$L_k f(x) = g_k(x) Q f(x) \text{ where } g_k(x) := f(Y_k|x). \tag{6.3}$$

For $0 \leq i \leq j$, let $L_{i,j} := L_i \dots L_j$ denote the compound operator. For η a probability measure on $\mathcal{X}$, set

$$\Phi_{Y_k}(\eta)(f) = \frac{\eta L_k f}{\eta L_k \mathbf{1}}.$$

Then the predictive distributions satisfy $\eta^{\eta_0}_{0|-1:0} = \eta_0$ and for $k \geq 1$,

$$\eta^{\eta_0}_{k|k-1:0} f = \frac{\eta^{\eta_0}_{k-1|k-2:0} L_{k-1} f}{\eta^{\eta_0}_{k-1|k-2:0} L_{k-1} \mathbf{1}} = \Phi_{Y_{k-1}}(\eta^{\eta_0}_{k-1|k-2:0})(f) \tag{6.4}$$

By iteration,

$$\mathbb{E}_{\eta_0}\left(f(X_k)|Y_{0:k-1}\right) = \eta^{\eta_0}_{k|k-1:0}(f) = \Phi_{Y_{k-1}} \circ \dots \circ \Phi_{Y_0}(\eta_0)(f) = \frac{\eta_0 L_{0,k-1} f}{\eta_0 L_{0,k-1} \mathbf{1}} \tag{6.5}$$

For δ_x the Dirac mass at x, we have

$$\eta^{\delta_x}_{k|k-1:i} f = \frac{\delta_x L_{i,k-1} f}{\delta_x L_{i,k-1} \mathbf{1}} = \Phi_{Y_{k-1}} \circ \dots \circ \Phi_{Y_i}(\delta_x) f. \tag{6.6}$$

We will simply set $\eta_{k|k-1:i}f(x) := \eta^{\delta_x}_{k|k-1:i}f$. Moreover we have the relations

$$\mathbb{E}_{\eta_0}\left(f(X_k)|Y_{0:k}\right) = \pi^{\eta_0}_{k|k:0}f = \frac{\eta^{\eta_0}_{k|k-1:0}(g_k f)}{\eta^{\eta_0}_{k|k-1:0}(g_k)} \tag{6.7}$$

$$\text{and } \eta^{\eta_0}_{k+1|k:0}(f) = \pi^{\eta_0}_{k|k:0}(Qf). \tag{6.8}$$

Note that for all y, $\Phi_y(\delta_x)(dx') = p(x,x')dx'$. For $\eta_0(dx) = h_0(x)dx$, with h_0 positive and continuous on $\mathcal{X}$,

$$\Phi_y(\eta_0)(dx') = \frac{\int_{\mathcal{X}} dx f(y|x)h_0(x)p(x,x')}{\int_{\mathcal{X}} dx f(y|x)p(x,x')}dx',$$

where the denominator is positive. Hence $\Phi_y(\eta)$ has a positive and continuous density when η is a Dirac mass or has a positive and continuous density. For these reasons and assumption **(A0)**, all denominators appearing in our formulae are positive.

Below, for simplicity, when no confusion is possible, we omit the sub- or superscript η_0 in the distributions. Denote the number of interacting particles by N. The distribution of the bootstrap particle filter for the prediction is denoted by $\eta^N_{k|k-1:0}(f)$ and the distribution of the bootstrap particle filter for the filter is denoted by $\pi^N_{k|k:0}(f)$. The following theorem central limit theorem is proved in Del Moral and Jacod (2001a).

Théorème 6.1 *For f a bounded measurable function and a given sequence $(Y_{0:k})$ of observations, the following convergences in distribution hold*

$$\sqrt{N}(\eta^N_{k|k-1:0}(f) - \eta_{k|k-1:0}(f)) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}(0, \Delta_{k|k-1:0}(f))$$

where

$$\Delta_{k|k-1:0}(f) = \sum_{i=0}^{k} \frac{\eta_{i|i-1:0}\left(\left(L_{i,k-1}\left(f - \eta_{k|k-1:0}f\right)\right)^2\right)}{(\eta_{i|i-1:0}L_{i,k-1}\mathbf{1})^2}, \tag{6.9}$$

and

$$\sqrt{N}(\pi^N_{k|k:0}(f) - \pi_{k|k:0}(f)) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}(0, \Gamma_{k|k:0}(f))$$

where

$$\Gamma_{k|k:0}(f) = \sum_{i=0}^{k} \frac{\eta_{i|i-1:0}\left(\left(L_{i,k-1}\left(g_k\left(f - \pi_{k|k:0}f\right)\right)\right)^2\right)}{(\eta_{k|k-1:0}(g_k))^2(\eta_{i|i-1:0}L_{i,k-1}\mathbf{1})^2}. \tag{6.10}$$

Note that, in Del Moral and Jacod (2001a), the above theorem is proved for a wider class of functions, including functions with polynomial growth.

In the sequel, we focus on the two asymptotic variances $\Delta_{k|k-1:0}(f)$ and $\Gamma_{k|k:0}(f)$ for f bounded, when considered as functions of $Y_{0:k}$. Proposition 6.1 gives the link between the two quantities. Recall that the initial distribution is fixed equal to η_0.

Proposition 6.1 *For f a bounded function, and k a non negative integer,*

$$\Gamma_{k|k:0}(f) = \Delta_{k|k-1:0}\left(\frac{g_k}{\eta_{k|k-1:0}(g_k)}\left(f - \pi_{k|k:0}f\right)\right) \tag{6.11}$$

and

$$\Delta_{k+1|k:0}(f) = \eta_{k+1|k:0}\left(\left(f - \eta_{k+1|k:0}f\right)^2\right) + \Gamma_{k|k:0}\left(Qf - \pi_{k|k:0}Qf\right). \tag{6.12}$$

Proof. The first formula is immediate from (6.9) and (6.10). Using (6.4)-(6.8), we get

$$\Delta_{k|k-1:0}(f) = \sum_{i=0}^{k} \eta_{i|i-1:0}\left(\left(\frac{L_{i,k-1}\mathbf{1}(.)}{\eta_{i|i-1:0}L_{i,k-1}\mathbf{1}}\right)^2\left(\eta_{k|k-1:i}f(.) - \eta_{k|k-1:0}f\right)^2\right)$$

$$\Gamma_{k|k:0}(f) = \sum_{i=0}^{k} \frac{\eta_{i|i-1:0}\left(\left(\frac{L_{i,k-1}\mathbf{1}(.)}{\eta_{i|i-1:0}L_{i,k-1}\mathbf{1}}\right)^2\left(\eta_{k|k-1:i}(g_kf)(.) - \pi_{k|k:0}f\right)^2\right)}{(\eta_{k|k-1:0}(g_k))^2}$$

Noting that

$$\eta_{i|i-1:0}\left(L_{i,k-1}\mathbf{1}\right)\,\eta_{k|k-1:0}\left(g_k\right) = \eta_{i|i-1:0}\left(L_{i,k}\mathbf{1}\right),$$

we derive

$$\begin{aligned}
\Delta_{k+1|k:0}(f) &= \eta_{k+1|k:0}\left(\left(f - \eta_{k+1|k:0}f\right)^2\right) \\
&\quad + \Delta_{k|k-1:0}\left(\frac{g_k}{\eta_{k|k-1:0}(g_k)}\left(Qf - \eta_{k+1|k:0}f\right)\right) \\
&= \eta_{k+1|k:0}\left(\left(f - \eta_{k+1|k:0}f\right)^2\right) + \Gamma_{k|k:0}\left(Qf - \eta_{k+1|k:0}f\right).
\end{aligned}$$

□

Note that Chopin (2004) and Künsch (2005) give these recursive formulae in a general context. We recall them in the specific case of filtering distributions, for convenience of the reader.

6.3 Tightness of the asymptotic variances

To stress the dependence on the observations (Y_k), we introduce another notation for $\eta^{\nu}_{k|k-1:0}$. For ν a probability measure, A a borelian set, $y_{0:k-1}$ a set of fixed real values, let us introduce

$$\begin{aligned}\eta_{\nu,k}[y_{0:k-1}](A) &= \frac{\mathbb{E}_\nu(\prod_{i=0}^{k-1} f(y_i|X_i)\mathbf{1}_A(X_k))}{\mathbb{E}_\nu(\prod_{i=0}^{n-1} f(y_i|X_i))} \\ &= \frac{\nu L_{0,k-1}\mathbf{1}_A}{\nu L_{0,k-1}\mathbf{1}} = \frac{\nu g_0 Q g_1 Q \dots g_{k-1} Q \mathbf{1}_A}{\nu g_0 Q g_1 Q \dots g_{k-1}}\end{aligned}$$

Here, $\mathbb{E}_\nu$ denotes the expectation with respect to the distribution of the chain (X_k) with initial distribution ν and $y_{0:k-1}$ are fixed values not involved in the expectation. To ensure that all expressions are well defined, we consider probability measures ν equal to either Dirac masses or probabilities with positive continuous densities on $\mathcal{X}$. In the second line, we have set $g_i(x) = f(y_i|x)$ and the formula explains the backward iterations of the operators (6.3) with $Y_k = y_k$.

The following proposition proves the exponential forgetting of the initial distribution for the predictive distributions.

Proposition 6.2 *Assume* **(A2)** *and* **(B1)** *and set* $\rho = 1 - \frac{\epsilon_-^2}{\epsilon_+^2}$. *Then for all non negative integer* k, *all probability distributions* ν *and* ν' *on* $\mathcal{X}$ *and all set* $y_{0:k-1}$ *of real values*

$$\|\eta_{\nu,k}[y_{0:k-1}] - \eta_{\nu',k}[y_{0:k-1}]\|_{TV} \leq \rho^k,$$

where $\|.\|_{TV}$ *denotes the total variation distance.*

Proof. The above result is generally proved for the filtering distribution (see *e.g.* Atar and Zeitouni (1997), Del Moral and Guionnet (2001) and Douc et al. (2009)). To prove it for the predictive distributions, we follow the scheme of Douc et al. (2009). Let $\bar{\mathcal{X}} = \mathcal{X} \times \mathcal{X}$ and denote by $\bar{Q}$ the Markov kernel on $\bar{\mathcal{X}}$ given by

$$\bar{Q}((x,x'), A \times A') = Q(x,A)Q(x',A').$$

Set $\bar{g}_i(x,x') = g_i(x)g_i(x')$. For two probability measures ν and ν', notice that

$$\eta_{\nu,k}[y_{0:k-1}](A) - \eta_{\nu',k}[y_{0:k-1}](A) = \Phi_{y_{k-1}} \circ \dots \circ \Phi_{y_0}(\nu)(\mathbf{1}_A) - \Phi_{y_{k-1}} \circ \dots \circ \Phi_{y_0}(\nu')(\mathbf{1}_A)$$

$$= \frac{\mathbb{E}_{\nu\otimes\nu'}(\prod_{i=0}^{k-1} \bar{g}_i(X_i,X_i')\mathbf{1}_A(X_k)) - \mathbb{E}_{\nu'\otimes\nu}(\prod_{i=0}^{k-1} \bar{g}_i(X_i,X_i')\mathbf{1}_A(X_k))}{\mathbb{E}_\nu(\prod_{i=0}^{k-1} g_i(X_i))\mathbb{E}_{\nu'}(\prod_{i=0}^{k-1} g_i(X_i))}$$

where (X_i) and (X_i') are two independent copies of the hidden Markov chain and $\mathbb{E}_{\nu\otimes\nu'}$ denotes the expectation with respect to the distribution of the chain

(X_i, X_i') with kernel $\bar{Q}$ and initial distribution $\nu \otimes \nu'$.
Set $\bar{\mu} = \mu \otimes \mu$, and $\bar{x} = (x, x')$. For $\bar{f}$ a measurable non negative function, we have

$$\epsilon_-^2 \bar{\mu}(\bar{f}) \leq \bar{Q}(\bar{x}, \bar{f}) \leq \epsilon_+^2 \bar{\mu}(\bar{f}).$$

Setting

$$\bar{Q}_0(\bar{x}, \bar{f}) = \epsilon_-^2 \bar{\mu}(\bar{f}) \quad \text{and} \quad \bar{Q}_1(\bar{x}, \bar{f}) = \bar{Q}(\bar{x}, \bar{f}) - \bar{Q}_0(\bar{x}, \bar{f}),$$

we deduce that

$$0 \leq \bar{Q}_1(\bar{x}, \bar{f}) \leq \rho \bar{Q}(\bar{x}, \bar{f}).$$

Now let us compute the numerator :

$$R_k(\nu, \nu', A) = \mathbb{E}_{\nu \otimes \nu'}(\prod_{i=0}^{k-1} \bar{g}_i(X_i, X_i') \mathbf{1}_A(X_k)) - \mathbb{E}_{\nu' \otimes \nu}(\prod_{i=0}^{k-1} \bar{g}_i(X_i, X_i') \mathbf{1}_A(X_k))$$

$$= \nu \otimes \nu'(\bar{g}_0 \bar{Q} \bar{g}_1 \bar{Q} \dots \bar{g}_{k-1} \bar{Q} \mathbf{1}_{A \times \mathcal{X}}) - \nu' \otimes \nu(\bar{g}_0 \bar{Q} \bar{g}_1 \bar{Q} \dots \bar{g}_{k-1} \bar{Q} \mathbf{1}_{A \times \mathcal{X}}).$$

It may be decomposed as

$$R_k(\nu, \nu', A) = \sum_{t_{0:k-1} \in \{0,1\}^k} R_k(A, t_{0:k-1})$$

where

$$R_k(A, t_{0:k-1}) := \nu \otimes \nu'(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \bar{Q}_{t_1} \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{A \times \mathcal{X}}) - \nu' \otimes \nu(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \bar{Q}_{t_1} \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{A \times \mathcal{X}}).$$

Assume that for an index i, $t_i = 0$. Then

$$\begin{aligned}
&\nu \otimes \nu'(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \bar{Q}_{t_1} \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{A \times \mathcal{X}}) \\
&= \nu \otimes \nu'(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \bar{Q}_{t_1} \dots \bar{g}_{i-1} \bar{Q}_{t_{i-1}} \bar{g}_i) \times \epsilon_-^2 \bar{\mu} \left(\bar{g}_{i+1} \bar{Q}_{t_{i+1}} \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{A \times \mathcal{X}} \right) \\
&= \nu' \otimes \nu(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \bar{Q}_{t_1} \dots \bar{g}_{i-1} \bar{Q}_{t_{i-1}} \bar{g}_i) \times \epsilon_-^2 \bar{\mu} \left(\bar{g}_{i+1} \bar{Q}_{t_{i+1}} \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{A \times \mathcal{X}} \right) \\
&= \nu' \otimes \nu(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \bar{Q}_{t_1} \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{A \times \mathcal{X}}) \\
&= \nu \otimes \nu'(\bar{g}_0 \bar{Q}_{t_0} \bar{g}_1 \bar{Q}_{t_1} \dots \bar{g}_{k-1} \bar{Q}_{t_{k-1}} \mathbf{1}_{\mathcal{X} \times A})
\end{aligned}$$

and $R_k(A, t_{0:k-1})$ vanishes except if all $t_i = 1$. Hence

$$R_k(\nu, \nu', A) = \nu \otimes \nu'(\bar{g}_0 \bar{Q}_1 \bar{g}_1 \bar{Q}_{t_1} \dots \bar{g}_{k-1} \bar{Q}_1 (\mathbf{1}_{A \times \mathcal{X}} - \mathbf{1}_{\mathcal{X} \times A})).$$

Therefore

$$\sup_A |R_k(\nu, \nu', A)| \leq \rho^k \mathbb{E}_{\nu \otimes \nu'}(\prod_{i=0}^{k-1} \bar{g}_i(X_i, X_i')).$$

The result follows.□

Remark. Applying the result of Proposition 6.2 with $g_i \equiv 1$ and $\nu' = \pi$, we get $\|\nu Q^k - \pi\|_{TV} \leq \rho^k$. Thus, **(A1)-(A2)** imply the geometric ergodicity of (X_k).

Let us make some comments. The result is given in Del Moral and Guionnet (2001), but the arguments of the proof are different and rely on stronger assumptions.

Proposition 6.3 *Assume* **(A0)-(A2) (B1)-(B2)** *For f a bounded measurable function and $(Y_{0:k})$ a sequence of observations, the following inequalities hold*

$$\Delta_{k|k-1:0}(f) \leq \|f\|_\infty^2 \sum_{i=0}^{k} \eta_{i|i-1:0} \left(\left(\frac{L_{i,k-1}\mathbf{1}}{\eta_{i|i-1:0} L_{i,k-1}\mathbf{1}} \right)^2 \right) \rho^{2(k-i)} \tag{6.13}$$

The following propositions give upper bounds for $\Delta_{k|k-1:0}(f)$.

Proof. Remark that, for all ν :

$$\eta_{\nu,k}[Y_{0:k-1}] = \Phi_{Y_{k-1}} \circ \cdots \circ \Phi_{Y_i}(\eta_{\nu,i}[Y_{0:i-1}]).$$

By Proposition 6.2, we deduce, for $\nu = \eta_0$:

$$|\eta_{k|k-1:i}(f)(x) - \eta^{\eta_0}_{k|k-1:0}(f)| \leq \|f\|_\infty \|\eta_{\delta_x,k-i}[Y_{i:k-1}] - \eta_{\eta_0,k}(Y_{0:k-1})\|_{TV}$$

$$\leq \|f\|_\infty \|\Phi_{y_{k-1}} \circ \cdots \circ \Phi_{y_i}(\delta_x) - \Phi_{y_{k-1}} \circ \cdots \circ \Phi_{y_i}(\eta_{\eta_0,i}[Y_{0:i-1}])\|_{TV}$$

$$\leq \|f\|_\infty \rho^{k-i}.$$

Using (6.13), we get the result. □
We stress the fact that Proposition 6.3 only relies on the exponential stability which may hold even if **(A2)** is not satisfied (see Douc et al. (2009)).

Proposition 6.4 *Under Assumption* **(A2)** *and for f measurable and bounded, it holds that*

$$\Delta_{k|k-1:0}(f) \leq \|f\|_\infty^2 \frac{\epsilon_+^2}{\epsilon_-^2} \sum_{i=0}^{k} \eta_{i|i-1:0} \left(\left(\frac{g_i}{\eta_{i|i-1:0} g_i} \right)^2 \right) \rho^{2(k-i)}. \tag{6.14}$$

Proof. We remark that for any probability measure ν

$$\epsilon_- \nu(g_i)\mu(g_{i+1}Q \dots g_{k-1}) \leq \nu g_i Q g_{i+1} Q \dots g_{k-1} \leq \epsilon_+ \nu(g_i)\mu(g_{i+1}Q \dots g_{k-1}).$$

By **(A2)**, since Q is Feller and the g_l's are positive continuous, $\mu(g_{i+1}Q\dots g_{k-1})$ is positive. Applying the left inequality with $\nu = \eta_{i|i-1:0}$ and the right inequality with $\nu = \delta_x$, it comes

$$\frac{L_{i,k-1}\mathbf{1}(x)}{\eta_{i|i-1:0}L_{i,k-1}\mathbf{1}} \leq \frac{\epsilon_+}{\epsilon_-}\frac{g_i(x)\mu(g_{i+1}Q\dots g_{k-1})}{\eta_{i|i-1:0}(g_i)\mu(g_{i+1}Q\dots g_{k-1})}.$$

Thus,

$$\Delta_{k|k-1:0}(f) \leq \|f\|_\infty^2 \frac{\epsilon_+^2}{\epsilon_-^2}\sum_{i=0}^{k}\eta_{i|i-1:0}\left(\left(\frac{g_i}{\eta_{i|i-1:0}g_i}\right)^2\right)\rho^{2(k-i)}.$$

□

We state the main result under the additional assumption **(B3)**. In Section 6.4, we show that, under **(A1)**, **(B3)** is especially easy to check.

Théorème 6.2 *Assume* **(A0)-(A2)**, **(B1)-(B3)**. *Then, for all bounded function f, the sequences of variances $(\Delta_{k|k-1:0}(f))$ and $(\Gamma_{k|k:0}(f))$ are tight.*

Proof. Using that $g_i \leq 1$

$$\eta_{i|i-1:0}\left(\left(\frac{g_i}{\eta_{i|i-1:0}g_i}\right)^2\right) \leq \frac{1}{\eta_{i|i-1:0}g_i}.$$

Setting $B_i = -\log(\eta_{i|i-1:0}g_i)$, Lemma C.1 (see the Appendix) implies that the sequence $\sum_{i=0}^{k} e^{B_i}\rho^{2(k-i)}$ is tight with respect to k. With Proposition 6.4, we deduce that $(\Delta_{k|k-1:0}(f))_{k\geq 0}$ is tight.

Using (6.11) and $g_k \leq 1$, we obtain :

$$\Gamma_{k|k:0}(f) \leq \frac{1}{(\eta_{k|k-1:0}g_k)^2}\Delta_{k|k-1:0}\left(f - \pi_{k|k:0}f\right).$$

Since $\|f - \pi_{k|k:0}f\|_\infty \leq 2\|f\|_\infty$, the first part implies that $(\Delta_{k|k-1:0}\left(f - \pi_{k|k:0}f\right))$ is tight. By **(B3)**, $(\eta_{k|k-1:0}(g_k))$ is also tight. The result follows. □

We have used in the proof that $f(y|x) \leq 1$ for all $x \in \mathcal{X}$. But as we claimed previously, up to the choice of the dominating measure κ this is not a restriction.

6.4 Discussion and examples

6.4.1 Checking of (B3)

Let us consider a hidden chain with state space $\mathcal{X} = [a, b]$ a compact interval of $\mathbb{R}$ satisfying **(A0)-(A2)** (for instance a discrete sampling of a diffusion on $[a, b]$

with reflecting boundaries). Under **(B2)**, $r(y) = \inf_{x\in\mathcal{X}} f(y|x)$ is well defined and positive. Thus, we have

$$r(Y_k) \leq \eta_{k|k-1:0}(g_k) \leq 1.$$

Therefore, the condition $\sup_{k\geq 0} \mathbf{E}\left|\log\left(r(Y_k)\right)\right|^{1+\delta} < \infty$ implies **(B3)**. In particular, when (Y_k) is stationary *i.e.* when the initial distribution of the chain is $\eta_0 = \pi$ the stationary distribution, the condition is simply $\mathbf{E}\left|\log\left(r(Y_0)\right)\right|^{1+\delta} < \infty$. Let us compute $r(y)$ in some typical examples.

Example 1. Assume that $Y_k = X_k + \varepsilon_k$ with $\varepsilon_k \sim_{i.i.d.} \mathcal{N}(0,1)$ and (X_k) independent of (ε_k). The observation kernel is $F(x, dy) = \mathcal{N}(x, 1)$. Choosing the dominating measure $\kappa(dy) = \frac{1}{\sqrt{2\pi}}dy$,

$$f(y|x) = \exp\left(-\frac{(y-x)^2}{2}\right) \geq r(y)$$

where

$$\left|\log(r(y))\right| \leq \frac{1}{2}\left((y-a)^2 + (y-b)^2\right).$$

Example 2. Assume that $Y_k = \sqrt{X_k}\varepsilon_k$ with $\varepsilon_k \sim_{i.i.d.} \mathcal{N}(0,1)$, (X_k) independent of (ε_k) and $0 < a < b$. The observation kernel is $F(x, dy) = \mathcal{N}(0, x)$. Note that

$$\frac{1}{\sqrt{2\pi b}}\exp\left(-\frac{y^2}{2a}\right) \leq \frac{1}{\sqrt{2\pi x}}\exp\left(-\frac{y^2}{2x}\right) \leq \frac{1}{\sqrt{2\pi a}}.$$

Taking $\kappa(dy) = \frac{1}{\sqrt{2\pi a}}dy$, we get that

$$\left|\log(r(y))\right| \leq C + \frac{y^2}{2a}.$$

Thus, assumption **(B3)** is a simple moment condition on the observations which is evidently satisfied on these examples.

6.4.2 The case of a diffusion on a compact manifold

Consider the stochastic differential equation

$$dZ_t = b(Z_t)dt + \sigma(Z_t)dW_t$$

with a one-dimensional observation process

$$Y_{t_i} = g(Z_{t_i}) + \varepsilon_i$$

where W is a standard Brownian motion, (ε_i) is an i.i.d. sequence of $\mathcal{N}(0,1)$ random variables, and b, σ are Lipschitz and g is smooth enough.

Assume that the diffusion process Z valued in a compact manifold M of dimension m embedded in $\mathbb{R}^d$. Assume that b and σ lead to a strictly elliptic generator on M, with heat kernel $G_t(x,y)$. We refer to Atar and Zeitouni (1997) and Davies (1989) for the following inequality

$$c_0 e^{-c_1/t} \leq G_t(x,y) \leq c_2 t^{-m/2}.$$

where c_0, c_1 and c_2 are numerical constants. Assume that $t_i = i\delta, \quad \delta > 0, \quad i \in \mathbb{N}$, hence the observations are equally spaced in time. Then we obtain the inequality for **(A2)** with μ a probability distribution with positive density with respect to Lebesgue measure on M, because the transition density of the hidden Markov chain is bounded from below by a positive value. Due to the underlying diffusion process, other assumptions on the chain are verified.

6.4.3 A toy-example

Consider two continuous densities u and v on $(0,1)$, a distribution π on $(0,1)$ with a continuous density with respect to Lebesgue measure and a real α of $(0,1)$. Define the Markov chain (X_k) by

$$X_0 \sim \pi, \qquad X_{k+1} = \mathbf{1}_{X_k<\alpha} U_{k+1} + \mathbf{1}_{X_k \geq \alpha} V_{k+1} \tag{6.15}$$

where (U_k) and (V_k) are two independent sequences of i.i.d. random variables, independent of X_0, with respective distributions $u(x)dx$ and $v(x)dx$. Set $p(x,x') = \mathbf{1}_{x<\alpha} u(x') + \mathbf{1}_{x \geq \alpha} v(x')$ the transition kernel density. The transition kernel admits an invariant distribution $\pi(x)dx$ with $\pi(x) = Au(x) + (1-A)v(x)$ and

$$A = \frac{\int_0^\alpha v(x)dx}{\int_\alpha^1 u(x)dx + \int_0^\alpha v(x)dx}.$$

For example, with

$$u(x) = \begin{cases} 6x \text{ if } x \in [0, \frac{1}{3}] \\ -3x+3 \text{ if } x \in]\frac{1}{3}, 1] \end{cases} \quad \text{and } v(x) = \begin{cases} 3x \text{ if } x \in [0, \frac{2}{3}] \\ -6x+6 \text{ if } x \in]\frac{2}{3}, 1] \end{cases}$$

the transition kernel Q of the chain (X_k) satisfies **(A2)** (see Figure 6.1) with $\mu(dx) = 4(x \wedge 1-x)dx$ and $\epsilon_- = \frac{1}{4}, \quad \epsilon_+ = \frac{3}{2}$:

$$\forall x \in \mathcal{X}, \forall B \in \mathcal{B}(\mathcal{X}) \quad \epsilon_- \mu(B) \leq Q(x,B) \leq \epsilon_+ \mu(B).$$

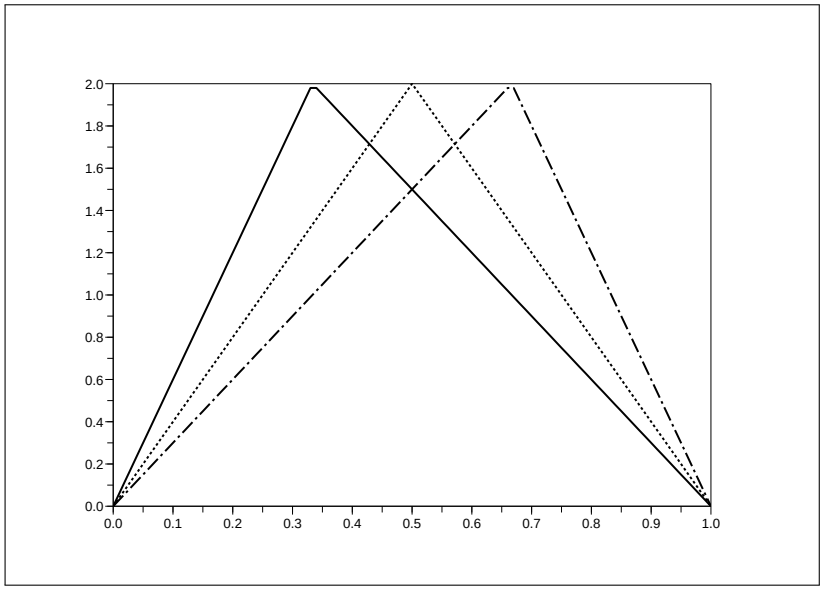

FIGURE 6.1 – Densities involved in the toy-example. Functions u (solid), v (bigdash dot) and $\frac{d\mu}{dx}$ (dash dot).

In Figure 6.1, the graph of u is plotted in solid line, the graph of v is plotted in bigdash dotted line, and the density of μ is plotted in dash dotted line.
For this example, the transition density, $p(x, x')$ is not bounded from below by a positive constant. This shows that **(A2)** is strictly weaker than the assumption of Theorem 5 in Chopin (2004). Although Assumption **(A0)** is not verified on this example, the proof of the tightness still holds. Indeed, all the denominators involved in the computations of the upper bounds are well-defined and positive. Assumption **(A1)** is clearly verified and the stationary distribution can be explicitely computed, and the bounds in **(A2)** too.

6.5 Numerical simulations

6.5.1 Simulations based on the toy-example

We consider Example 6.4.3 with the observations $Y_k = X_k + \varepsilon_k$ where $(\varepsilon_k)_k$ is a sequence of i.i.d. $\mathcal{N}(0, (0.5)^2)$ random variables. In figure 6.2, we have plotted in plain line, with square marks, a trajectory of the hidden Markov chain, in longdashed line, with diamond marks, the noisy observations. We have plotted in dashed line, with plus marks, the result of the bootstrap particle filter associated to these observations, with $N = 500$ particles and $\varphi(x) = x$. We observe that the

result of the particle filter is close to the hidden chain, uniformly in time.

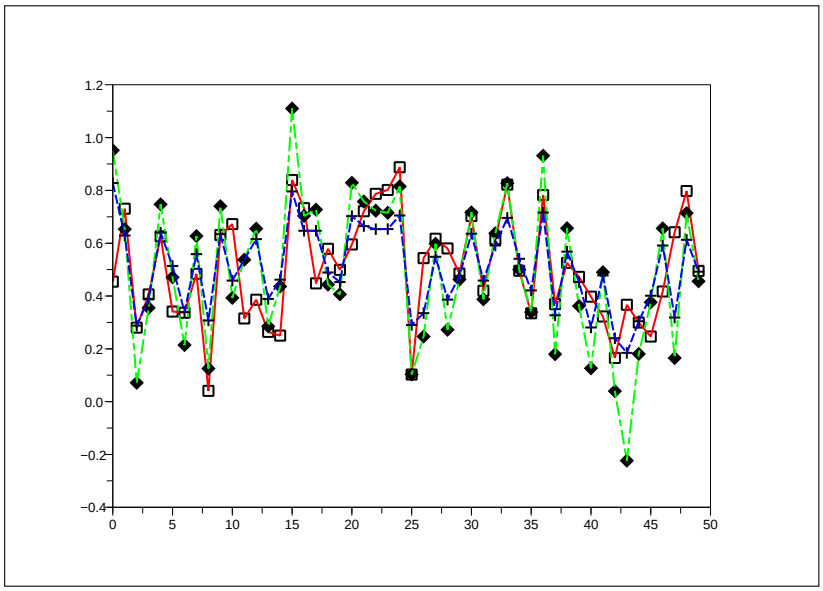

FIGURE 6.2 – Toy-example ($\alpha = 0.4$). Hidden Markov chain (plain, $\square$ marks). Observations (longdashed line, diamond marks). Particle filter (dashed line, $+$ marks)

We propose a study of the variances based on Monte-Carlo simulations.

We simulate $J = 50$ trajectories $(y_0^{(j)}, \dots, y_T^{(j)})$ for $j = 1 \dots J$ and $T = 200$. For each trajectory, we compute $L = 50$ realizations of the particular Monte-Carlo method, named $(\pi_{k|k:0}^{(j),N}\varphi)_l$ for $l = 1 \dots L$ and $k = 0 \dots T$. Hence we set

$$\hat{\Gamma}_{k|k:0}^{(j),N}(\varphi) = \frac{N}{L}\sum_{l=1}^{L}\left((\pi_{k|k:0}^{(j),N}\varphi)_l - \frac{1}{L}\sum_{l=1}^{L}(\pi_{k|k:0}^{(j),N}\varphi)_l\right)^2. \tag{6.16}$$

	$\hat{\Gamma}_{50\|50:0}^{(1),N}(\varphi)$	$\hat{\Gamma}_{100\|100:0}^{(1),N}(\varphi)$	$\hat{\Gamma}_{150\|150:0}^{(1),N}(\varphi)$	$\hat{\Gamma}_{200\|200:0}^{(1),N}(\varphi)$
$N = 100$	6.95×10^{-2}	4.00×10^{-2}	4.14×10^{-2}	5.27×10^{-2}
$N = 250$	5.91×10^{-2}	4.01×10^{-2}	3.35×10^{-2}	4.27×10^{-2}
$N = 500$	6.46×10^{-2}	3.24×10^{-2}	3.69×10^{-2}	4.60×10^{-2}

TABLE 6.1 – Computation of (6.16) for different values of N and k.

In Table 6.1, the quantity (6.16) is computed for one trajectory, and different values of the number of particules N and the number of observations k. We can notice that the value remains stable.

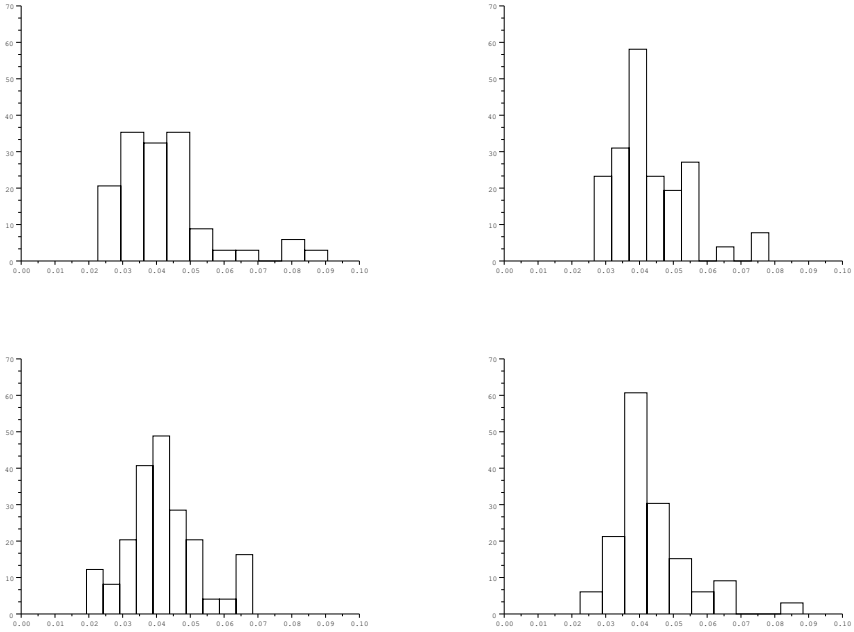

FIGURE 6.3 – Histograms of the random variables (6.16) for $k = 50$ (top left), 100 (top right), 150 (bottom left), 200 (bottom right).

In Figure 6.3, the histograms of the distribution of the random variables $(\hat{\Gamma}^{(j),500}_{k|k:0}(\varphi), j = 1 \dots J)$ for $k \in \{50, 100, 150, 200\}$ are plotted. This confirms the tightness property : the support remains within a small compact set and the distributions are concentrated around an unique mode.

6.6 Acknowledgements

The author wishes to thank Pr. Genon-Catalot, his PhD advisor, for her help and the numerous discussions. He also wishes to thank Prs. Del Moral and Jacod for their advices and bibliographical references. He wishes finally to thank Mahendra Mariadassou for discussions about the numerical examples, and the anonymous referees for their suggestions and one bibliographical reference.

Appendices

Annexe C

Appendix

C.1 Bootstrap particle filter

The aim is to build a sequence of measures $(\eta^N_{k|k-1:0})_k$, where N is the number of interacting particles, so that $\eta^N_{k|k-1:0}f$ is a good approximation of $\eta_{k|k-1:0}f$ for f bounded. We assume that the distribution of X_0 is known and we set it as $\eta_0 = \eta_{0|-1:0}$. We assume that we are able to simulate random variables under η_0 and under $Q(x, dx')$.

Step 0 : Simulate $(X^j_0)_{1\leq j\leq N}$ i.i.d. with distribution η_0 and compute $\eta^N_{0|-1:0} = \frac{1}{N}\sum_{j=1}^N \delta_{X^j_0}$.

Step 1-a : Simulate X'^j_0 i.i.d. with distribution $\pi^N_{0|0:0} = \sum_{j=1}^N \frac{g_0(X^j_0)}{\sum_{j=1}^N g_0(X^j_0)}\delta_{X^j_0}$.

Step 1-b : Simulate N random variables $(X^j_1)_j$ independantly with $X^j_1 \sim Q(X'^j_0, dx)$. Set $\eta^N_{1|0:0} = \frac{1}{N}\sum_{i=1}^N \delta_{X^j_1}$.

Step k-a : (updating) Suppose that $\eta^N_{k|k-1:0}$ is known. Simulate $(X^j_k)_{1\leq j\leq N}$ i.i.d. with distribution $\eta^N_{k|k-1:0}$ and simulate X'^j_k i.i.d.with distribution $\pi^N_{k|k:0} = \sum_{j=1}^N \frac{g_k(X^j_k)}{\sum_{j=1}^N g_k(X^j_k)}\delta_{X^j_k}$.

Step k-b : (prediction) Simulate X^N_{k+1} independantly with $X^N_{k+1} \sim Q(X'^j_k, dx)$. Set $\eta^N_{k+1|k:0} = \frac{1}{N}\sum_{j=1}^N \delta_{X^j_{k+1}}$.

C.2 Tightness lemma

The following lemma is proved (with $\delta = 1$) in Del Moral and Jacod (2001b).

Lemme C.1 *(Tightness lemma) Let $\alpha \in (0,1)$ and consider two sequences $(A_{i,k})_{1\leq i\leq k}$ and $(B_{i,k})_{1\leq i\leq k}$ of non negative random variables such that*

$$\sup_{i,k} \mathbb{E}(A_{i,k}) + \sup_{i,k} \mathbb{E}(B_{i,k}^{1+\delta}) = K < \infty \tag{C.1}$$

then the sequence

$$\Upsilon_k = \sum_{i=1}^{k} \alpha^{k-i} A_{i,k} e^{B_{i,k}} \tag{C.2}$$

is tight.

Proof. Choose $\gamma > 1$ such that $\alpha\gamma < 1$. Set $\Omega_{j,k} = \cap_{i=1}^{k-j}\{|B_{i,k}| \leq (k-i)\log\gamma\}$ for $1 \leq j \leq k$. Set also $\epsilon_j = \sum_{i\geq j} \frac{1}{i^{1+\delta}}$. Then, with the Markov inequality for $B_{i,k}$ we have

$$\mathbb{P}(\Omega_{j,k}^c) \leq \frac{K}{(\log\gamma)^{1+\delta}} \sum_{i=1}^{k-j} \frac{1}{(k-i)^{1+\delta}} \leq \frac{K\epsilon_j}{(\log\gamma)^{1+\delta}}.$$

On $\Omega_{j,k}$ we have

$$\begin{aligned}
\sum_{i=1}^{k} \alpha^{k-i} A_{i,k} e^{B_{i,k}} &= \sum_{i=1}^{k-j} \alpha^{k-i} A_{i,k} e^{B_{i,k}} + \sum_{i=k-j+1}^{k} \alpha^{k-i} A_{i,k} e^{B_{i,k}} \\
&\leq \sum_{i=1}^{k-j} (\gamma\alpha)^{k-i} A_{i,k} + \sum_{i=k-j+1}^{k} \alpha^{k-i} A_{i,k} e^{B_{i,k}}
\end{aligned}$$

Finally, with the Markov inequality for $A_{i,k}$, we get for $1 \leq j \leq k$

$$\mathbb{P}(\Upsilon_k > M) \leq \frac{K\epsilon_j}{(\log\gamma)^{1+\delta}} + \frac{2K}{M(1-\alpha\gamma)} + \sum_{i=k-j+1}^{k} \mathbb{P}(A_{i,k} e^{B_{i,k}} > \frac{M}{2j})$$

With our assumption, the sequence $(A_{i,k}e^{B_{i,k}})_{1\leq i\leq k}$ is tight. For $\epsilon > 0$ we first choose j then M, hence the result. □

C.3 The Gaussian case

In the context of the one-dimensional Kalman filter model, Assumptions **(A0)-(A1) (B1)-(B2)** hold but not **(A2)**. On the other hand, we are able to study the expression (6.13) of $\Delta_{k|k-1:0}(f)$ by explicit computations. In Del Moral and Jacod (2001b), (6.13) is proved to be tight by direct computations too. We do

the analogous calculus for (6.13) which is simpler. Recall the model : the hidden Markov chain is a Gaussian AR(1) process

$$\begin{aligned} X_0 &\sim \mathcal{N}(0, \sigma_s^2) \\ X_{k+1} &= aX_k + \beta U_{k+1} \end{aligned}$$

where (U_k) is a sequence of $\mathcal{N}(0,1)$ independent random variables, independent of X_0. The observations are given by

$$Y_k = bX_k + \beta' V_k$$

where (V_k) is a sequence of $\mathcal{N}(0,1)$ independent random variables, independent of (X_k). We assume that $|a| < 1$ and $\sigma_s^2 = \frac{\beta^2}{1-a^2}$. Hence, the process (X_k, Y_k) is stationary.

C.3.1 Preliminary computations

Denote by η_{m,σ^2} the Gaussian distribution $\mathcal{N}(m, \sigma^2)$ and by ϕ_{m,σ^2} its density. Due to the identity

$$\begin{aligned} \frac{1}{\sigma^2}(x-m)^2 + \frac{1}{v^2}(x-u)^2 &= \left(\frac{1}{\sigma^2} + \frac{1}{v^2}\right)\left(x - \frac{\frac{m}{\sigma^2} + \frac{u}{v^2}}{\frac{1}{\sigma^2} + \frac{1}{v^2}}\right)^2 - \frac{\left(\frac{m}{\sigma^2} + \frac{u}{v^2}\right)^2}{\frac{1}{\sigma^2} + \frac{1}{v^2}} \\ &\quad + \left(\frac{m^2}{\sigma^2} + \frac{u^2}{v^2}\right), \end{aligned}$$

we get

$$\eta_{m,\sigma^2}(\phi_{u,v^2}) = \frac{1}{\sqrt{2\pi\,(\sigma^2 + v^2)}} \exp\left(-\frac{(m-u)^2}{2(v^2+\sigma^2)}\right). \tag{C.3}$$

The prediction operator L_k is given by $L_k(x,.) = \frac{1}{b}\phi_{\frac{Y_k}{b}, \frac{\beta'^2}{b^2}}(x)\mathcal{N}(ax, \beta^2)$. To compute (6.13), following Del Moral and Jacod (2001b) we search the compound operator $L_{i,j} = L_i \dots L_j$ in the following form :

$$L_{i,j}(x,.) = u_{i,j}\phi_{v_{i,j}, w_{i,j}}(x)\mathcal{N}(\theta_{i,j}x + \gamma_{i,j}, \delta_{i,j}).$$

After some technical computations, the parameters are recursively given by :

$$u_{i,i} = \frac{1}{b}, \quad v_{i,i} = \frac{Y_i}{b}, \quad w_{i,i} = \frac{\beta'^2}{b^2}, \quad \theta_{i,i} = a, \quad \gamma_{i,i} = 0, \quad \delta_{i,i} = \beta^2.$$

For $i < j$

$$\begin{aligned} \theta_{i,j} &= \tfrac{a w_{i+1,j}\theta_{i+1,j}}{\beta^2 + w_{i+1,j}}, & \gamma_{i,j} &= \gamma_{i+1,j} + \theta_{i+1,j}\left(\tfrac{\beta^2 v_{i+1,j}}{\beta^2 + w_{i+1,j}}\right), \\ \delta_{i,j} &= \delta_{i+1,j} + \theta_{i+1,j}^2\left(\tfrac{\beta^2 w_{i+1,j}}{\beta^2 + w_{i+1,j}}\right), & v_{i,j} &= \tfrac{\frac{Y_i}{b}(\beta^2 + w_{i+1,j}) + a\frac{\beta'^2}{b^2}v_{i+1,j}}{(\beta^2 + w_{i+1,j}) + \frac{\beta'^2}{b^2}a^2}, \\ w_{i,j} &= \tfrac{\beta'^2}{b^2}\tfrac{\beta^2 + w_{i+1,j}}{(\beta^2 + w_{i+1,j}) + \frac{\beta'^2}{b^2}a^2}, \end{aligned}$$

$$u_{i,j} = \frac{u_{i+1,j}}{\sqrt{2\pi\left((\beta^2 + w_{i+1,j}) + \frac{\beta'^2}{b^2}a^2\right)}} \exp\left(-\frac{1}{2}\frac{(v_{i+1,j} - a\frac{Y_i}{b})^2}{(\beta^2 + w_{i+1,j}) + \frac{\beta'^2}{b^2}a^2}\right).$$

It appears that the recursion for $w_{i,j}$ is driven by an autonomous algorithm, which has a unique fixed point $\underline{w}$ defined as the solution of

$$\underline{w} = \frac{\beta'^2}{b^2}\frac{\beta^2 + \underline{w}}{(\beta^2 + \underline{w}) + \frac{\beta'^2}{b^2}a^2}.$$

The downward recursions for all parameters can be solved by elementary but extremely cumbersome computations. Since we just want to illustrate the way tightness is obtained for (6.13), we do not proceed to the exact computations. Instead, we introduce a *stabilized* operator

$$\underline{L}_k(x, .) = \frac{1}{b}\phi_{\frac{Y_k}{b},\underline{w}}(x)\mathcal{N}(ax, \beta^2)$$

and compute the compound operator $\underline{L}_{i,j} = \underline{L}_i \dots \underline{L}_j$. We denote by $\underline{\Delta}_{k|k-1:0}(f)$ the *stabilized* asymptotic variance, *i.e.* (6.13) computed with $\underline{L}_k(., .)$ instead of the exact L_k.

C.3.2 Solving recursions for the stabilized operators

Define

$$\underline{L}_{i,j}(x, .) = u_{i,j}\phi_{v_{i,j},\underline{w}}(x)\mathcal{N}(\theta_{i,j}x + \gamma_{i,j}, \delta_{i,j}) \tag{C.4}$$

and set

$$\alpha = \frac{\underline{w}}{\beta^2 + \underline{w}}, \quad \tau = \frac{\beta^2 + \underline{w}}{\beta^2 + \underline{w} + \frac{\beta'^2}{b^2}a^2}. \tag{C.5}$$

The recursions are as follows :

$$\theta_{i,j} = a\alpha\theta_{i+1,j}, \quad \gamma_{i,j} = \gamma_{i+1,j} + (1-\alpha)\theta_{i+1,j}v_{i+1,j},$$
$$\delta_{i,j} = \delta_{i+1,j} + \beta^2\alpha\theta^2_{i+1,j}, \quad v_{i,j} = \tau\frac{Y_i}{b} + a\alpha v_{i+1,j})$$

where the initial values are as previously, except that $w_{i,i} = \underline{w}$. Now the recursions are easily solved. We get :

$$\theta_{i,j} = a(a\alpha)^{j-i}, \qquad \delta_{i,j} = \beta^2(1 + a\sum_{l=0}^{j-i-1}(a\alpha)^{2l+1}), \tag{C.6}$$

$$\gamma_{i,j} = (1-\alpha)a\sum_{l=0}^{j-i-1}(a\alpha)^l v_{j-l,j}, \qquad v_{i,j} = \frac{\tau}{b}\left(\sum_{l=1}^{j-i}(a\alpha)^{j-i-l}Y_{j-l}\right) + (a\alpha)^{j-i}\frac{Y_j}{b}.$$

We finally obtain closed formulae for $\gamma_{i,j}$ and $u_{i,j}$.

By (6.13) we must compute

$$f_{i,k-1}(x) = \frac{\underline{L}_{i,k-1}\mathbf{1}(x)}{\eta_{i|i-1:0}\underline{L}_{i,k-1}\mathbf{1}} = \frac{u_{0,i-1}\phi_{v_{0,i-1},\underline{w}}(x)}{u_{0,i-1}\eta_{i|i-1:0}(\phi_{v_{i,k},\underline{w}})}$$

Hence the term $u_{0,i-1}$ is compensated. Then, we obtain $\eta_{i|i-1:0} = \mathcal{N}(m_{0,i-1}, s^2_{0,i-1})$ with

$$m_{0,i-1} = \gamma_{0,i-1} + \theta_{0,i-1}\frac{\sigma_s^2 v_{0,i-1}}{\sigma_s^2 + v_{0,i-1}} \text{ and } s^2_{0,i-1} = \delta_{0,i-1} + \theta^2_{0,i-1}\frac{\sigma_s^2\underline{w}}{\sigma_s^2 + \underline{w}}.$$

Finally

$$f^2_{i,k-1}(x) = \left(1 + \frac{s^2_{0,i-1}}{\underline{w}}\right)\exp\left(-\frac{(x - v_{i,k-1})^2}{\underline{w}}\right)\exp\left(\frac{(m_{0,i-1} - v_{i,k-1})^2}{s^2_{0,i-1} + \underline{w}}\right). \quad \text{(C.7)}$$

Now we study

$$\left(\eta_{k|k-1:i}f(x) - \eta_{k|k-1:0}f\right)^2.$$

The distributions $\eta^{\delta_x}_{k|k-1:i}$ and $\eta_{k|k-1:0}$ are Gaussian with variances respectively given by $\delta_{i,k-1}$ and $s^2_{0,k-1}$. By (C.6), these variances belong to a compact interval $[k_1, k_2] \subset (0, +\infty)$. By Lemma C.2 of the Appendix, we have

$$|\eta_{k|k-1:i}f(x) - \eta_{k|k-1:0}f| \leq K Z_{i,k} \quad \text{(C.8)}$$

with

$$Z_{i,k} = |\theta_{i,k-1}x - \theta_{0,k-1}\frac{\sigma_s^2 v_{0,k-1}}{\sigma_s^2 + \underline{w}} + \gamma_{i,k-1} - \gamma_{0,k-1}| + |\delta_{i,k-1} - \delta_{0,k-1} - \frac{\sigma_s^2\underline{w}\theta^2_{0,k-1}}{\sigma_s^2 + \underline{w}}|, \quad \text{(C.9)}$$

and K is a constant depending on $\|f\|_\infty$ and k_1, k_2.

Notice that :

$$\begin{aligned}\gamma_{i,k-1} - \gamma_{0,k-1} &= (1-\alpha)a\left(\sum_{l=0}^{k-i-1}(a\alpha)^l v_{k-1-l,k-1} - \sum_{l=0}^{k-1}(a\alpha)^l v_{k-1-l,k-1}\right)\\ &= (1-\alpha)(a\alpha)^{k-1-i}\sum_{l=k-1-i}^{k-1}(a\alpha)^{l-(k-1-i)}v_{k-1-l,k-1}\end{aligned}$$

and

$$\begin{aligned}\delta_{i,k-1} - \delta_{0,k-1} &= \beta^2 a\left(\sum_{l=0}^{k-i-1}(a\alpha)^{2l+1} - \sum_{l=0}^{k-1}(a\alpha)^{2l+1}\right)\\ &= \beta^2 a\sum_{l=k-i}^{k-1}(a\alpha)^{2l+1} \asymp (a\alpha)^{2(k-i)}\end{aligned}$$

In addition, the quantity $\left(1+\frac{s^2_{0,i-1}}{\underline{w}}\right)$ is uniformly bounded. We have

$$\underline{\Delta}_{k|k-1:0}(f) = \sum_{i=0}^{k} \underline{\Delta}_{i,k}$$

with

$$\underline{\Delta}_{i,k} = \eta_{i|i-1,0}(f^2_{i,k-1}(.)(\eta_{k|k-1,i}f(.) - \eta_{k|k-1,0}f)^2).$$

Now we use (C.3) and (C.7)-(C.9) and the above computations to get

$$\underline{\Delta}_{k|k-1:0}(f) \leq M \sum_{i=0}^{k} \alpha^i e^{B_{i,k}} A_{i,k}$$

where M is a constant depending on $\|f\|_\infty$ and $A_{i,k}, B_{i,k}$ are random variables depending on the observations $Y_{0:k}$ through $v_{i,j}$ and $\gamma_{i,j}$. Due to the stationarity, one can see that

$$\sup_{i\leq j} \mathbf{E}|v_{i,j}|^2 < \infty \text{ and } \sup_{i\leq j} \mathbf{E}|\gamma_{i,j}|^2 < \infty$$

where the expectation is taken with respect to the Y_k random variables. This allows to prove $\sup_{i\leq k} \mathbf{E}(A^2_{i,k} + B^2_{i,k}) < \infty$. We conclude the proof with Lemma C.1 (see appendix). Analogously, we prove the tightness of $\underline{\Gamma}_{k|k:0}(f)$.

C.4 Simulations based on a Gaussian AR(1) model

Consider the observations given by

$$Y_i = X_i + \varepsilon_i.$$

where (ε_i) is a sequence of i.i.d. $\mathcal{N}(0, 0.5^2)$ random variables and $X_i = aX_{i-1} + \beta U_i, \quad a = 0.5, \beta = 0.5$. Figure C.1 presents in plain line with square marks a trajectory of the hidden chain, in longdashed line, with diamond marks, the observations. We have plotted in dashed line, with plus marks, the result of the particle filter with $N = 500$ particles and $f(x) = x$. We also observe that the result of the particle filter is close to the hidden chain, uniformly in time.

C.4.1 Inequality between Gaussian distributions

Lemme C.2 *Let f a measurable function and $C > 0$ such that $\forall x \in \mathbb{R}, \quad |f(x)| \leq C$. Then for $m, m' \in \mathbb{R}$ and $\sigma, \sigma' \in [\epsilon, \frac{1}{\epsilon}]$ there is a constant $K > 0$ where K only depends on C and ε, such that*

$$\left|\mathcal{N}(m, \sigma^2)(f) - \mathcal{N}(m', \sigma'^2)(f)\right| \leq K\left(|m - m'| + |\sigma^2 - \sigma'^2|\right).$$

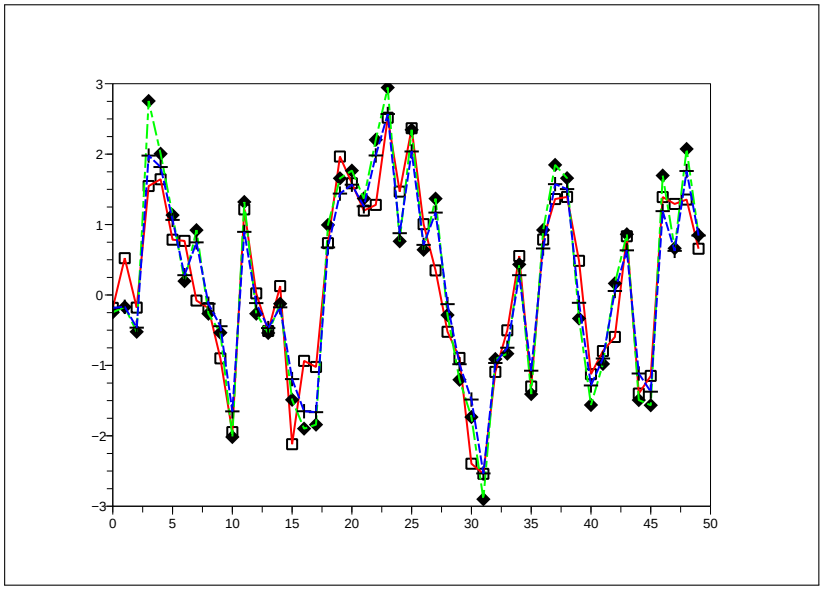

FIGURE C.1 – Kalman model. Hidden Markov chain (plain, square marks). Observations (longdashed line, diamond marks). Particle filter (dashed line, plus marks)

We start with the case $\sigma = \sigma'$:

$$\begin{aligned} & \left| \int_{\mathbb{R}} f(x)(\phi_{m,\sigma^2}(x) - \phi_{m',\sigma^2}) dx \right| \\ \leq \quad & C \int_{\mathbb{R}} \left| \frac{1}{\sqrt{2\pi}\sigma} (\exp(-\frac{1}{2}(\frac{x-m}{\sigma})^2) - \exp(-\frac{1}{2}(\frac{x-m'}{\sigma})^2)) \right| dx. \end{aligned}$$

With the changing of variables $x = \sigma z$

$$\begin{aligned} & \int_{\mathbb{R}} \left| \frac{1}{\sqrt{2\pi}\sigma} (\exp(-\frac{1}{2}(\frac{x-m}{\sigma})^2) - \exp(-\frac{1}{2}(\frac{x-m'}{\sigma})^2)) \right| dx \\ = \quad & \int_{-\infty}^{\frac{m}{\sigma}+\frac{t}{2}} \frac{1}{\sqrt{2\pi}} \left(\exp(-\frac{1}{2}(x - \frac{m}{\sigma})^2) - \exp(-\frac{1}{2}(x - \frac{m}{\sigma} - t)^2) \right) dx \\ & + \int_{\frac{m}{\sigma}+\frac{t}{2}}^{+\infty} \frac{1}{\sqrt{2\pi}} \left(\exp(-\frac{1}{2}(x - \frac{m}{\sigma} - t)^2) - \exp(-\frac{1}{2}(x - \frac{m}{\sigma})^2) \right) dx \end{aligned}$$

where $t = \frac{m'}{\sigma} - \frac{m}{\sigma}$ and we assumed $m' > m$. With m and σ fixed, we consider the former quantity as a function of t : let $g(t) = g_1(t) + g_2(t)$, where $g_1(t) = G_1(t, \frac{m}{\sigma} + \frac{t}{2})$, with G_1 is defined by

$$G_1(t, y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} \left(\exp(-\frac{1}{2}(x - \frac{m}{\sigma})^2) - \exp(-\frac{1}{2}(x - \frac{m}{\sigma} - t)^2) \right) dx.$$

and with similar notations $g_2(t) = G_2(t, \frac{m}{\sigma} + \frac{t}{2})$. Then :

$$\frac{\partial G_1}{\partial y}(t, \frac{m}{\sigma} + \frac{t}{2}) = \frac{\partial G_2}{\partial y}(t, \frac{m}{\sigma} + \frac{t}{2}) = 0$$

and

$$g_1'(t) = \frac{\partial G_1}{\partial t}(t, \frac{m}{\sigma} + \frac{t}{2}) + \frac{1}{2}\frac{\partial G_1}{\partial y}(t, \frac{m}{\sigma} + \frac{t}{2}).$$

Differentiating the two members, it comes

$$\begin{aligned} |g'(t)| &\leq \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} |x - \frac{m}{\sigma} - t| \exp(-\frac{1}{2}(x - \frac{m}{\sigma} - t)^2) dx \\ &\leq \int_{\mathbb{R}} |x| \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2) dx \end{aligned}$$

where the last inequality holds for $t \in [0, \frac{m'-m}{\sigma}]$. By applying the mean value theorem to g on $[0, \frac{m'-m}{\sigma}]$ we have

$$|g(\frac{m'-m}{\sigma}) - g(0)| \leq \left(\int_{\mathbb{R}} |x| \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2) dx \right) |\frac{m'-m}{\sigma}|$$

and then

$$|\mathcal{N}(m, \sigma^2)(f) - \mathcal{N}(m', \sigma^2)(f)| \leq K_\epsilon |m - m'|.$$

Analogous computations work when $m = m'$. With $s = \frac{1}{\sigma}$ et $s' = \frac{1}{\sigma'}$, we have :

$$\begin{aligned} & \left| \int_{\mathbb{R}} f(x) \frac{1}{\sqrt{2\pi}} \left(s \exp(-\frac{1}{2}s^2(x-m)^2) - s' \exp(-\frac{1}{2}s'^2(x-m)^2) \right) dx \right| \\ \leq\ & C \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \left| \left(s \exp(-\frac{1}{2}s^2(x-m)^2) - s' \exp(-\frac{1}{2}s'^2(x-m)^2) \right) \right| dx \\ \leq\ & C \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \left| \left(s \exp(-\frac{1}{2}s^2x^2) - st \exp(-\frac{1}{2}s^2t^2x^2) \right) \right| dx \end{aligned}$$

with $t = \frac{s'}{s}$, assuming $s' > s$. The sign changes occur at :

$$x = \pm \sqrt{\frac{2 \log t}{s^2(t^2 - 1)}}.$$

Thus, we have

$$\begin{aligned}
&\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \left| \left(s\exp(-\frac{1}{2}s^2x^2) - st\exp(-\frac{1}{2}s^2t^2x^2) \right) \right| dx \\
\leq\ &\int_{-\infty}^{-\sqrt{\frac{2\log t}{s^2(t^2-1)}}} \frac{1}{\sqrt{2\pi}} \left(s\exp(-\frac{1}{2}s^2x^2) - st\exp(-\frac{1}{2}s^2t^2x^2) \right) dx \\
&+ \int_{-\sqrt{\frac{2\log t}{s^2(t^2-1)}}}^{+\sqrt{\frac{2\log t}{s^2(t^2-1)}}} \frac{1}{\sqrt{2\pi}} \left(st\exp(-\frac{1}{2}s^2t^2x^2) - s\exp(-\frac{1}{2}s^2x^2) - \right) dx \\
&+ \int_{+\sqrt{\frac{2\log t}{s^2(t^2-1)}}}^{+\infty} \frac{1}{\sqrt{2\pi}} \left(s\exp(-\frac{1}{2}s^2x^2) - st\exp(-\frac{1}{2}s^2t^2x^2) \right) dx.
\end{aligned}$$

Proceeding as above, we derive :

$$\left| \mathcal{N}(m,\sigma^2)(f) - \mathcal{N}(m,\sigma'^2)(f) \right| \leq K_\epsilon |\sigma^2 - \sigma'^2|$$

remarking that $\sigma, \sigma^2 \in [\epsilon, \frac{1}{\epsilon}]$ and $|\frac{1}{\sigma} - \frac{1}{\sigma'}| \leq C_\epsilon |\sigma^2 - \sigma'^2|$.

□

Conclusion et perspectives

Dans cette thèse, plusieurs problèmes d'estimation paramétrique pour une diffusion (X_t) discrétisée et bruitée ont été étudiés, dans des contextes asymptotiques différents. On a essentiellement considéré des observations de la forme

$$Y_{i\delta} = HX_{i\delta} + \rho\varepsilon_{i\delta}$$

avec $H = 1$ dans le cas unidimensionnel, ou plus généralement H une forme linéaire dans le cas d'observations partielles d'une diffusion multidimensionnelle.

L'un des prolongements possibles de la deuxième partie est d'envisager une structure des observations plus générale que celle du bruit additif : en considérant une diffusion (X_t) à valeurs dans un intervalle (l, r), un noyau $F(x, dy)$ tel que $Y_{i\delta}$ soit distribué conditionnellement à la trajectoire (X_t) selon $F(X_{i\delta}, dx)$, avec $\mathbb{E}(Y_{i\delta}|(X_t)) = X_{i\delta}$, la méthode basée sur la construction de moyennes locales peut permettre de traiter un modèle plus général, comme par exemple celui basé sur la diffusion de Wright et Fisher décrit dans Chaleyat-Maurel and Genon-Catalot (2009).

Une autre perspective envisageable est de construire un contraste pour l'estimation des paramètres de la diffusion (X_t), lorsque les observations sont données par $Z_{i\delta} = \int_{i\delta}^{(i+1)\delta} dY_t$, où

$$dY_t = X_t dt + \rho dV_t$$

avec (V_t) un mouvement brownien indépendant de (X_t). Le cas où ρ tend suffisamment vite vers 0 peut être déjà envisagé à l'aide des moyennes locales, et il semble intéressant de pouvoir considérer d'autres formes de réduction du bruit, comme la méthode de Zhang et al. (2005).

Bibliographie

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions : a closed-form approximation approach. *Econometrica*, 70(1) :223–262.

Atar, R. and Zeitouni, O. (1997). Exponential stability for nonlinear filtering. *Ann. Inst. H. Poincaré Probab. Statist.*, 33(6) :697–725.

Baltazar-Larios, F. and Sørensen, M. (2009). Maximum likelihood estimation for integrated diffusion processes. *Preprint*, 1(1).

Balvay, D., Tropres, I., Billet, R., Joubert, A., Peoc'h, M., Cuenod, C., and Le Duc, G. (2009). Mapping the zonal organization of tumor perfusion and permeability in a rat glioma model by using dynamic contrast-enhanced synchrotron radiation CT. *Radiology*, 250 :692–702.

Beskos, A., Papaspiliopoulos, O., and Roberts, G. O. (2006). Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12(6) :1077–1098.

Bibby, B. M. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, 1(1-2) :17–39.

Bisdas, S., Konstantinou, G.-N., Lee, P.-S., Thng, C.-H., Wagenblast, J., Baghi, M., and Koh, T.-S. (2007). Dynamic contrast-enhanced CT of head and neck tumors : perfusion measurements using a distributed-parameter tracer kinetic model. Initial results and comparison with deconvolution-based analysis. *Phys Med Biol*, 52(20) :6181–96.

Brix, G., Kiessling, F., Lucht, R., Darai, S., Wasser, K., Delorme, S., and Griebel, J. (2004). Microcirculation and microvasculature in breast tumors : pharmacokinetic analysis of dynamic MR image series. *Magn Reson Med.*, 52-2 :420–9.

Brochot, C., Bessoud, B., Balvay, D., Cuenod, C., Siauve, N., and Bois, F. (2006). Evaluation of antiangiogenic treatment effects on tumors' microcirculation by

bayesian physiological pharmacokinetic modeling and magnetic resonance imaging. *Magn Reson Imaging.*, 24 :1059–1067.

Brockwell, P. and Davis, R. (1991). *Time Series : Theory and Methods*. Springer.

Buadu, L., Murakami, J., Murayama, S., Hashiguchi, N., Sakai, S., Masuda, K., Toyoshima, S., Kuroki, S., and Ohno, S. (1996). Breast lesions : correlation of contrast medium enhancement patterns on MR images with histopathologic findings and tumor angiogenesis. *Radiology*, 200 :639–649.

Buckley, D. (2002). Uncertainty in the analysis of tracer kinetics using dynamic contrast-enhanced T1-weighted MRI. *Magn Reson Med*, 47 :601–606.

Buonaccorsi, G., O'Connor, J., Caunce, A., Roberts, C., Cheung, S., Watson, Y., Davies, K., Hope, L., Jackson, A., Jayson, G., and Parker, G. (2007). Tracer kinetic model-driven registration for dynamic contrast-enhanced MRI time-series data. *Magn Reson Med*, 58 :1010–1019.

Calamante, F., Gadian, D., and Connelly, A. (2003). Quantification of bolus-tracking MRI : Improved characterization of the tissue residue function using Tikhonov regularization. *Magn Reson Med*, 50 :1237–1247.

Canet, E., Revel, D., Forrat, R., Baldy-Porcher, C., de Lorgeril, M., Sebbag, L., Vallee, J., Didier, D., and Amiel, M. (1993). Superparamagnetic iron oxide particles and positive enhancement for myocardial perfusion studies assessed by subsecond T1-weighted MRI. *Magn Reson Imaging*, 11 :1139–1145.

Cappé, O., Moulines, E., and Rydén, T. (2005). *Inference in hidden Markov models*. Springer Series in Statistics. Springer, New York. With Randal Douc's contributions to Chapter 9 and Christian P. Robert's to Chapters 6, 7 and 13, With Chapter 14 by Gersende Fort, Philippe Soulier and Moulines, and Chapter 15 by Stéphane Boucheron and Elisabeth Gassiat.

Chaleyat-Maurel, M. and Genon-Catalot, V. (2006). Computable infinite-dimensional filters with applications to discretized diffusion processes. *Stochastic Process. Appl.*, 116(10) :1447–1467.

Chaleyat-Maurel, M. and Genon-Catalot, V. (2009). Filtering the Wright-Fisher diffusion. *ESAIM Probab. Stat.*, 13 :197–217.

Cheng, H. (2008). Investigation and optimization of parameter accuracy in dynamic contrast-enhanced MRI. *J Magn Reson Imaging*, 28 :736–743.

Chopin, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.*, 32(6) :2385–2411.

Cuenod, C., Leconte, I., Siauve, N., Resten, A., Dromain, C., Poulet, B., Frouin, F., Clement, O., and Frija, G. (2001). Early changes in liver perfusion caused by occult metastases in rats : detection with quantitative CT. *Radiology*, 218 :556–561.

Davies, E. B. (1989). *Heat kernels and spectral theory*, volume 92 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge.

de Bazelaire, C., Siauve, N., Fournier, L., Frouin, F., Robert, P., Clement, O., de Kerviler, E., and Cuenod, C. (2005). Comprehensive model for simultaneous MRI determination of perfusion and permeability using a blood-pool agent in rats rhabdomyosarcoma. *Eur Radiol*, 15 :2497–505.

Del Moral, P. (2004). *Feynman-Kac formulae*. Probability and its Applications (New York). Springer-Verlag, New York. Genealogical and interacting particle systems with applications.

Del Moral, P. and Guionnet, A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. H. Poincaré Probab. Statist.*, 37(2) :155–194.

Del Moral, P. and Jacod, J. (2001a). Interacting particle filtering with discrete observations. In *Sequential Monte Carlo methods in practice*, Stat. Eng. Inf. Sci., pages 43–75. Springer, New York.

Del Moral, P. and Jacod, J. (2001b). Interacting particle filtering with discrete-time observations : asymptotic behaviour in the Gaussian case. In *Stochastics in finite and infinite dimensions*, Trends Math., pages 101–122. Birkhäuser Boston, Boston, MA.

Del Moral, P., Jacod, J., and Protter, P. (2001). The Monte-Carlo method for filtering with discrete-time observations. *Probab. Theory Related Fields*, 120(3) :346–368.

Delzescaux, T., Frouin, F., De Cesare, A., Philipp-Foliguet, S., Zeboudj, R., Janier, M., Todd-Pokropek, A., and Herment, A. (2001). Adaptive and self-evaluating registration method for myocardial perfusion assessment. *Magma*, 13 :28–39.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Jr. R. Stat. Soc. B*, 39 :1–38.

Ditlevsen, S. and De Gaetano, A. (2005). Stochastic vs. deterministic uptake of dodecanedioic acid by isolated rat livers. *Bull. Math. Biol.*, 67 :547–561.

Ditlevsen, S. and Ditlevsen, O. (2007). Parameter estimation from observations of first-passage times of the Ornstein-Uhlenbeck process and the Feller process. *Probabilistic Engineering Mechanics*, 23(1).

Ditlevsen, S. and Lansky, P. (2005). Estimation of the input parameters in the ornstein-uhlenbeck neuronal model. *Phys. Rev. E*, 71(1) :011907.

Ditlevsen, S., Yip, K., and Holstein-Rathlou, N. (2005). Parameter estimation in a stochastic model of the tubuloglomerular feedback mechanism in a rat nephron. *Math Biosc*, 194 :49–69.

Donnet, S. and Samson, A. (2008). Parametric inference for mixed models defined by stochastic differential equations. *ESAIM Probability & Statistics*, 12 :196–218.

Douc, R., Fort, G., Moulines, E., and Priouret, P. (2009). Forgetting of the initial distribution for Hidden Markov Models. *Stochastic Process. Appl.*, 119(4) :1235–1256.

Douc, R., Guillin, A., and Najim, J. (2005). Moderate deviations for particle filtering. *Ann. Appl. Probab.*, 15(1B) :587–614.

Doucet, A., de Freitas, N., and Gordon, N., editors (2001). *Sequential Monte Carlo methods in practice*. Statistics for Engineering and Information Science. Springer-Verlag, New York.

Fan, X., Medved, M., Karczmar, G., Yang, C., Foxley, S., Arkani, S., Recant, W., Zamora, M., Abe, H., and Newstead, G. (2007). Diagnosis of suspicious breast lesions using an empirical mathematical model for dynamic contrast-enhanced MRI. *Magn Reson Imaging*, 25 :593–603.

Favetto, B. (2009). On the asymptotic variance in the Central Limit Theorem for particle filters. Preprint MAP5 2009-14, to appear in ESAIM P & S.

Favetto, B. and Samson, A. (2010). Parameter estimation for a bidimensional partially observed Ornstein-Uhlenbeck process with biological application. *Scand. J. Statist.*, 37(2) :200–220.

Favetto, B., Samson, A., Balvay, D., Thomassin, I., Genon-Catalot, V., Cuenod, C.-A., and Yves, R. (2009). Blood micro-circulation parameters extraction from Dynamic Contrast Enhanced MRI data using stochastic differential equations . Submitted.

Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics*, 20(4) :547–557.

Fournier, L., Thiam, R., Cuénod, C.-A., Medioni, J., Trinquart, L., Balvay, D., Banu, E., Balcaceres, J., Frija, G., and Oudard, S. (2007). Dynamic contrast-enhanced CT (DCE-CT) as an early biomarker of response in metastatic renal cell carcinoma (mRCC) under anti-angiogenic treatment. *J. of Clinical Oncology - ASCO Annual Meeting Proceedings (Post-Meeting Edition)*, 25.

Fritz-Hansen, T., Rostrup, E., Sondergaard, L., Ring, P., Amtorp, O., and Larsson, H. (1998). Capillary transfer constant of Gd-DTPA in the myocardium at rest and during vasodilation assessed by MRI. *Magn Reson Med*, 40 :922–929.

Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion processes from discrete observations. *Statistics*, 21(1) :99–116.

Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 29(1) :119–151.

Genon-Catalot, V., Jeantheau, T., and Laredo, C. (2003). Conditional likelihood estimators for hidden Markov models and stochastic volatility models. *Scand. J. Statist.*, 30(2) :297–316.

Genon-Catalot, V. and Laredo, C. (2006). Leroux's method for general hidden Markov models. *Stochastic Process. Appl.*, 116(2) :222–243.

Gloter, A. (2000). Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM Probab. Statist.*, 4 :205–227 (electronic).

Gloter, A. (2006). Parameter estimation for a discretely observed integrated diffusion process. *Scand. J. Statist.*, 33(1) :83–104.

Gloter, A. and Jacod, J. (2001). Diffusions with measurement errors. II. Optimal estimators. *ESAIM Probab. Statist.*, 5 :243–260 (electronic).

Goh, V., Halligan, S., Hugill, J., Gartner, L., and Bartram, C. (2005). Quantitative colorectal cancer perfusion measurement using dynamic contrast-enhanced multidetector-row computed tomography : effect of acquisition time and implications for protocols. *J. Comput. Assist. Tomogr.*, 29 :59–63.

Goh, V., Padhani, A. R., and Rasheed, S. (2007). Functional imaging of colorectal cancer angiogenesis. *Lancet Oncol.*, 8(3) :245–55.

Guo, W., Wang, Y., and B., M. B. (1999). A signal extraction approach to modeling hormone time series with pulses and a changing baseline. *Journal of the American Statistical Association*, 94(447) :746–756.

Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York. Probability and Mathematical Statistics.

Höpfner, R. and Brodda, K. (2006). A stochastic model and a functional central limit theorem for information processing in large systems of neurons. *J. Math. Biol.*, 52(4) :439–457.

Ibragimov, I. A. and Has′minskiĭ, R. Z. (1981). *Statistical estimation*, volume 16 of *Applications of Mathematics*. Springer-Verlag, New York. Asymptotic theory, Translated from the Russian by Samuel Kotz.

Idoux, E., Serafin, M., Fort, P., Vidal, P., Beraneck, M., Vibert, N., MÃ$\frac{1}{4}$hlethaler, M., and Moore, L. (2006). Oscillatory and intrinsic membrane properties of guinea pig nucleus prepositus hypoglossi neurons in vitro. *J Neurophysiol*, 96(1) :175–196.

Jacobsen, M. (2002). Optimality and small Δ-optimality of martingale estimating functions. *Bernoulli*, 8(5) :643–668.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009). Microstructure noise in the continuous case : the pre-averaging approach. *Stochastic Process. Appl.*, 119(7) :2249–2276.

Jain, R., Scarpace, L., Ellika, S., Schultz, L., Rock, J., Rosenblum, M., Patel, S., Lee, T., and Mikkelsen, T. (2007). First-pass perfusion computed tomography : initial experience in differentiating recurrent brain tumors from radiation effects and radiation necrosis. *Neurosurgery*, 61 :778–86.

Kelcz, F., Furman-Haran, E., Grobgeld, D., and Degani, H. (2002). Clinical testing of high-spatial-resolution parametric contrast-enhanced MR imaging of the breast. *AJR Am J Roentgenol*, 179 :1485–1492.

Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.*, 24(2) :211–229.

Kessler, M. (2000). Simple and explicit estimating functions for a discretely observed diffusion process. *Scand. J. Statist.*, 27(1) :65–82.

Kessler, M. and Sørensen, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, 5(2) :299–314.

Kiessling, F., Greschus, S., Lichy, M., Bock, M., Fink, C., Vosseler, S., Moll, J., Mueller, M., Fusenig, N., Traupe, H., and Semmler, W. (2004). Volumetric computed tomography (VCT) : a new technology for noninvasive, high-resolution monitoring of tumor angiogenesis. *Nat. Med.*, 10 :1133–8.

Krishnamurthi, G., Stantz, K.-M., Steinmetz, R., Gattone, V.-H., Minsong, C., Hutchins, G.-D., and Yun, L. (2005). Functional imaging in small animals using X-ray computed tomography : study of physiologic measurement reproducibility. *IEEE Trans. on Medical Imaging*, 24(7) :832–43.

Künsch, H. R. (2001). State space and hidden Markov models. In *Complex stochastic systems (Eindhoven, 1999)*, volume 87 of *Monogr. Statist. Appl. Probab.*, pages 109–173. Chapman & Hall/CRC, Boca Raton, FL.

Künsch, H. R. (2005). Recursive Monte Carlo filters : algorithms and theoretical analysis. *Ann. Statist.*, 33(5) :1983–2021.

Liptser, R. S. and Shiryaev, A. N. (2001). *Statistics of random processes. I*, volume 5 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, expanded edition. General theory, Translated from the 1974 Russian original by A. B. Aries, Stochastic Modelling and Applied Probability.

Lucht, R., Delorme, S., Hei, J., Knopp, M., Weber, M., Griebel, J., and Brix, G. (2005). Classification of signal-time curves obtained by dynamic magnetic resonance mammography : statistical comparison of quantitative methods. *Invest Radiol*, 40 :442–447.

Materne, R., Smith, A., Peeters, F., Dehoux, J., Keyeux, A., Horsmans, Y., and Van Beers, B. (2002). Assessment of hepatic perfusion parameters with dynamic MRI. *Magn Reson Med*, 47 :135–142.

Miles, K. A. (2003). Functional CT imaging in oncology. *Eur. Radiol.*, 13(5) :134–8.

Nasel, C., Azizi, A., Veintimilla, A., Mallek, R., and Schindler, E. (2000). A standardized method of generating time-to-peak perfusion maps in dynamic-susceptibility contrast-enhanced MR imaging. *AJNR Am J Neuroradiol*, 21 :1195–1198.

Ostergaard, L. (2005). Principles of cerebral perfusion imaging by bolus tracking. *J Magn Reson Imaging*, 22 :710–717.

Oudjane, N. and Rubenthaler, S. (2005). Stability and uniform particle approximation of nonlinear filters in case of non ergodic signals. *Stoch. Anal. Appl.*, 23(3) :421–448.

Padhani, A. R., Harvey, C. J., and Cosgrove, D. (2005). Angiogenesis imaging in the management of prostate cancer. *Nat. Clin. Pract. Urol.*, 2(12) :596–607.

Pedersen, A. (1994). Statistical analysis of gaussian diffusion processes based on incomplete discrete observations. *Research Report, Department of Theoretical Statistics, University of Aarhus*, 297.

Pedersen, A. R. (1995a). Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli*, 1(3) :257–279.

Pedersen, A. R. (1995b). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Statist.*, 22(1) :55–71.

Picchini, U., Ditlevsen, S., and De Gaetano, A. (2006). Modeling the euglycemic hyperinsulinemic clamp by stochastic differential equations. *J. Math. Biol.*, 53 :771–796.

Picchini, U., Ditlevsen, S., and De Gaetano, A. (2008). Maximum likelihood estimation of a time-inhomogeneous stochastic differential model of glucose dynamics. *Mathematical Medicine and Biology*, 25 :141–155.

Port, R., Knopp, M., and Brix, G. (2001). Dynamic contrast-enhanced MRI using Gd-DTPA : interindividual variability of the arterial input function and consequences for the assessment of kinetics in tumors. *Magn Reson Med*, 45 :1030–1038.

Pradel, C., Siauve, N., Bruneteau, G., Clement, O., de Bazelaire, C., Frouin, F., Wedge, S., Tessier, J., Robert, P., Frija, G., and Cuenod, C. (2003). Reduced capillary perfusion and permeability in human tumour xenografts treated with

the VEGF signalling inhibitor ZD4190 : an in vivo assessment using dynamic MR imaging and macromolecular contrast media. *Magn Reson Imaging*, 21 :845–851.

Robert, C. P. and Casella, G. (2004). *Monte Carlo statistical methods*. Springer Texts in Statistics. Springer-Verlag, New York, second edition.

Rogers, L. C. G. and Williams, D. (2000). *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge. Itô calculus, Reprint of the second (1994) edition.

Rosen, M. and Schnall, M. (2007). Dynamic contrast-enhanced magnetic resonance imaging for assessing tumor vascularity and vascular effects of targeted therapies in renal cell carcinoma. *Clin Cancer Res.*, 13(2) :770–6.

Rozenholc, Y., Reiß, M., Balvay, D., and Cuenod, C.-A. (2009). Growing time-homogeneous neighbourhoods for denoising and clustering dynamic contrast enhanced-CT sequences. *Université Paris Descartes, Prepublication MAP5 UMR CNRS 8145*, nÂ° 2009-.

Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of econometrics*, 63 :289–306.

Segal, M. and Weinstein, E. (1989). A new method for evaluating the log-likelihood gradient, the hessian, and the fisher information matrix for linear dynamic systems. *IEEE Trans. Inf. Theory*, 35 :682–687.

Shames, D., Kuwatsuru, R., Vexler, V., Muhler, A., and Brasch, R. (1993). Measurement of capillary permeability to macromolecules by dynamic magnetic resonance imaging : a quantitative noninvasive technique. *Magn Reson Med*, 29 :616–622.

Shumway, R. and Stoffer, D. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Series Anal*, 3 :253–264.

Sørensen, A., Tievsky, A., Ostergaard, L., Weisskoff, R., and Rosen, B. (1997). Contrast agents in functional MR imaging. *J Magn Reson Imaging*, 7(1) :47–55.

Sørensen, M. (2009). Efficient estimation for ergodic diffusions sampled at high frequency. Preprint.

Sørensen, M. (2010). Estimating functions for diffusion-type processes. Preprint.

Stoer, J. and Bulirsch, R. (1993). *Introduction to numerical analysis*, volume 12 of *Texts in Applied Mathematics*. Springer-Verlag, New York, second edition.

Thomassin, I. (2008). Etude des tumeurs annexielles du pelvis feminin en IRM fonctionnelle, mise au point des techniques et applications cliniques. *Physical PhD, Ecole doctorale Sciences et Technologies de l'Information des Télécomunication et des Systèmes*.

Thomassin-Naggara, I., Balvay, D., Darai, E., Bazot, M., and Cuenod, C. (in press). Dynamic contrast enhanced MR imaging to assess physiologic variations of myometrial perfusion. *in press in European Radiology*.

Thomassin-Naggara, I., Bazot, M., Darai, E., Callard, P., Thomassin, J., and Cuenod, C. (2008). Epithelial ovarian tumors : value of dynamic contrast-enhanced MR imaging and correlation with tumor angiogenesis. *Radiology*, 248 :148–159.

Tofts, P.-S. (1997). Modelling tracer kinetics in dynamic Gd-DTPA MR imaging. *J Magn. Reson. Imaging*, 7(1) :91–101.

Tse, G., Chaiwun, B., Wong, K., Yeung, D., Pang, A., Tang, A., and Cheung, H. (2007). Magnetic resonance imaging of breast lesions–a pathologic correlation. *Breast Cancer Res Treat*, 103 :1–10.

Vallee, J., Sostman, H., MacFall, J., Wheeler, T., Hedlund, L., Spritzer, C., and Coleman, R. (1997). MRI quantitative myocardial perfusion with compartmental analysis : a rest and stress study. *Magn Reson Med*, 38 :981–989.

Van Handel, R. (2009). Uniform time average consistency of Monte Carlo particle filters. *Stochastic Process. Appl.*, 119(11) :3835–3861.

Wintermark, M., Flanders, A., Velthuis, B., Meuli, R., van Leeuwen, M., Goldsher, D., Pineda, C., Serena, J., van der Schaaf, I., Waaijer, A., Anderson, J., Nesbit, G., Gabriely, I., Medina, V., Quiles, A., Pohlman, S., Quist, M., Schnyder, P., Bogousslavsky, J., Dillon, W., and Pedraza, S. (2006). Perfusion-CT assessment of infarct core and penumbra : receiver operating characteristic curve analysis in 130 patients suspected of acute hemispheric stroke. *Stroke*, 37 :979–985.

Wu, C. (1983). On the convergence property of the EM algorithm. *Ann. Statist.*, 11 :95–103.

Yoshida, N. (1992). Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.*, 41(2) :220–242.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005). A tale of two time scales : determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.*, 100(472) :1394–1411.

Zhu, A., Holalkere, N., Muzikansky, A., Horgan, K., and Sahani, D. (2008). Early antiangiogenic activity of bevacizumab evaluated by computed tomography perfusion scan in patients with advanced hepatocellular carcinoma. *Oncologist*, 13 :120–5.