

Méthodes de Monte Carlo
et
Chaînes de Markov pour la simulation

Mémoire présenté pour l'obtention de l'Habilitation à Diriger les Recherches

Gersende FORT

Novembre 2009

Table des matières

Articles présentés dans ce mémoire	6
Introduction	7
Summary	10
Notations	16
1 Méthodes de Monte Carlo	17
1.1 Exemples de méthodes de réduction de variance	18
1.1.1 Méthode de stratification	19
1.1.2 Echantillonnage d'importance	20
Exemple de méthode adaptative : Mixture-Population Monte Carlo (M-PMC)	20
1.2 Exemples de méthodes MCMC	21
1.2.1 Echantillonneurs de Hastings-Metropolis à marche aléatoire symétrique (SRWM)	21
Exemple d'algorithme adaptatif: <i>Adaptive Metropolis</i>	22
1.2.2 Echantillonneur Metropolis-within-Gibbs	23
1.2.3 Echantillonneur <i>Metropolis-Adjusted Langevin Algorithm</i> (MALA)	23
Choix du pas en fonction de l'état courant	23
1.2.4 Exemple d'algorithme MCMC à adaptation externe: l'échantillonneur <i>Equi-Energy</i>	24
2 Ergodicité des processus de Markov	27
2.1 Critères de V -ergodicité sous-géométrique	29
2.1.1 Condition de drift sur le noyau de transition P	30
2.1.2 Condition de drift sur un itéré du noyau de transition P	32
2.1.3 Condition sur le modèle fluide d'une chaîne de Markov	33
Définition du modèle fluide	34
Existence du modèle fluide	35
Stabilité du modèle fluide et ergodicité des chaînes de Markov	36
Conditions suffisantes pour la stabilité du modèle fluide	36
Conclusion	37
2.1.4 Condition de drift sur le générateur d'un processus de Markov fort	38
2.2 Contrôle explicite de V -ergodicité	39
2.2.1 Contrôle explicite par couplage	40
2.2.2 Contrôle explicite d'ergodicité géométrique	41
2.2.3 Contrôle explicite d'ergodicité sous-géométrique	42
2.3 Applications : vitesse de convergence d'algorithmes MCMC	44

2.3.1	Ergodicité sous-géométrique de l'algorithme Hastings-Metropolis à marche aléatoire symétrique	44
2.3.2	Ergodicité sous-géométrique des diffusions de Langevin - Algorithmes de simulation MALA	46
3	Optimisation des méthodes de simulation	51
3.1	Etude de la phase transiente d'algorithmes MCMC	52
3.1.1	Caractérisation des limites fluides d'un algorithme de Hastings-Metropolis	54
	Limites fluides pour des densités sur-exponentielles	55
	Limites fluides pour des densités sous-exponentielles	58
3.1.2	Limites fluides d'un algorithme Metropolis-within-Gibbs	61
3.1.3	Conclusion	64
3.2	Convergence d'algorithmes adaptatifs à dynamique markovienne	65
3.2.1	Ergodicité de l'échantillonneur adaptatif	66
3.2.2	Loi forte des grands nombres	68
3.2.3	Application à la convergence d'un algorithme de Hastings-Metropolis adaptatif	70
3.2.4	Application à la convergence de l'échantillonneur <i>Equi-Energy</i>	71
3.3	Méthode de stratification adaptative	73
3.3.1	Variance asymptotique de l'estimateur stratifié	74
3.3.2	Algorithme de stratification adaptative	75
3.3.3	Application à l'ingénierie financière	76
4	Application des méthodes de simulation à l'analyse statistique de données complexes	83
4.1	Combiner méthodes de Monte Carlo et méthodes de champ moyen pour l'inférence dans des champs de Markov cachés	84
4.1.1	Modèle de champ de Markov caché pour la segmentation d'images	85
4.1.2	Algorithme Monte Carlo VEM (MCVEM)	86
4.1.3	Application du MCVEM à la segmentation d'images	87
4.2	Estimation de paramètres cosmologiques	91
4.2.1	M-PMC sur données simulées	92
4.2.2	Comparaison de M-PMC et de l'échantillonneur <i>Adaptive Metropolis</i> sur données simulées	94
4.2.3	M-PMC pour l'estimation de paramètres cosmologiques	96
Bibliographie		101
Liste des figures		108
Liste des tableaux		111
A Récapitulatif des activités		115
A.1	Curriculum Vitae	115
A.2	Encadrement, animation et administration de la recherche	117

A.2.1	Responsabilités dans l'animation de programmes et projets	117
A.2.2	Autres responsabilités	117
A.3	Diffusion de la culture scientifique	117
A.3.1	Participation à l'organisation de conférences, workshops, numéros spéciaux de revues	117
A.3.2	Organisation de séminaires	117
A.4	Enseignement, formation	117
A.4.1	Enseignement (à partir de 2001)	117
A.4.2	Thèses co-dirigées	118
A.4.3	Participation à des jurys de thèse	118
A.5	Transfert technologique, Relations industrielles et valorisation	118
A.5.1	Relations Industrielles	118
A.6	Production scientifique	118
A.6.1	Revues avec comité de lecture	118
A.6.2	Chapitres de livre	119
A.6.3	Actes de conférences internationales avec comité de lecture	119
A.6.4	Rapports de Recherche	120
A.6.5	Thèse de Doctorat	120
A.6.6	Conférences invitées	120
A.6.7	Exposés dans des conférences ou séminaires (à partir de 2001)	121

Articles présentés dans ce mémoire

1. C. Andrieu and G. Fort. Explicit control of subgeometric ergodicity. Technical Report 05:17, University of Bristol, UK, 2005.
2. Y. Atchadé and G. Fort. Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. *Bernoulli*, 2009. Accepted for publication.
3. Y. Atchadé, G. Fort, E. Moulines, and P. Priouret. Adaptive Markov chain Monte Carlo : Theory and Methods. Technical report, 2009.
4. S. Connor and G. Fort. State-dependent Foster-Lyapunov criteria for subgeometric convergence of Markov chains. *Stoch. Process. Appl.*, 119:4176–4193, 2009.
5. R. Douc, G. Fort, and A. Guillin. Subgeometric rates of convergence of f-ergodic strong Markov processes. *Stoch. Process Appl.*, 119(3):897–923, 2009.
6. R. Douc, G. Fort, E. Moulines, and P. Soulier. Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.*, 14(3):1353–1377, 2004.
7. P. Etoré, G. Fort, B. Jourdain, and E. Moulines. On adaptive stratification. *Annals of Operations Research*, 2009. Accepted for publication.
8. F. Forbes and G. Fort. Combining Monte Carlo and Mean-Field-like methods for inference in Hidden Markov Random Fields. *IEEE Transactions on Image Processing*, 16(3):824–837, 2007.
9. G. Fort. Computable bounds for V-geometric ergodicity of Markov transition kernels. Technical Report RR 1047-M, Univ. J. Fourier, France, 2002.
10. G. Fort. Fluid limit-based tuning of some hybrid MCMC samplers. Technical report, 2007.
11. G. Fort, S. Meyn, E. Moulines, and P. Priouret. The ODE method for the stability of skip-free Markov Chains with applications to MCMC. *Ann. Appl. Probab.*, 18(2):664–707, 2008.
12. G. Fort and G.O. Roberts. Subgeometric ergodicity of strong Markov processes. *Ann. Appl. Probab.*, 15(2):1565–1589, 2005.
13. D. Wraith, M. Kilbinger, K. Benabed, O. Cappé, J.F. Cardoso, G. Fort, S. Prunet, and C.P. Robert. Estimation of cosmological parameters using adaptive importance sampling. *Phys. Rev. D*, 80(2):18 pages, 2009.

Introduction

Les méthodes de Monte Carlo sont des méthodes d'intégration numérique qui utilisent l'aléatoire. Elles permettent de résoudre de nombreux problèmes autrement insolubles comme par exemple l'évaluation d'intégrales sur des domaines complexes et/ou de grande dimension, le calcul de fonctionnelles de processus stochastiques ou l'exploration de densités de probabilité complexes. Les grandes avancées méthodologiques dans le domaine des méthodes de Monte Carlo sont souvent dues aux physiciens ([82, 78, 115] par exemple). Néanmoins, ces propositions ne couvrent pas tous les domaines d'application : la recherche méthodologique autour des méthodes de simulation est donc un axe de recherche d'actualité en prise avec de nombreux domaines scientifiques (physique statistique, ingénierie, statistiques computationnelles, finance, ...)

Les algorithmes de Monte Carlo dépendent de paramètres d'implémentation dont le choix est crucial pour la convergence de l'algorithme et pour son efficacité. Par exemple, dans les algorithmes de type échantillonnage d'importance, les tirages se font sous une loi instrumentale et ces points sont ensuite corrigés - par un mécanisme de pondération - pour approcher une loi cible. Dans ce cas, la loi instrumentale est un paramètre d'implémentation. Dans les algorithmes de type Monte Carlo par chaînes de Markov (MCMC), il s'agit de construire une chaîne de Markov stationnaire, de mesure invariante égale à la loi cible. Du choix du noyau de transition dépend les propriétés d'ergodicité de la chaîne et donc l'efficacité de l'échantillonneur : dans cet exemple, le choix du noyau de transition est un paramètre d'implémentation.

Choisir des paramètres appropriés est un problème difficile, qui demande une certaine expertise de l'utilisateur. En dépit de la facilité d'implémentation des algorithmes de simulation, cet aspect de la mise en oeuvre est un frein à la large diffusion des méthodes de Monte Carlo. En conséquence, les nouveaux algorithmes de simulation s'orientent vers des procédures adaptatives, c'est-à-dire des procédures itératives au cours desquelles l'algorithme corrige la valeur des paramètres d'implémentation en fonction de son comportement passé. La procédure d'adaptation est guidée par la recherche des paramètres d'implémentation qui optimisent un critère d'efficacité de l'échantillonneur.

Proposer une méthodologie, si possible assez générale pour avoir un domaine d'application très large, est une chose. Vérifier que l'algorithme fait ce pour quoi il a été développé en est une autre. L'étude de la convergence des méthodes de simulation adaptatives s'avère plus complexe que celle des échantillonneurs non-adaptatifs. Par exemple, les tirages issus d'une procédure MCMC adaptative ne sont plus la réalisation d'une chaîne de Markov homogène, et l'étude de ces algorithmes ne relève plus seulement de résultats sur l'ergodicité d'un noyau de transition. La recherche sur les méthodes de simulation doit donc aussi permettre d'aider à la compréhension fine du fonctionnement des échantillonneurs pour identifier les paramètres d'implémentation optimaux; et permettre d'aider

à l'analyse du comportement des algorithmes proposés.

Mes travaux s'inscrivent pour l'essentiel dans le cadre de la recherche sur les méthodes de Monte Carlo: ils sont motivés par une compréhension fine de la dynamique des échantillonneurs et de l'influence des paramètres d'implémentation sur cette dynamique. Du fait des orientations prises durant ma thèse, ces travaux comportent une forte composante "markovienne" développée tout d'abord dans le but d'étudier les algorithmes MCMC, mais qui maintenant se révèle aussi nécessaire pour l'étude des algorithmes MCMC adaptatifs ou plus généralement des algorithmes de simulation adaptatifs à dynamique markovienne ([74, 115] par exemple). Je m'intéresse aussi à des aspects méthodologiques et applicatifs, aspects que j'ai développés principalement dans le cadre de deux projets ANR (projets ADAP'MC et ECOSSTAT).

Ce document d'habilitation ne fait pas le point sur toutes mes activités de recherche depuis la soutenance de ma thèse - en juin 2001 - mais plutôt sur les thématiques de recherche que j'avais mises en avant dans les projets de recherche sur lesquels j'ai été recrutée au CNRS - à l'automne 2001 - et sur lesquels ma demande de mutation du LMC (UMR 5523) au LTCI (UMR 5141) a été acceptée - au printemps 2005 -. La liste de mes publications est donnée en Annexe A, la liste des articles présentés dans ce document est donnée page 6.

La description de mes travaux relatifs aux méthodes de simulation est organisée en trois chapitres, précédés d'un chapitre général.

Le Chapitre 1 décrit quelques méthodes de simulation classiques, et quelques extensions adaptatives de ces méthodes. Ce chapitre ne se veut pas une présentation exhaustive des méthodes de Monte Carlo: seuls les échantillonneurs dont l'étude a motivé les travaux exposés dans les Chapitres 2 à 4 sont présentés.

Le Chapitre 2 est consacré à la description de mes travaux relatifs à la théorie des processus Markoviens. Ces travaux sont suffisamment généraux pour s'appliquer à l'étude des processus Markoviens pas nécessairement issus de la modélisation d'algorithmes de simulation. Dans une première partie, je présente les résultats relatifs à la caractérisation de l'ergodicité de chaînes de Markov, et, dans une moindre mesure, de processus de Markov; ces caractérisations sont de deux formes: conditions de contraction du noyau de transition, et conditions de stabilité du modèle fluide du processus Markovien. Dans une seconde partie, je présente les résultats obtenus sur le contrôle explicite de convergence des noyaux de transition vers la loi stationnaire. Enfin, la troisième partie est consacrée à l'application de ces résultats à l'étude d'un échantillonneur MCMC et à l'étude de la vitesse de convergence de diffusions de Langevin (diffusions utilisées dans la construction des algorithmes de simulation *Metropolis-Adjusted Langevin Algorithm*).

Le Chapitre 3 est consacré à la description de mes travaux relatifs à l'optimisation des algorithmes de simulation. Dans la première partie, je présente les résultats d'une étude du comportement de certains échantillonneurs en phase transiente, basée sur l'étude de leur modèle fluide. Le modèle fluide est en quelque sorte "le squelette" d'une chaîne de Markov et donne donc une observation fine de la

dynamique de la chaîne. Utilisé dans le contexte des méthodes de simulation, le modèle fluide permet de comprendre le rôle des paramètres d'implémentation. Dans la seconde partie, je présente les résultats relatifs à la stabilité des algorithmes adaptatifs à dynamique markovienne, stabilité en termes de convergence de la loi du processus d'intérêt vers la loi cible, et d'existence d'une loi des grands nombres. Ces conditions de stabilité sont ensuite appliqués à l'étude de la convergence d'un échantillonneur MCMC adaptatif, et d'un échantillonneur MCMC non-linéaire. Enfin, la troisième partie est consacrée à l'étude de l'algorithme de stratification, méthode de simulation dite "de réduction de variance". Cette étude est menée en trois étapes: une étape théorique avec la dérivation d'un critère d'efficacité et l'identification des valeurs optimales des paramètres d'implémentation; une étape méthodologique avec la proposition d'une procédure de stratification adaptative; et une étape de validation, par application de cette méthode adaptative à la résolution d'un problème de *pricing* d'options.

Le Chapitre 4 est consacré aux applications des méthodes de simulation pour le traitement de données complexes. Deux applications sont considérées. La première relève de l'inférence statistique dans un modèle de champ de Markov caché; dans ce cas, la complexité des données est une conséquence de la dimension des variables cachées et de leur structure de dépendance probabiliste. La seconde relève de l'inférence statistique - au sens bayésien - en cosmologie : dans ce cadre applicatif, la vraisemblance des observations n'a pas d'expression explicite mais elle peut être évaluée au prix d'un temps de calcul très long (plusieurs secondes pour une seule évaluation). Il s'agit donc d'explorer la loi a posteriori par une méthode d'échantillonnage robuste à la forte corrélation entre les paramètres (au sens bayésien du terme), et facilement parallélisable.

Enfin, en Annexe A, sont récapitulées mes activités administratives, d'enseignement et de recherche.

Summary

Monte Carlo methods are stochastic numerical integration methods. They are widely used for solving problems that are otherwise intractable, such as for example evaluating high-dimensional integrals, computing complex functionals of a stochastic process, drawing samples from complex probability distributions. Physicists have been at the forefront of the methodological research in the field of Monte Carlo methods, and have provided algorithms which have been successfully applied to some complex sampling problems in physics ([82, 78, 115] for example). Nevertheless these methods are far from covering all the complex sampling situations. This keeps the methodological researches in simulation more than ever useful, and directly connected to many scientific communities (statistical physics, engineering, computational statistics, quantitative finance, …)

Monte Carlo algorithms depend on many design parameters that need to be finely tuned in order that the algorithm converges efficiently. For example, importance sampling algorithms draw samples under a proposal distribution and these outputs are then corrected - by a weighting-type correction - in order to approximate a target density. In this example, the proposal distribution is a design parameter. Markov Chain Monte Carlo (MCMC) algorithms consist in sampling a Markov chain with stationary distribution set to the target density. The ergodic properties of the chain - and thus, the efficiency of the sampler - depend on the transition kernel: in this example, the transition kernel is a design parameter.

The choice of these design parameters is mostly based on trial-and-error and is time-consuming. More importantly, properly addressing the choice of the design parameters requires a good understanding of the algorithms in order to be able to diagnose and to solve convergence problems. This expertise is the main factor in preventing the mass-diffusion of the Monte Carlo methodology. Recently, proposals have been made towards partial automation of the choice of the design parameters. These self-calibrating simulation methods are said to be *adaptive*: the design parameters are updated “on the fly”, based on the past outputs of the sampler. The dynamic of the adaptation is driven by finding the design parameters optimizing an efficiency criterion.

Developing a methodology, if possible general enough to have a large scope of application, is one thing. Checking that the algorithm really answers the question it has been developed for, is another. Proving the convergence of adaptive samplers is more complex than proving the convergence of non-adaptive ones. For example, the outputs of an adaptive MCMC sampler are not drawn from a (homogeneous) Markov chain so that proving the convergence of these samplers is not only based on Markov chains theory results. Research on simulation methods will help in the understanding of the dynamic of the samplers in order to identify the optimal design parameters; and also help in analyzing the behavior of the new samplers.

My research interests mainly focus on simulation methods, more precisely on a precise understanding of the dynamic of the samplers and of the role of the design parameters on this dynamic. I develop activities in parallel on the theory of Markov chains because I think the analysis of the behavior of algorithms has to be established on solid theoretical foundations. This interest in the “Markovian research interest” was originally motivated by the study of MCMC algorithms. Now it now appears crucial to analyze the adaptive MCMC algorithms and more generally, adaptive simulation methods with a Markovian dynamic.

In this document, I want to focus on the results of Monte Carlo simulations and Markov chain theory, which are the themes I described in the research program when I applied for my current position at the *Centre National de la Recherche Scientifique* (CNRS). I also restrict this description to the results obtained after my PhD defense (June 2001). The full list of my publications is given in Annex A, and the list of the works discussed in this document is shown on page 6.

This document is divided into four chapters: the first one general and the three others are devoted to my works. Most of these results are published; the papers are not included in this document but are attached to this document for convenience. I decided to take one or two main results from each paper and to comment them.

Chapter 1 describes some standard Monte Carlo methods. The goal of this chapter is not to provide an exhaustive description of the Monte Carlo methods. The introduction is limited to the algorithms that will be used in the following chapters, with an emphasis on the description of their design parameters. The first section is about two variance reduction techniques (see e.g. [61, 29]) - stratification and importance sampling - and about an adaptive importance sampling algorithm proposed by [37]. The second section is devoted to some MCMC samplers: the Hastings-Metropolis sampler, with an emphasis on the symmetric random walk Hastings-Metropolis (SRWM); the Metropolis-within-Gibbs sampler; and the Metropolis-Adjusted Langevin Algorithm (MALA). In this section two adaptive algorithms with Markov-type dynamic are also described: the Adaptive Metropolis sampler proposed by Haario et al.[64, 65] and the Equi-Energy sampler, proposed by Kou et al. [74].

Chapter 2 is devoted to the *Ergodicity of Markov processes* (published papers [4, 5, 7, 22] and technical reports [1, 11]). The analysis of simulation methods with Markovian-type dynamics relies on the theory of Markov processes. For example, MCMC algorithms sample a Markov chain with given stationary distribution. The efficiency of the sampler is thus closely related to the ergodicity of the chain; questions of interest are for example, the convergence of the chain to the stationary distribution, the explicit control of this convergence, or the role of the quantities that define the transition kernel, in this convergence. Ergodicity also implies Law of Large Numbers and a Central Limit theorem which are fundamental asymptotic results to prove the validity of the Monte Carlo methods, and to control the Monte Carlo approximation.

Section 2.1: The originality of the results is the development of criteria for ergodicity of Markov chains at a subgeometric rate. We propose two approaches: the first one is via the contraction properties of the transition kernel. We first derive a characterization based on a drift inequality satisfied by the one-iterated kernel (see section 2.1 and paper [7]); a second characterization is a drift

inequality satisfied by a state-dependent-iterated transition kernel (section 2.1.2 and paper [4]). We finally propose criteria for subgeometric ergodicity of strong Markov processes in terms of a drift inequality verified by the generator of the process (section 2.1.4 and papers [22, 5]). The second approach is via the fluid model of the chain : it is a scaling technique which consists of simplifying the behavior of the chain by an appropriate scaling in time, in space and in the initial condition. The chain is thus studied through time-continuous processes, called fluid limits. We show that the chain is ergodic at a subgeometric rate provided the fluid limits are stable; we also prove that the paths of the fluid model follow the solutions of an Ordinary Differential Equations (ODE) so that stability of the fluid limits is implied by the stability of the ODE (section 2.1.3 and paper [16]).

Section 2.2: An approach which has been intensively explored during the last decades to understand the dynamic of a Markov chain, is to compute explicit bounds for the convergence to stationarity. We derive such bounds in the case of geometric ergodicity (section 2.2.2 and technical report [11]) and subgeometric ergodicity (section 2.2.3 and paper [1]). These works extend previous results obtained during my PhD ([10, 19, 18]). In practice, these bounds are defined as the tails probability of some random time (in our construction, it is a “coupling” time); we thus deduce control of ergodicity from bounds on moments of this coupling time. Unfortunately, the bounds drastically deteriorate with the dimension of the sampling space (the same phenomenon occurs for all the bounds derived in the literature). When applied to the study of Monte Carlo samplers, these bounds are too intricate to provide information on the role of the design parameters. Nevertheless, they allow the comparison of Markov chains - in terms of ergodicity - and we will see in Chapter 3 that they are really useful to prove the ergodicity of adaptive simulation methods.

Section 2.3: Applications of these results to MCMC samplers are detailed. We first apply the subgeometric ergodicity criteria to prove the convergence of the SRWM algorithm when the target distribution is sub-exponential in the tails (section 2.3.1 and paper [7]). These works extend previous works on polynomial ergodicity obtained during my PhD ([10, 17, 19]). We then consider application of the subgeometric ergodicity criteria for (time continuous) Markov process, to the study of Langevin diffusion (section 2.3.2 and papers [22, 5]). We show that when the target distribution is sub-exponential in the tails, the Langevin diffusion cannot be geometrically ergodic. For *tempered* Langevin diffusion however, a convenient choice of the temperature involves geometric ergodicity.

Chapter 3 is devoted to the *Optimization of simulation methods* (published papers [2, 8, 16] and technical reports [3, 13]). Monte Carlo methods depend upon design parameters and these parameters have to be carefully tuned in order to improve the efficiency of the samplers. Adaptive Monte Carlo methods include some update steps in order to modify “on the fly” these design parameters. The development of such algorithms requires an understanding of the role of the design parameters on the efficiency of the sampler; and the definition of the “optimal design parameters” in order to guide the update of the parameters to these optimal values. These novel methods have to be analyzed to establish their convergence properties since adding an updating step in a classical Monte Carlo algorithm does not guarantee any more the convergence to the target distribution. This chapter is devoted to the (partial !) answers we obtained about adaptive samplers.

Section 3.1: We use the fluid limit model to understand the role of the design parameters of some MCMC samplers (namely, the SRWM in section 3.1.1 and paper [16]; and the Metropolis-within-

Gibbs in section 3.1.2 and technical report [13]) on the transient phase of the algorithms. This study gives new insights on the influence of the design parameters; nevertheless it has to be carried on to obtain non deterministic limit that will give information on the fluctuations of the chains from its deterministic “skeleton”.

Section 3.2: We analyze the behavior of adaptive Monte Carlo algorithms with Markovian dynamics; we provide sufficient conditions to prove the ergodicity of adaptive samplers i.e. convergence of the distribution of the process of interest to the target distribution (section 3.2.1 and technical report [3]), and a strong law of large numbers (section 3.2.2 and technical report [3]). These conditions are general enough to address the case of internal adaptation (the adaptation is based on the past behavior of the process of interest) and the case of external adaptation (the adaptation is based on the process of interest and on auxiliary processes). These results are then applied to the analysis of the behavior of the Adaptive Metropolis algorithm when the target density is sub-exponential in the tails (section 3.2.3 and paper [2]); and of the Equi-Energy sampler (section 3.2.4 and paper [3]). These two algorithms are examples of resp. internal and external adaptation.

Section 3.3: We develop and analyze the behavior of an adaptive stratification procedure (see paper [8]). We first derive an efficiency criterion for this algorithm, which is the asymptotic variance of the stratified estimator when the number of strata and the number of samples tend to infinity (see section 3.3.1). This criterion depends on the design parameters: the optimization of this criterion with respect to the parameters can not be explicitly solved. We propose an adaptive (stratification) algorithm where the design parameters are updated by a stochastic approximation dynamic, in order to optimize the efficiency criterion (section 3.3.2). We finally compare this algorithm to other stratification methods and to the crude Monte Carlo method when applied to the computation of the price of path-dependent options (section 3.3.3).

Chapter 4 is devoted to the *Applications of Simulation methods to the statistical analysis of complex data* (published papers [9, 23]).

Section 4.1: The first application is about the inference and restauration of missing data in a Hidden Markov Random Field model: in that case, complexity is a consequence of the Markovian dependence of the missing variables, and of the dimension of these variables. In missing data models, inference and restauration are usually solved by the Expectation Maximization (EM) algorithm; in this complex application, this algorithm is intractable. We propose an algorithm- called “Monte Carlo Variational EM” (MCVEM) - which roughly speaking, introduces Mean-Field-type Variational approach and simulation methods in the EM iterations. MCVEM is a stochastic algorithm and we identify the (almost-sure) asymptotic limits of the MCVEM paths. We finally compare MCVEM to other EM-type algorithms which combine relaxation methods and simulation methods. These results are published in [9].

Section 4.2: The second application is about (Bayesian) inference of cosmological parameters. In that case, the complexity of the model implies that the evaluation of the likelihood is really time-consuming (for some data sets, one evaluation takes a few seconds). Simulation methods that can be easily implemented on parallel computers are thus required for the exploration of the posterior density. We discuss the interest of an adaptive Importance Sampling method (M-PMC) to solve this inference problem and compare M-PMC to adaptive MCMC samplers. The main advantage of the

PMC method is the parallelization of the posterior calculations; the low variance of the Monte Carlo estimates; and the existence of simple diagnostics of convergence. These results are published in [23].

Notations

Espace d'état

Lorsque ce n'est pas précisé, X désigne un espace polonais que l'on munit de sa tribu borélienne \mathcal{X} . Dans la plupart des applications considérées dans ce mémoire, $\mathsf{X} \subseteq \mathbb{R}^d$.

Noyaux de transition

Soit $P : \mathsf{X} \times \mathcal{X} \rightarrow [0, 1]$ un noyau de transition sur X .

On note P^n le noyau de transition P itéré n fois et défini par récurrence

$$P^n(x, A) := \int P^{n-1}(x, dy)P(y, A).$$

Pour une probabilité μ sur X et une fonction $f : \mathsf{X} \rightarrow \mathbb{R}^+$ mesurable, μP^n et $P^n f$ sont respectivement une probabilité sur X et une fonction mesurable positive définies par

$$\mu P^n(A) := \int \mu(dx)P^n(x, A), \quad P^n f(x) := \int P^n(x, dy)f(y).$$

Normes

Soit $V : \mathsf{X} \rightarrow [1, +\infty)$ une fonction. On définit la V -norme d'une fonction $f : \mathsf{X} \rightarrow \mathbb{R}$ par

$$|f|_V := \sup_{\mathsf{X}} \frac{|f|}{V}; \tag{1}$$

et la distance en norme V entre deux probabilités μ_1, μ_2 sur $(\mathsf{X}, \mathcal{X})$ par

$$\|\mu_1 - \mu_2\|_V := \sup_{\{f \text{ mesurable, } |f|_V \leq 1\}} |\mu_1(f) - \mu_2(f)|. \tag{2}$$

La distance en variation totale, notée $\|\mu_1 - \mu_2\|_{\text{TV}}$, est la distance en norme V lorsque $V = 1$. En particulier, si $1 \leq V_1 \leq V_2$ alors $\|\mu_1 - \mu_2\|_{V_1} \leq \|\mu_1 - \mu_2\|_{V_2}$.

Processus

Les processus sont définis sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. On utilisera la notation $\mathbb{P}_\mu, \mathbb{E}_\mu$ pour désigner resp. la probabilité canonique et l'espérance sous cette probabilité, d'un processus de Markov de loi initiale μ . Lorsque $\mu = \delta_x$, on écrira simplement $\mathbb{P}_x, \mathbb{E}_x$.

Chapitre 1

Méthodes de Monte Carlo

Les méthodes de Monte Carlo sont des méthodes d'intégration numérique qui utilisent l'aléatoire. Si l'objectif est d'approcher une intégrale \mathcal{I} - incalculable explicitement - les méthodes de Monte Carlo sont basées sur la re-écriture de \mathcal{I} en une quantité de la forme

$$\int_{\mathbb{X}} \phi(x) \pi(x) d\lambda(x), \quad (1.1)$$

où λ est une mesure positive sur \mathbb{X} et $\pi : \mathbb{X} \rightarrow \mathbb{R}^+$ est une fonction mesurable telle que $\int_{\mathbb{X}} \pi d\lambda$ vaut 1. L'approximation de (1.1) est basée sur la Loi des Grands Nombres.

Si l'on sait obtenir une réalisation de v.a. i.i.d. $\{X_n, n \geq 0\}$ de loi $\pi d\lambda$, l'estimateur usuel de Monte Carlo de la quantité \mathcal{I} est défini par

$$\frac{1}{n} \sum_{k=1}^n \phi(X_k),$$

et la précision de cet estimateur est donnée par un Théorème Central Limite - quand il existe -

$$\sqrt{n} \left\{ \frac{1}{n} \sum_{k=1}^n \phi(X_k) - \pi(\phi) \right\} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \text{Var}(\phi(X_0))).$$

Si l'on ne sait pas obtenir une réalisation de v.a. i.i.d. de loi $\pi d\lambda$, une alternative consiste à obtenir une réalisation d'une chaîne de Markov $\{X_n, n \geq 0\}$ telle que la loi de X_n converge, en un certain sens, vers $\pi d\lambda$ de telle sorte que l'on dispose pour ce processus d'une Loi des Grands Nombres et éventuellement, d'un Théorème Central Limite (cf. [84, Chapitre 17] pour l'existence de théorèmes limites pour les chaînes de Markov).

Par exemple, les méthodes de réduction de variance sont des méthodes de simulation de v.a. i.i.d. (ou simplement indépendantes); on pourra consulter [97, 61, 29] pour une présentation de ces méthodes. Les méthodes de Monte Carlo par chaînes de Markov sont des méthodes de simulation d'une chaîne de Markov; on pourra consulter [97, 114] pour une présentation de ces méthodes.

L'utilisation de ces méthodes d'approximation numérique nécessite de

- (a) savoir obtenir des réalisations de v.a.; c'est un des critères de choix des lois instrumentales.
- (b) savoir contrôler l'erreur d'approximation ; une approche possible est de calculer des intervalles de confiance (gaussiens), ce qui nécessite de considérer le point suivant.

(c) étudier le comportement asymptotique des méthodes de simulation; vérifier en particulier que l'on a une loi des Grands Nombres, un Théorème Central Limite, ···, pour le processus simulé $\{X_n, n \geq 0\}$.

Etant donnée l'intégrale \mathcal{I} , la décomposition sous la forme (1.1) n'est pas unique: le choix de la loi instrumentale dans les méthodes de simulation de v.a. indépendantes ou le choix du noyau de transition dans les méthodes de simulation d'une chaîne de Markov est en pratique guidé par le point (a): l'utilisateur averti des méthodes de Monte Carlo doit choisir les différents paramètres d'implémentation afin de garantir le bien-fondé de la méthode et optimiser son efficacité (resp. vérifier les points (b) et (c)).

Pour identifier des paramètres d'implémentation plus indiqués que d'autres pour résoudre le problème numérique considéré, une approche possible est de tester l'efficacité de la méthode de simulation pour différentes valeurs des paramètres, puis de ne retenir que le plus performant. Cette approche a des limites: tout d'abord elle est très coûteuse en temps; ensuite, elle demande une certaine expertise de l'utilisateur qui doit non seulement savoir diagnostiquer les problèmes de convergence mais aussi savoir les résoudre. Il est en effet illusoire de penser qu'un mauvais choix de paramètres d'implémentation peut être compensé par un grand nombre de simulations.

Pour s'affranchir de cette expertise, il faut des méthodes qui déterminent "automatiquement" les paramètres d'implémentation adaptés au problème numérique considéré. Depuis une dizaine d'années, des **méthodes de simulation adaptatives** ont été proposées dans lesquelles les paramètres d'implémentation sont modifiés au fur et à mesure de l'algorithme, en fonction de son comportement [108, 65, 115, 24, 28, 38, 74, 26]. Le développement de tels échantillonneurs adaptatifs et l'étude de leur comportement asymptotique est actuellement un des enjeux majeurs de la recherche sur les méthodes de simulation.

Ce chapitre présente quelques méthodes de simulation - certaines adaptatives - dont l'étude est à la base des travaux de recherche exposés dans ce mémoire. Il se limite donc à une description succincte des échantillonneurs qui seront évoqués dans les chapitres suivants. Pour simplifier l'exposé, nous supposons que $X = \mathbb{R}^d$ et que λ désigne la mesure de Lebesgue sur \mathbb{R}^d .

1.1 Exemples de méthodes de réduction de variance

Les méthodes de réduction de variance que nous présentons sont basées sur la simulation de v.a. indépendantes sous une loi instrumentale $q d\lambda$; ces tirages sont ensuite pondérés pour approcher la loi cible $\pi d\lambda$.

Supposons que l'on souhaite estimer $\mathbb{E}[\phi(X)]$ où $X \sim \pi d\lambda$ et ϕ est une fonction à valeur réelle telle que $\mathbb{E}[\phi^2(X)] < +\infty$.

1.1.1 Méthode de stratification

Etant données une v.a. Y et $\{\mathcal{A}_i, i \in \{1, \dots, I\}\}$ une partition de l'espace d'état de Y , la quantité $\mathbb{E}[\phi(X)]$ est approchée par l'estimateur

$$\sum_{i=1}^I \mathbb{P}(Y \in \mathcal{A}_i) \frac{1}{n_i} \sum_{j=1}^{n_i} \phi(Y_j^{(i)}) ;$$

où les v.a. $\{Y_j^{(i)}, j \leq n_i, i \in \{1, \dots, I\}\}$ sont indépendantes et $Y_j^{(i)} \sim \mathbb{P}(X \in \cdot | Y \in \mathcal{A}_i)$. Sa mise en oeuvre dépend d'un certain nombre de paramètres d'implémentation:

- de la variable de stratification Y , d'un ensemble de I strates $\{\mathcal{A}_i, i \in \{1, \dots, I\}\}$ formant une partition de l'espace des valeurs de Y .
- de l'allocation $\{n_i, i \in \{1, \dots, I\}\}$ des tirages à chacune des strates. Etant donné un nombre total de tirages n , cette allocation est définie par la donnée d'une probabilité $\mathbf{Q} := \{q_i, i \in \{1, \dots, I\}\}$ sur $\{1, \dots, I\}$ telle que $n_i = \lfloor n \sum_{j=1}^i q_j \rfloor - \lfloor n \sum_{j=1}^{i-1} q_j \rfloor$.

L'estimateur stratifié est sans biais si pour tout i , $n_i > 0$. Sa variance est

$$\sum_{i=1}^I \{\mathbb{P}(Y \in \mathcal{A}_i)\}^2 \frac{1}{n_i} \text{Var}(\phi(X)|Y \in \mathcal{A}_i)$$

et quand $n \rightarrow +\infty$ et I est fixé (ou I tend vers $+\infty$ assez lentement), cette variance est équivalente à

$$\frac{1}{n} \sum_{i=1}^I \frac{\{\mathbb{P}(Y \in \mathcal{A}_i)\}^2}{q_i} \text{Var}(\phi(X)|Y \in \mathcal{A}_i) ;$$

Appliquée avec l'allocation proportionnelle donnée par $q_i = \mathbb{P}(Y \in \mathcal{A}_i)$, la variance de l'estimateur stratifié est équivalente, quand $n \rightarrow +\infty$ à

$$\frac{1}{n} \sum_{i=1}^I \mathbb{P}(Y \in \mathcal{A}_i) \text{Var}(\phi(X)|Y \in \mathcal{A}_i)$$

quantité qui est toujours inférieure à $n^{-1}\text{Var}(\phi(X))$: l'estimateur stratifié est donc de variance plus faible que l'estimateur de Monte Carlo standard calculé avec n tirages. La variance dépend de l'allocation \mathbf{Q} : les autres paramètres d'implémentation étant fixés, elle est minimale pour l'allocation optimale \mathbf{Q}^* donnée par

$$q_i^* \propto \mathbb{P}(Y \in \mathcal{A}_i) \sqrt{\text{Var}(\phi(X)|Y \in \mathcal{A}_i)} . \quad (1.2)$$

En pratique, dans les situations où l'on utilise les méthodes de simulation pour approcher $\mathbb{E}[\phi(X)]$, l'allocation optimale \mathbf{Q}^* n'est pas calculable, principalement parce que les variances conditionnelles $\text{Var}(\phi(X)|Y \in \mathcal{A}_i)$ ne sont pas connues.

Vers des méthodes de stratification adaptative

L'efficacité de cette méthode de simulation mesurée en terme de réduction de variance dépend des paramètres d'implémentation.

► Un des axes de recherche présenté dans ce mémoire consiste à calculer l'expression de la variance asymptotique (l'asymptotique étant ici à comprendre quand n et I tendent vers $+\infty$). L'optimisation de ce critère en les paramètres d'implémentation n'est pas explicite : nous avons donc proposé une méthode itérative de stratification adaptative qui, au fur et à mesure des itérations, adapte les paramètres d'implémentation de façon à converger vers l'optimum de la variance asymptotique. Cette recherche est présentée Chapitre 3.

1.1.2 Echantillonnage d'importance

Les méthodes d'échantillonnage d'importance reposent sur l'égalité suivante : pour toute loi instrumentale $q d\lambda$ telle que $\pi \prec q$, on a

$$\mathbb{E}[\phi(X)] = \mathbb{E}\left[\frac{\pi(Y)}{q(Y)} \phi(Y)\right]$$

où $Y \sim q d\lambda$. Par suite, une approximation de $\mathbb{E}[\phi(X)]$ est donné par un des estimateurs d'échantillonnage d'importance

$$\frac{1}{n} \sum_{k=1}^n \frac{\pi(Y_k)}{q(Y_k)} \phi(Y_k), \quad \text{ou} \quad \left(\sum_{j=1}^n \frac{\pi(Y_j)}{q(Y_j)} \right)^{-1} \sum_{k=1}^n \frac{\pi(Y_k)}{q(Y_k)} \phi(Y_k);$$

où les v.a. $\{Y_j, j \geq 1\}$ sont i.i.d. de loi $q d\lambda$. Le second estimateur dit "version auto-normalisée" ne requiert que la connaissance du ratio π/q , et en particulier de la loi cible π , à une constante multiplicative près.

La variance de cet estimateur (par exemple de la version non auto-normalisée) est donnée par

$$\frac{1}{n} \left(\mathbb{E}\left[\phi^2(X) \frac{\pi(X)}{q(X)}\right] - \{\mathbb{E}[\phi(X)]\}^2 \right).$$

Elle dépend donc du choix de la loi de proposition $q d\lambda$. En particulier, lorsque $\phi \geq 0$, cette variance est nulle en prenant pour loi de proposition $q_* \propto \phi \pi$. Ce choix n'est évidemment pas possible en pratique puisqu'il requiert la connaissance de la quantité d'intérêt $\pi(\phi)$. Néanmoins, cette expression de q_* signifie que le choix de q doit être guidé par un critère de proximité au produit $\phi \pi$.

Exemple de méthode adaptative : Mixture-Population Monte Carlo (M-PMC)

L'algorithme M-PMC [37], cas particulier de l'algorithme Population Monte Carlo [38], est une méthode d'échantillonnage d'importance adaptative pour le choix de la densité instrumentale q (voir aussi [116, 95]). La loi de proposition q est choisie parmi une famille \mathcal{Q} de densités de type mélange: l'adaptation consiste à modifier les poids et les composantes du mélange. La règle d'adaptation est guidée par la recherche d'une densité de proposition q_* qui, parmi la classe de densités considérée, minimise la divergence de Kullback entre la densité instrumentale q et la densité cible π , ce qui revient à résoudre le problème d'optimisation $\max_{q \in \mathcal{Q}} \int \ln q(y) \pi(y) \lambda(dy)$.

Ce problème peut formellement être relié au problème $\max_{q \in \mathcal{Q}} \sum_{k=1}^n \ln q(Y_k)$ - qui est celui que l'on résoud par l'algorithme Expectation Maximization (EM, [51]) pour déterminer l'estimateur du maximum de vraisemblance de la loi (de mélange) q à partir de réalisations i.i.d. $\{Y_k, k \leq n\}$.

Par suite, l'optimisation $\max_{q \in \mathcal{Q}} \int \ln q(y) \pi(y) \lambda(dy)$ peut être résolue par un algorithme de type Monte Carlo EM dans lequel toutes les étapes d'intégration sous π sont approchées par des

estimateurs d'échantillonnage d'importance avec pour loi de proposition, l'estimation courante $q^{(t)}$. Les étapes de maximisation sont en particulier explicites lorsque \mathcal{Q} est l'ensemble des mélanges de gaussiennes, ou des mélanges de t -distribution [37].

- Nous présenterons Chapitre 4 une application de M-PMC à l'estimation bayésienne de paramètres cosmologiques.

1.2 Exemples de méthodes de Monte Carlo par Chaînes de Markov (MCMC)

Les méthodes de Monte Carlo par chaînes de Markov (MCMC) sont basées sur la simulation d'une chaîne de Markov $\{X_n, n \geq 0\}$ dont $\pi d\lambda$ est probabilité invariante.

1.2.1 Echantillonneurs de Hastings-Metropolis à marche aléatoire symétrique (SRWM)

Etant donné un noyau de transition sur X de densité $q(x, y)$ (par rapport à la mesure λ), une transition $X_n \rightarrow X_{n+1}$ est décrite par

(a) conditionnellement à X_n , simuler $Y \sim q(X_n, \cdot) d\lambda$.

(b) poser

$$X_{n+1} = \begin{cases} Y & \text{avec probabilité } \alpha(X_n, Y) \\ X_n & \text{avec probabilité } 1 - \alpha(X_n, Y) \end{cases}$$

où le *ratio d'acceptation-rejet* α est définie par $\alpha(x, y) := 1 \wedge \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}$.

C'est donc un algorithme proche des méthodes de simulation dites "méthodes de rejet", dans lesquelles on obtient des réalisations de v.a. indépendantes sous une loi de proposition, réalisations que l'on corrige par un mécanisme d'acceptation-rejet [97, 61]. La grande souplesse des algorithmes de Hastings-Metropolis est de ne requérir que la connaissance de π à une constante multiplicative près. C'est une propriété fondamentale qui explique le succès de ces méthodes et leur utilisation dans des contextes de simulation complexes.

Le noyau de transition de ces échantillonneurs est donné par

$$P(x, A) = \int_A \alpha(x, y) q(x, y) d\lambda(y) + \delta_x(A) \int_{\mathsf{X}} (1 - \alpha(x, y)) q(x, y) d\lambda(y).$$

Dans le cas où le noyau de transition ne dépend pas de x , l'échantillonneur est dit "indépendant".

Dans le cas où la loi de la proposition q est de la forme $q(x, y) = q(y, x)$, le ratio d'acceptation $\alpha(x, y)$ se simplifie en $1 \wedge [\pi(y)/\pi(x)]$. Si de plus q est symétrique, l'échantillonneur est dit "à marche aléatoire symétrique" (*Symmetric Random Walk Metropolis*, SRWM). Le noyau de transition s'écrit alors

$$P(x, A) = \int_{A-x} \alpha(x, x+y) q(y) d\lambda(y) + \delta_x(A) \int_{\mathsf{X}} (1 - \alpha(x, x+y)) q(y) d\lambda(y).$$

où $A - x := \{y \in \mathsf{X}, x+y \in A\}$ et $\alpha(x, y) = 1 \wedge [\pi(y)/\pi(x)]$.

La loi de proposition q est un paramètre d'implémentation. La figure 1.1 montre l'évolution des trajectoires d'une chaîne SRWM à valeur dans \mathbb{R} , lorsque $q = \mathcal{N}(0, \sigma^2)$ pour différentes valeurs de σ^2 .

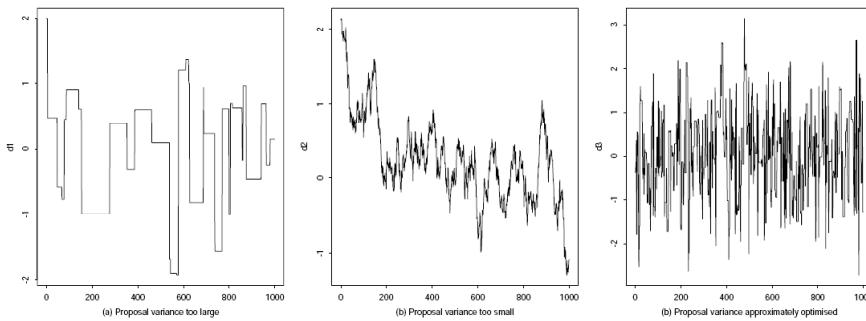


Fig. 1.1: Figures extraites de [114]. Réalisation d'une chaîne de Hastings-Metropolis à marche aléatoire symétrique, de loi de proposition gaussienne $\mathcal{N}(0, \sigma^2)$ pour σ^2 grand [gauche], σ^2 petit [centre] et σ^2 approprié [droite]

Si σ^2 est trop grand, presque tous les déplacements proposés sont refusés (car $\pi(x + y) \ll \pi(x)$) et la chaîne reste longtemps au même point. Si σ^2 est trop faible, la chaîne se déplace très lentement et elle mettra longtemps à converger vers π .

► L'étude du comportement de ces échantillonneurs repose sur l'étude du comportement des chaînes de Markov. Nous avons donc développé des résultats généraux en théorie des chaînes de Markov afin d'établir la vitesse de convergence des chaînes (et donc des échantillonneurs) vers leur mesure invariante π ; afin de contrôler explicitement cette convergence; et afin d'identifier le rôle des paramètres d'implémentation dans cette convergence. Ces résultats sont présentés Chapitre 2.

Exemple d'algorithme adaptatif: *Adaptive Metropolis*

Haario et al. ont proposé un algorithme de Hastings-Metropolis à marche aléatoire symétrique adaptatif, dans lequel la matrice de variance-covariance de la loi de proposition gaussienne $\mathcal{N}_d(0, \Sigma)$ est adaptée en ligne [64, 65]. Cet algorithme se résume par

$$\begin{aligned}\mu_{k+1} &= \mu_k + \frac{1}{k+1}(X_{k+1} - \mu_k) \\ \Gamma_{k+1} &= \Gamma_k + \frac{1}{k+1}((X_{k+1} - \mu_k)(X_{k+1} - \mu_k)^T - \Gamma_k),\end{aligned}$$

où la loi conditionnelle de X_{k+1} sachant 'le passé' est un noyau de transition SRWM de loi cible π et de loi de proposition $\mathcal{N}_d(0, [2.38]^2 d^{-1} \Gamma_k)$. La convergence de cet algorithme a été étudié lorsque la loi cible est une densité sur-exponentielle.

Cet algorithme est un exemple du cadre général suivant: on se donne une famille de noyaux de transition $\{P_\theta, \theta \in \Theta\}$ tels que pour tout θ , $\pi P_\theta = \pi$; on construit un processus d'intérêt $\{X_n, n \geq 0\}$ tel que $\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) = P_{\theta_n}(X_n, A)$ où $\mathcal{F}_n := \sigma(X_k, \theta_k, k \leq n)$. $\{X_n, n \geq 0\}$ n'est plus une chaîne de Markov: l'étude du comportement asymptotique de cet algorithme ne relève donc plus seulement de la théorie des chaînes de Markov.

► Nous avons étudié la convergence de ces algorithmes adaptatifs dits à “adaptation interne” du fait que le paramètre θ_n ne dépend que du passé du processus d’intérêt $(X_0, \dots, X_{n-1}, X_n, \theta_0)$; et appliqué ces résultats à l’étude de la convergence de l’algorithme *Adaptive Metropolis* lorsque la densité cible π est une densité sous-exponentielle. Ces résultats sont présentés Chapitre 3.

1.2.2 Echantillonneur Metropolis-within-Gibbs

Les algorithmes de Hastings-Metropolis et l’échantillonneur de Gibbs (voir par exemple [97]) sont utilisés comme éléments de base d’algorithmes dits **hybrides**. L’échantillonneur Metropolis-within-Gibbs en est un exemple.

Soit $\{\omega_1, \dots, \omega_d\}$ une probabilité sur $\{1, \dots, d\}$; on note $\pi(x_i | x_{j,j \neq i})$ la i -ème loi conditionnelle. Une itération de l’algorithme est décrite par

- choisir une composante $I \in \{1, \dots, d\}$ selon la loi de probabilité $\{\omega_1, \dots, \omega_d\}$.
- laisser les $d - 1$ autres composantes inchangées et mettre à jour la composante I selon un noyau de Hastings-Metropolis à marche aléatoire symétrique sur \mathbb{R} , de densité cible $\pi(x_I | x_{j,j \neq I})$.

Les paramètres d’implémentation de cet algorithme sont donc les poids de sélection $\{\omega_1, \dots, \omega_d\}$ et les paramètres d’implémentation des d échantillonneurs de Hastings-Metropolis.

► Nous étudierons Chapitre 3 l’influence de ces paramètres d’implémentation sur la dynamique de l’algorithme en phase transiente.

1.2.3 Echantillonneur *Metropolis-Adjusted Langevin Algorithm* (MALA)

Dans l’échantillonneur de Langevin, la loi de proposition utilise de l’information sur la densité cible et plus précisément l’information contenue dans le gradient $\nabla \log \pi$.

La loi de proposition est construite comme une discrétisation de type Euler d’une diffusion de Langevin de probabilité stationnaire $\pi d\lambda$. Dans l’algorithme MALA,

$$q(x, \cdot) = \mathcal{N}_d \left(x + \frac{h^2}{2} \nabla \log \pi(x), h^2 \text{Id}_d \right),$$

pour un pas $h > 0$. Notons que bien que la chaîne de noyau de transition $q(x, dy)$ soit une discrétisation d’une diffusion de loi stationnaire π , π n’est pas mesure invariante pour cette chaîne. Par exemple, dans le cas $d = 1$ et $\pi = \mathcal{N}(0, 1)$, $q(x, dy) = \mathcal{N}(x + \frac{h^2}{2} \nabla \log \pi(x), h^2)$ (pour tout x). Cela justifie l’ajout d’une correction de type acceptation-rejet des tirages sous $q(x, \cdot)$.

Choix du pas en fonction de l’état courant

Le pas h est un paramètre d’implémentation. D’un côté on souhaiterait prendre h suffisamment petit pour que le noyau de proposition q ait une mesure invariante proche de π . D’un autre côté, plus h est petit et plus il faut d’itérations pour que la chaîne $\{X_k, k \geq 0\}$ ressemble à la diffusion de Langevin $\{L_t, t \in \mathbb{R}^+\}$ à un instant t donné (en quelque sorte, “ $X_{\lfloor t/h \rfloor} \approx L_t$ ”)

► Nous présenterons Chapitre 2 des résultats sur la vitesse de convergence des diffusions de Langevin vers π , lorsque la matrice de diffusion $\sigma(x)$ est de la forme $\pi^{-s}(x) \text{Id}$ ou $|\ln \pi|^s(x) \text{Id}$, $s > 0$; et discuterons de l'influence de s sur cette vitesse de convergence. Nous verrons qu'en deçà d'une “température critique” s_* cette vitesse ne peut pas être géométrique; mais qu'au-delà, elle est géométrique.

1.2.4 Exemple d'algorithme MCMC à adaptation externe: l'échantillonneur *Equi-Energy*

L'algorithme *Equi-Energy* proposé par Kou et al. [74], est un algorithme adaptatif d'un esprit différent de ceux présentés jusqu'ici: au lieu d'améliorer les performances de l'échantillonneur en modifiant les paramètres d'implémentation en fonction du comportement passé de l'algorithme, on améliore les performances de l'échantillonneur en introduisant une interaction avec un processus auxiliaire construit pour avoir de bonnes propriétés de convergence.

L'échantillonneur *Equi-Energy* est basé sur la simulation de K processus $X^{(1)}, \dots, X^{(k)}, \dots, X^{(K)}$ en parallèle, dont on attend de chacun qu'ils convergent vers une loi cible $\pi^{(k)}$. La dernière loi est la loi cible ($\pi^{(K)} = \pi$), tandis que les autres lois cibles sont en particulier choisies pour que le processus $X^{(k)}$ ait de bonnes propriétés de convergence. Par exemple, $\pi^{(k)} = \pi^{\beta_k}$ où $\{\beta_k, k \leq K-1\}$ est une famille croissante de $]0, 1[$. Les processus sont construits de façon à interagir avec les processus adjacents (une idée adaptée des algorithmes de *tempering* [60, 78]); on commence par construire le processus le plus “froid” $X^{(1)}$, puis itérativement, on construit $X^{(k)}$ à l'aide de $X^{(k-1)}$.

Il suffit donc de décrire l'algorithme dans le cas $K = 2$. On suppose que les interactions ne se font qu'avec certains points du processus adjacent; on introduit donc une fonction de sélection $g : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}^+$. On note $\tilde{\pi}$ et π , les mesures cibles des processus $(k-1)$ et k ; et $\{Y_n, n \geq 0\}, \{X_n, n \geq 0\}$ les processus $X^{(k-1)}, X^{(k)}$.

Etant donné un noyau de transition P sur \mathbb{X} de mesure invariante π et un réel $\epsilon \in]0, 1[$, la loi de X_{n+1} conditionnellement à $(X_0, \dots, X_n, Y_0, \dots, Y_n)$ est décrite par

- a) avec probabilité $(1 - \epsilon)$: $X_{n+1} \sim P(X_n, \cdot)$.
- b) avec probabilité ϵ :
 - choisir $Z \in \{Y_1, \dots, Y_n\}$ avec probabilité $\mathbb{P}(Z = Y_k) \propto g(X_n, Y_k)$.
 - poser $X_{n+1} = Z$ avec probabilité $\alpha(X_n, Z)$, et $X_{n+1} = X_n$ sinon, où
$$\alpha(x, y) := 1 \wedge \frac{\pi(y)\tilde{\pi}(x)}{\tilde{\pi}(y)\pi(x)};$$

ce qui se résume par le noyau de transition $P_{\theta_n}(X_n, \cdot) = (1 - \epsilon)P(X_n, \cdot) + \epsilon\bar{P}_{\theta_n}(X_n, \cdot)$ où

$$\bar{P}_{\theta}(x, \cdot) = \frac{1}{\theta[g(x, \cdot)]} \left\{ \alpha(x, y)g(x, y)\theta(dy) + \delta_x(dy) \left(1 - \int \alpha(x, y)g(x, y)\theta(dy) \right) \right\},$$

$$\theta_n := n^{-1} \sum_{k=1}^n \delta_{Y_k} \text{ et } \theta_n[g(x, \cdot)] := \int \theta_n(dy) g(x, y).$$

La fonction de sélection est de la forme $g(x, y) \propto \sum_{i=1}^I \mathbb{1}_{\mathbb{X}_i \times \mathbb{X}_i}(x, y)$ où $\{\mathbb{X}_1, \dots, \mathbb{X}_I\}$ est une partition de \mathbb{X} . Kou et al. prennent pour partition des ensembles d'énergie de la forme $\mathbb{X}_i = \{x \in$

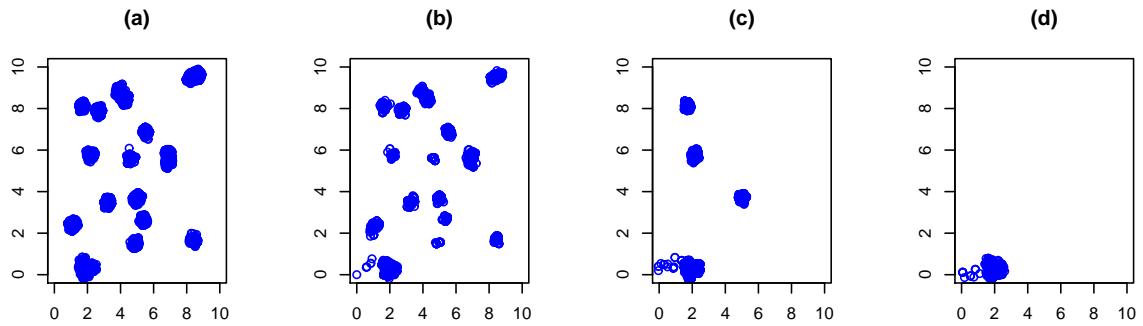


Fig. 1.2: [de gauche à droite] Densité cible, échantillonneur *Equi-Energy*, échantillonneur *Parallel Tempering* [60] et échantillonneur SRWM

$\mathbb{X}, a_i \leq -\ln \pi(x) < b_i\}$ si bien que seules les interactions avec des points d'énergie voisine de l'état courant X_n ne sont autorisées.

Le ratio d'acceptation-rejet $\alpha(x, y)$ est choisi de façon à ce que lorsque $\theta = \tilde{\pi}$, \bar{P}_θ soit le noyau d'une chaîne de Hastings-Metropolis indépendant de mesure cible π et de loi de proposition $\tilde{\pi}$. Ainsi, $\pi \bar{P}_{\tilde{\pi}} = \pi$ puis $\pi P_{\tilde{\pi}} = \pi$. L'intuition est donc que si $\{\theta_n, n \geq 0\}$ converge vers $\tilde{\pi}$ - dans une limite à préciser - le processus $\{X_n, n \geq 0\}$ se comportera comme une chaîne de probabilité invariante π .

Le choix de la partition, des différentes mesures cibles $\pi^{(k)}$ et de la probabilité d'interaction ϵ jouent un rôle sur l'efficacité de cet algorithme ([74, 30]).

La Figure 1.2 compare le résultat de plusieurs échantillonneurs pour la simulation d'un mélange de lois gaussiennes dans \mathbb{R}^2 , après un même nombre d'itérations: l'algorithme SRWM est resté bloqué sur un mode de la densité cible, tandis que l'échantillonneur *Equi-Energy* a visité tous les modes (cf. [74, 3] pour la description de cette simulation).

Cet algorithme est un exemple du cadre général suivant: on se donne une famille de noyaux de transition $\{P_\theta, \theta \in \Theta\}$, une filtration $\{\mathcal{F}_n, n \geq 0\}$ et un processus auxiliaire $\{Y_n, n \geq 0\}$ adapté; on construit un processus d'intérêt adapté $\{X_n, n \geq 0\}$ tel que $\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) = P_{\theta_n}(X_n, A)$. $\{X_n, n \geq 0\}$ n'est plus une chaîne de Markov: l'étude du comportement asymptotique de cet algorithme ne relève donc plus seulement de la théorie des chaînes de Markov.

- Nous avons étudié la convergence de ces algorithmes adaptatifs dits à “adaptation externe” du fait que le paramètre θ_n dépend du passé du processus d'intérêt et de celui d'un processus auxiliaire. Cette étude a été menée dans le cas où chaque noyau P_θ possède une probabilité invariante π_θ telle que la suite $\{\pi_{\theta_n}, n \geq 0\}$ converge vers π dans une limite à préciser.

Nous avons appliqué ces résultats à l'étude de la convergence de l'algorithme *Equi-Energy*, et plus précisément à l'étude de la convergence de la loi de X_n vers π et de l'existence d'une loi forte des Grands Nombres. Ces résultats sont présentés Chapitre 3.

Chapitre 2

Ergodicité des processus de Markov

Les résultats que nous présentons dans ce chapitre sont issus des articles

- [4] S. Connor and G. Fort. State-dependent Foster-Lyapunov criteria for subgeometric convergence of Markov chains. *Stoch. Process. Appl.*, 119:4176–4193, 2009.
- [5] R. Douc, G. Fort, and A. Guillin. Subgeometric rates of convergence of f-ergodic strong Markov processes. *Stoch. Process Appl.*, 119(3):897–923, 2009.
- [7] R. Douc, G. Fort, E. Moulines, and P. Soulier. Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.*, 14(3):1353–1377, 2004.
- [22] G. Fort and G.O. Roberts. Subgeometric ergodicity of strong Markov processes. *Ann. Appl. Probab.*, 15(2):1565–1589, 2005.

et des rapports de recherche

- [1] C. Andrieu and G. Fort. Explicit control of subgeometric ergodicity. Technical Report 05:17, University of Bristol, UK, 2005.
- [11] G. Fort. Computable bounds for V-geometric ergodicity of Markov transition kernels. Technical Report RR 1047-M, Univ. J. Fourier, France, 2002.

Motivation. L'étude de la convergence des algorithmes de simulation à dynamique markovienne s'appuie sur la théorie des chaînes de Markov.

Par exemple, les méthodes de simulation Monte Carlo par Chaînes de Markov (MCMC) sont des procédures itératives pour l'obtention d'une réalisation d'une chaîne de Markov de loi stationnaire donnée. Par suite, étudier les propriétés de convergence de cet algorithme revient à étudier l'ergodicité des chaînes de Markov : *(i)* établir la convergence de la loi de la chaîne vers sa mesure invariante; *(ii)* contrôler explicitement cette convergence; *(iii)* identifier le rôle des quantités qui définissent le noyau de transition dans la convergence des chaînes vers leur loi stationnaire; *(iv)*

De ces résultats de convergence, on peut en particulier déduire des théorèmes limites comme la Loi des Grands Nombres ou le Théorème Central Limite pour des chaînes de Markov. Appliqués dans le contexte de la simulation, ces théorèmes justifient le bien-fondé des méthodes de Monte Carlo à dynamique markovienne et précisent la qualité d'approximation.

Ce chapitre est consacré à la présentation des résultats obtenus en théorie des chaînes de Markov et à leur application à l'étude de convergence de quelques échantillonneurs MCMC.

Points forts de notre contribution. Nous avons établi des critères d'ergodicité à une vitesse sous-géométrique, en norme de variation totale et plus généralement en norme V , pour des chaînes de Markov et processus de Markov: ces résultats sont présentés section 2.1. Nous proposons deux approches pour établir l'ergodicité.

La première approche est basée sur des propriétés de contraction du noyau de transition de la chaîne de Markov, ou du générateur du processus de Markov. Ces conditions, connues dans la littérature sous le nom de “conditions de drift”, sont surtout développées pour caractériser l'ergodicité des chaînes à une vitesse géométrique. L'originalité de notre contribution est d'avoir étendu ces critères à la caractérisation de l'ergodicité sous-géométrique. Ces travaux sont présentés sections 2.1.1 et 2.1.2 (ergodicité des chaînes) et section 2.1.4 (ergodicité des processus).

La seconde approche est basée sur l'étude du modèle fluide i.e. l'étude d'un processus à temps continu, limite d'une famille de processus cadlag obtenus par renormalisations en temps, en espace, et en le point initial de la chaîne de Markov. Les modèles fluides sont des techniques très répandues pour l'étude de la stabilité de processus de Markov modélisant le comportement de réseaux complexes. L'originalité de notre contribution est d'avoir étendu cette notion de modèle fluide, à certaines chaînes de Markov; et d'avoir établi le lien entre stabilité du modèle fluide et ergodicité de la chaîne. Ces travaux sont présentés section 2.1.3.

La section 2.2 est consacrée au contrôle explicite d'ergodicité: quantifier l'écart entre la loi de la chaîne de Markov à l'instant n , et sa loi stationnaire. Nous avons construit de telles bornes principalement sous deux hypothèses sur le noyau de transition: une hypothèse de minoration du noyau de transition; et une hypothèse d'existence d'une solution à une condition de drift. La section 2.2.2 (resp. 2.2.3) traite du cas où la condition de drift est de type géométrique (resp. sous-géométrique). L'originalité de notre travail réside dans le contrôle explicite de convergence pour des chaînes ergodiques à une vitesse sous-géométrique; les premiers travaux sur ce sujet ont été établis durant ma thèse [10] et, suite à de nouvelles caractérisations de l'ergodicité sous-géométrique, nous avons généralisé ces premiers résultats.

Dans la section 2.3, nous présentons l'application de ces critères d'ergodicité à l'étude des méthodes MCMC et à l'étude de la vitesse de convergence de diffusions de Langevin. En section 2.3.1, on établit l'ergodicité sous-géométrique d'un algorithme de Hastings-Metropolis à marche aléatoire symétrique lorsque la densité cible est une densité sous-exponentielle (dans les queues). En section 2.3.2, on montre que par un choix adéquat de la matrice de diffusion, on peut accélérer la convergence des diffusions de Langevin vers leur mesure stationnaire. Ces travaux sont motivés par les applications à l'implémentation des échantillonneurs MALA.

2.1 Critères de V -ergodicité sous-géométrique

Une chaîne de Markov de noyau de transition P est dite V -ergodique (ou simplement ergodique lorsque $V = 1$) si il existe une unique probabilité π telle que pour π -presque tout x , on a

$$\lim_n \|P^n(x, \cdot) - \pi\|_V = 0 . \quad (2.1)$$

Si de plus la convergence a lieu à la vitesse $r = \{r(n), n \geq 0\}$

$$\lim_n r(n) \|P^n(x, \cdot) - \pi\|_V = 0 ,$$

pour une suite croissante telle que $\lim_n r(n) = +\infty$ on parlera de (V, r) -ergodicité. On trouve dans la littérature beaucoup de travaux l'ergodicité et la V -ergodicité à un taux géométrique, les premiers travaux remontant aux années 1960 [63]. En revanche très peu de résultats existent sur l'ergodicité à des vitesses plus lentes, telles que des vitesses logarithmiques, polynomiales ou sous-exponentielles. Nous nous sommes donc intéressés à des conditions garantissant la V -ergodicité à de telles vitesses et plus généralement, à la vitesse sous-géométrique.

Définition 1 (taux sous-géométrique). *Un taux de convergence $r = \{r(n), n \geq 0\}$ est dit sous-géométrique si il existe une fonction croissante $r_\star : \mathbb{N} \rightarrow [1, +\infty)$ telle que*

$$\frac{\log r(n)}{n} \downarrow 0 , \quad 0 < \liminf_n \frac{r(n)}{r_\star(n)} \leq \limsup_n \frac{r(n)}{r_\star(n)} < +\infty .$$

Dans la littérature sur les chaînes de Markov, des conditions suffisantes pour la V -ergodicité reposent essentiellement sur des critères de récurrence d'un ensemble small¹.

Définition 2 (ensemble small). *Un ensemble mesurable \mathcal{C} est (n, ν, ϵ) -small (ou simplement small) si il existe un entier $n \in \mathbb{N}_*$, une mesure de probabilité ν et un réel $\epsilon \in]0, 1[$ tels que*

$$P^n(x, \cdot) \geq \mathbb{I}_{\mathcal{C}}(x) \epsilon \nu(\cdot) .$$

Dans le cas d'un espace d'état général, la psi-irréductibilité de la chaîne entraîne l'existence d'ensembles small [84, Théorème 5.2.2]. Ces critères de récurrence d'un ensemble small sont de deux formes. Dans la première, le critère est exprimé en termes d'existence de moments de $\tau_{\mathcal{C}}$, le temps de retour dans \mathcal{C}

$$\tau_{\mathcal{C}} := \inf\{n \geq 1, X_n \in \mathcal{C}\} ;$$

dans la seconde, il est exprimé à l'aide d'une condition de drift i.e. une condition de contraction du noyau P à l'extérieur de \mathcal{C} : il existe un réel $b \in \mathbb{R}^+$ et une fonction $f : \mathbb{X} \rightarrow [1, +\infty[$ tels que

$$PV - V \leq -f + b\mathbb{I}_{\mathcal{C}} .$$

Ces deux formes ont en fait très liées puisque dans certains cas - comme le cas d'une vitesse constante $r(n) = 1$ ou le cas des vitesses géométriques - on sait établir l'équivalence de l'existence d'un moment du temps de retour, et de l'existence d'une solution à une condition de drift. En revanche, l'équivalence reste un problème ouvert dans le cas des vitesses sous-géométriques.

La première caractérisation de la V -ergodicité sous-géométrique des chaînes de Markov est due à Tuominen & Tweedie [113, Théorème 2.1].

¹on pourra consulter [84] pour les notions classiques en théorie des chaînes de Markov

Théorème 3. Soit P un noyau de transition psi-irréductible, apériodique. Si il existe une fonction $V : \mathbb{X} \rightarrow [1, +\infty)$, un taux sous-géométrique $r = \{r(n), n \geq 0\}$ et un ensemble small \mathcal{C} tels que

$$\sup_{x \in \mathcal{C}} \mathbb{E}_x \left[\sum_{k=0}^{\tau_{\mathcal{C}}-1} r(k) V(X_k) \right] < +\infty, \quad (2.2)$$

alors la chaîne possède une unique mesure de probabilité invariante π et pour π -presque tout x , $\lim_n r(n) \|P^n(x, \cdot) - \pi\|_V = 0$.

Dans le contexte de la simulation, et plus précisément des algorithmes MCMC, il est très difficile de calculer des moments du temps de retour $\tau_{\mathcal{C}}$. Des critères de type “conditions de drift” sont donc privilégiés.

2.1.1 Condition de drift sur le noyau de transition P

Nous avons montré que le contrôle de moments de la forme (2.2) se déduisait d'une condition de drift. Cette activité de recherche sur la caractérisation de la V -ergodicité sous-géométrique est la suite de travaux effectués durant ma thèse relatifs à la caractérisation de la V -ergodicité à une vitesse polynomiale ([17, 19, 10]).

Pour un noyau P psi-irréductible et apériodique, une condition de drift de la forme $PW \leq W - 1 + b\mathbb{I}_{\mathcal{C}}$ pour un ensemble small \mathcal{C} entraîne l'ergodicité [84, Chapitre 10]; une condition de la forme $PW \leq W - V + b\mathbb{I}_{\mathcal{C}}$ entraîne la V -ergodicité [84, Chapitre 14]; une condition de la forme $PV \leq V - (1 - \lambda)V + b\mathbb{I}_{\mathcal{C}}$ pour $\lambda \in]0, 1[$ entraîne la V -ergodicité à une vitesse géométrique.

Nous avons étudié les conditions de la forme $PV \leq V - \phi \circ V + b\mathbb{I}_{\mathcal{C}}$ pour une fonction $\phi : [1, +\infty) \rightarrow]0, +\infty[$ concave, croissante et différentiable. Le Théorème 5 ci-dessous établit le lien entre cette forme de condition de drift et la V -ergodicité à un taux sous-géométrique.

Soit

$$r_{\phi}(n) := \phi \circ H_{\phi}^{-1}(n) \quad \text{où} \quad H_{\phi}(t) := \int_1^t \frac{dx}{\phi(x)}. \quad$$

Par exemple,

- (taux logarithmiques) si $\phi(t) \sim c(1 + \log(t))^{\alpha}$ pour $c \in]0, 1]$ et $\alpha \geq 0$ alors $r_{\phi}(t) \sim \log^{\alpha}(t)$;
- (taux polynomiaux) si $\phi(t) \sim ct^{1-\alpha}$ pour $c > 0$ et $\alpha \in]0, 1[$ alors $r_{\phi}(t) \sim t^{1/\alpha-1}$;
- (taux sous-exponentiels) si $\phi(t) \sim ct/[\log t]^{\alpha}$ pour $c, \alpha > 0$ alors $r_{\phi}(t) \sim t^{-\alpha/(1+\alpha)} \exp(\{c(1 + \alpha)t\}^{1/(1+\alpha)})$.

Proposition 4. [7, Proposition 2.2] Supposons qu'il existe une fonction $V : \mathbb{X} \rightarrow [1, +\infty[$; une fonction $\phi : [1, +\infty) \rightarrow]0, +\infty[$ concave, croissante et différentiable; un réel $b \in \mathbb{R}^+$ et un ensemble mesurable \mathcal{C} tels que

$$PV \leq V - \phi \circ V + b\mathbb{I}_{\mathcal{C}}.$$

Alors pour tout $x \in \mathbb{X}$,

$$\mathbb{E}_x \left[\sum_{k=0}^{\tau_{\mathcal{C}}-1} \phi \circ V(X_k) \right] \leq V(x) + b\mathbb{I}_{\mathcal{C}}(x) \quad \mathbb{E}_x \left[\sum_{k=0}^{\tau_{\mathcal{C}}-1} r_{\phi}(k) \right] \leq V(x) + b \frac{r_{\phi}(0)}{r_{\phi}(1)} \mathbb{I}_{\mathcal{C}}(x).$$

Ces moments modulés du temps de retour $\tau_{\mathcal{C}}$ sont de la forme (2.2) : le premier correspond au cas $r \leftarrow 1$ et $V \leftarrow \phi \circ V$; le second correspond au cas $r \leftarrow r_\phi$ et $V \leftarrow 1$.

En utilisant une inégalité de convexité, on peut obtenir des moments modulés interpolés: plus précisément soit \mathcal{Y} l'ensemble des paires de fonctions (Ψ_1, Ψ_2) définies sur $[1, +\infty)$, croissantes et telles que $\lim_{+\infty} \Psi_1 = +\infty$ ou $\lim_{+\infty} \Psi_2 = +\infty$ et pour tout $x, y \geq 1$

$$\Psi_1(x)\Psi_2(y) \leq x + y .$$

Alors sous les hypothèses de la Proposition 4

$$\mathbb{E}_x \left[\sum_{k=0}^{\tau_{\mathcal{C}}-1} \Psi_1(r_\phi(k)) \Psi_2(\phi \circ V(X_k)) \right] \leq 2V(x) + b \left(1 + \frac{r_\phi(1)}{r_\phi(0)} \right) \mathbb{1}_{\mathcal{C}}(x) ,$$

ce qui donne un moment de la forme (2.2) avec $r \leftarrow \Psi_1 \circ r_\phi$ et $V \leftarrow \Psi_2 \circ \phi \circ V$. Des exemples de paires éléments de \mathcal{Y} sont les fonctions $\Psi_1(x) = x$, $\Psi_2(x) = 1$; $\Psi_1(x) \propto x^{1/p}$, $\Psi_2(x) \propto x^{1/q}$ avec $1/p + 1/q = 1$; et plus généralement, les inverses des paires de Young [75, Chapitre 1].

Pour appliquer le Théorème 3, il faut vérifier que $\Psi_1 \circ r_\phi$ est un taux sous-géométrique, ce qui est vrai si $\lim_{+\infty} \phi' = 0$ [7, Lemme 2.7]. Il faut aussi vérifier que l'on peut toujours supposer que $\sup_{\mathcal{C}} V < +\infty$ [7, Lemme 2.4].

Nous avons donc le résultat suivant:

Théorème 5. [7, Théorème 2.8] Soit P un noyau psi-irréductible et apériodique. Supposons qu'il existe une fonction $V : \mathsf{X} \rightarrow [1, +\infty[$; une fonction $\phi : [1, +\infty) \rightarrow]0, +\infty[$ concave, croissante et différentiable; un réel $b \in \mathbb{R}^+$ et un ensemble small \mathcal{C} tels que $PV \leq V - \phi \circ V + b\mathbb{1}_{\mathcal{C}}$. Alors la chaîne possède une probabilité invariante π et pour tout $x \in \mathsf{X}$ et toute paire $(\Psi_1, \Psi_2) \in \mathcal{Y}$,

$$\lim_n \Psi_1 \circ r_\phi(n) \|P^n(x, \cdot) - \pi\|_{\Psi_2 \circ \phi \circ V} = 0 .$$

Ce théorème appliqué avec la paire $(\Psi_1(x), \Psi_2(x)) = (x, 1)$ permet d'établir l'ergodicité à la vitesse sous-géométrique r_ϕ et en particulier à des vitesses logarithmiques, polynomiales ou sous-exponentielles selon l'expression de ϕ .

Appliqué avec la paire $[\Psi_1(x), \Psi_2(x)] \sim [x^{1/p}, x^{1/q}]$ pour $1/p + 1/q = 1$, il montre que la vitesse de convergence dépend de la norme de convergence : plus la norme est forte et plus la vitesse est faible, et inversement. Par exemple, dans le cas où $\phi(t) \sim t^{1-\alpha}$, $\alpha \in]0, 1[$, on déduit du Théorème 5 que pour tout $\kappa \in [1, 1/\alpha]$

$$\lim_n n^{\kappa-1} \|P^n(x, \cdot) - \pi\|_{V^{1-\alpha\kappa}} = 0 .$$

Ce résultat d'ergodicité polynomiale avait déjà été établi par [69, 19]; nous le retrouvons comme un cas particulier d'une théorie plus générale qui établit l'ergodicité à des taux sous-géométriques pas nécessairement polynomiaux.

En appliquant ce Théorème dans le cas “limite” où $\phi(t) = 1$ (resp. $\phi(t) = (1-\lambda)t$ pour $\lambda \in]0, 1[$), on a $r_\phi(t) \sim 1$ (resp. $r_\phi(t) \sim \kappa^t$ pour $\kappa > 1$) et on retrouve les critères d'ergodicité (resp. V-ergodicité géométrique). Bien que ce passage à la limite ne soit pas rigoureux, il nous permet de vérifier que nos résultats s'inscrivent dans la continuité de ceux existants.

2.1.2 Condition de drift sur un itéré du noyau de transition P

L'ergodicité et la V -ergodicité géométrique se déduisent aussi de conditions de drift portant sur un itéré du noyau, l'indice d'itération dépendant de l'état courant. Plus précisément, Meyn & Tweedie [85] ont étudié des conditions de la forme

$$\begin{aligned} P^{n(x)}V(x) &\leq V(x) + b\mathbb{1}_{\mathcal{C}}(x), \\ P^{n(x)}V(x) &\leq V(x) - n(x) + b\mathbb{1}_{\mathcal{C}}(x), \\ P^{n(x)}V(x) &\leq \beta^{n(x)} V(x) + b\mathbb{1}_{\mathcal{C}}(x), \end{aligned}$$

et montré que si un noyau de transition psi-irréductible apériodique vérifie une de ces conditions de drift pour un ensemble small \mathcal{C} , alors la chaîne est resp. Harris-recurrente, positive Harris-recurrente et V -géométriquement ergodique.

Nous nous sommes intéressés à ce qui pouvait être déduit d'une condition de la forme

$$P^{n(x)}V(x) \leq \beta V(x) + b\mathbb{1}_{\mathcal{C}}(x) \quad (2.3)$$

concernant l'ergodicité du noyau P . La condition de drift (2.3) exprime que la chaîne, échantillonnée à des instants aléatoires $\{\tau^l, l \geq 0\}$ construits par la relation de récurrence $\tau^l = n(X_{\tau^{l-1}}) + \tau^{l-1}$ et $\tau^0 = 0$, vérifie une condition de drift de type géométrique.

Nous avons montré qu'une condition de drift de la forme (2.3) permet de contrôler des moments du temps de retour dans \mathcal{C} , et en particulier, des moments sous-géométriques.

Théorème 6. [4, Théorème 3.2] Supposons qu'il existe des fonctions $V : \mathbf{X} \rightarrow [1, +\infty)$ et $n : \mathbf{X} \rightarrow \mathbb{N}_*$, des constantes $\beta \in]0, 1[$ et $b < +\infty$, et un ensemble mesurable \mathcal{C} tels que $P^{n(x)}V(x) \leq \beta V(x) + b\mathbb{1}_{\mathcal{C}}(x)$. Si il existe une fonction strictement croissante $R :]0, +\infty[\rightarrow]0, +\infty[$ vérifiant une des conditions

- (i) $t \mapsto R(t)/t$ est décroissante et $R \circ n \leq V$.
 - (ii) R est une fonction convexe et continûment différentiable telle que R' est log-concave et $R^{-1}(V) - R^{-1}(\beta V) \geq n$
- alors il existe une constante c finie telle que $\mathbb{E}_x[R(\tau_{\mathcal{C}})] \leq c\{V(x) + b\mathbb{1}_{\mathcal{C}}(x)\}$.

Une condition suffisante pour vérifier le cas (i) est qu'il existe une fonction concave ξ telle que $\xi \circ n = V$; dans ce cas $R(t) \sim \xi(t)$. Puisque les fonctions concaves sont sous-linéaires, le cas (i) traite les situations où $n >> V$.

Une condition suffisante pour vérifier la relation $R^{-1}(V) - R^{-1}(\beta V) \geq n$ dans le cas (ii) est qu'il existe ξ telle que $\xi \circ n = V$ et $t \mapsto \xi(t)/t$ est décroissante; dans ce cas, $R^{-1}(t) \sim \int_1^t u^{-1}\xi(u)du$. Ainsi, le cas (ii) traite les situations où n/V est décroissante (en dehors d'un ensemble borné).

Ainsi, en fonction de la relation de domination entre la vitesse d'échantillonnage $x \mapsto n(x)$ et la fonction de drift $x \mapsto V(x)$, on peut déduire d'une inégalité de la forme (2.3) des moments du temps de retour $\tau_{\mathcal{C}}$ concaves ou convexes. Par exemple,

(taux logarithmiques) si $n(x) \propto W[\ln W]^{-\alpha}$ pour $\alpha > 0$ et $V \propto [\ln W]^{\alpha}$ alors la condition (i) est vérifiée avec $R(t) \sim [\ln t]^{\alpha}$;

(taux polynomiaux) si $n(x) \propto W^\alpha(x)$ pour $\alpha \in]0, 1]$ et $V(x) \propto W^{1-\alpha}(x)$ alors une des deux conditions s'applique selon le signe de $\alpha - 1/2$, et on a $R(t) \sim t^{1/\alpha-1}$;

(taux sous-exponentiels) si $n(x) \propto [\ln W(x)]^\alpha$ pour $\alpha > 0$ et $V(x) \propto W[\ln W]^\alpha$ alors la condition (ii) est vérifiée avec $R(t) \sim \exp(ct^{1/(1+\alpha)})$.

Le Théorème 6 combiné avec le Théorème 3 donne des conditions suffisantes d'ergodicité sous-géométrique. Nous nous sommes intéressés à la réciproque: si la chaîne est ergodique, peut-on établir une condition de drift de la forme (2.3) ? Nous avons obtenu le résultat

Proposition 7. [4, Théorème 2.1] *Supposons qu'il existe une fonction croissante $r : \mathbb{N} \rightarrow]0, +\infty[$ telle que $\lim_{n \rightarrow \infty} r = +\infty$, des fonctions mesurables $V, W : \mathcal{X} \rightarrow [1, +\infty[$ et une constante $C < +\infty$ telles que*

$$\forall (x, x') \in \mathcal{X} \times \mathcal{X}, \quad r(k) \|P^k(x, \cdot) - P^k(x', \cdot)\|_V \leq C\{W(x) + W(x')\}, \quad (2.4)$$

$$\exists x_0 \in \mathcal{X}, \quad \sup_{k \geq 0} P^k V(x_0) < +\infty. \quad (2.5)$$

Soit $\beta \in]0, 1[$ et $n : \mathcal{X} \rightarrow \mathbb{N}$ vérifiant $n(x) \geq r^{-1}([CW(x)] / [\beta V(x)])$ où r^{-1} désigne l'inverse généralisée du taux r . Alors pour tout $\beta' \in]\beta, 1[$, il existe un ensemble mesurable \mathcal{C} tels que $P^n(x) \leq \beta' V(x) + b \mathbb{1}_{\mathcal{C}}(x)$.

La condition (2.5) est en particulier vérifiée dès que P est psi-irréductible, apériodique et possède une unique probabilité invariante π telle que $\pi(V) < +\infty$. La condition (2.4) est liée à l'ergodicité de la chaîne : par exemple, lorsque W est bornée, (2.4) est une condition sur le coefficient de Dobrushin du noyau P^k et elle entraîne l'ergodicité uniforme $\sup_x \|P^n(x, \cdot) - \pi\|_{TV} \rightarrow 0$.

D'une part, sous les hypothèses de la Proposition 4, nous savons contrôler les moments sous-géométriques du temps de retour dans \mathcal{C} . D'autre part, sous ces mêmes hypothèses, nous avons montré que les conditions (2.4) et (2.5) sont vérifiées avec $r(n) = n$ et $V \leftarrow \phi \circ V$ [4, Proposition 2.3]. On peut alors appliquer la Proposition 7 et obtenir des contrôles de moments de temps de retour dans \mathcal{C} . La comparaison de ces deux méthodes, pour des formes standard de la fonction concave ϕ (par exemple, $\phi(t) \sim t^{1-\alpha}$, $\phi(t) \sim [\ln t]^\alpha$, $\phi(t) \sim t / [\ln t]^\alpha$) donne les mêmes résultats. En ce sens, bien que les conditions du Théorème 6 et de la Proposition 7 ne soient que des conditions suffisantes, elles semblent assez minimales.

2.1.3 Condition sur le modèle fluide d'une chaîne de Markov

Les techniques de limites fluides pour l'analyse des chaînes de Markov ont été introduites par [76] puis améliorées par [93]. Depuis, et en particulier ces dix dernières années, ces techniques sont très utilisées pour l'étude de la stabilité de réseaux complexes (cf. [45, 83] et références citées).

Nous avons étendu la technique des *limites fluides* à une classe de chaînes de Markov à valeur dans un espace euclidien \mathcal{X} de dimension d , dans le but d'étudier le comportement asymptotique de la chaîne à partir de l'étude des limites fluides associées.

La première étape consiste à construire une approximation fluide de la chaîne $\{X_n, n \geq 0\}$, par interpolation et renormalisation en temps, en espace et en le point initial de $\{X_n, n \geq 0\}$: pour $r, \alpha \geq 0$

on définit le processus à temps continu issu de x

$$\eta_r^\alpha(t; x) := \frac{X_{\lfloor tr^{1+\alpha} \rfloor}}{r}, \quad \eta_r^\alpha(t; 0) = \frac{X_0}{r} = x, \quad (2.6)$$

Sous certaines conditions, la famille de processus $\{\eta_r^\alpha, r \geq 0\}$ possède des valeurs d'adhérence - au sens d'une limite à préciser - qui constitue le **modèle fluide**. La seconde étape consiste à montrer que la chaîne initiale $\{X_n, n \geq 0\}$ est ergodique dès que le modèle fluide est **stable**. Enfin, il faut caractériser la stabilité du modèle fluide: nous avons considéré le cas où les limites fluides sont solutions d'une équation différentielle ordinaire (EDO), de sorte que la stabilité des limites fluides est liée à la stabilité de l'EDO.

L'intuition est en effet la suivante: on écrit

$$X_{k+1} = X_k + \Delta(X_k) + \epsilon_{k+1} \quad \text{avec } \Delta(x) = \mathbb{E}_x [X_1 - x] \quad \epsilon_{k+1} = X_{k+1} - \mathbb{E}[X_{k+1} | \mathcal{F}_k] \quad (2.7)$$

le processus $\{\epsilon_k, k \geq 0\}$ étant un incrément de martingale. D'après (2.6), le processus normalisé vérifie

$$\begin{aligned} \eta_r^\alpha \left[\frac{k+1}{r^{1+\alpha}}, x \right] &= \eta_r^\alpha \left[\frac{k}{r^{1+\alpha}}, x \right] + \frac{1}{r} \Delta \left(r \eta_r^\alpha \left[\frac{k}{r^{1+\alpha}}, x \right] \right) + \frac{1}{r} \epsilon_{k+1} \\ &= \eta_r^\alpha \left[\frac{k}{r^{1+\alpha}}, x \right] + \frac{1}{r^{1+\alpha}} h \left(\eta_r^\alpha \left[\frac{k}{r^{1+\alpha}}, x \right] \right) + \frac{\epsilon_{k+1}}{r} + \frac{1}{r^{1+\alpha}} \left\{ h \left(\eta_r^\alpha \left[\frac{k}{r^{1+\alpha}}, x \right] \right) - r^\alpha \Delta \left(r \eta_r^\alpha \left[\frac{k}{r^{1+\alpha}}, x \right] \right) \right\} \end{aligned}$$

en ayant posé $h(u) = \lim_{r \rightarrow +\infty} r^\alpha \Delta(r u)$. Le processus normalisé $t \mapsto \eta_r^\alpha(t, x)$ peut donc être vu comme une version bruitée de la relation déterministe

$$\mu \left(\frac{k+1}{r^{1+\alpha}} \right) = \mu \left(\frac{k}{r^{1+\alpha}} \right) + \frac{1}{r^{1+\alpha}} h \left(\mu \left(\frac{k}{r^{1+\alpha}} \right) \right), \quad \mu(0) = x,$$

issue de la discréétisation de l'EDO $\dot{\mu} = h(\mu)$ de condition initiale x . Par suite, l'étude du comportement limite (quand $r \rightarrow +\infty$) de la famille de processus $\{\eta_r^\alpha, r \geq 0\}$ est liée à celle de l'EDO $\dot{\mu} = h(\mu)$. Quant à la limite fluide, limite des processus η_r^α quand $r \rightarrow +\infty$, c'est en quelque sorte le "squelette" de la chaîne de Markov $\{X_k, k \geq 0\}$: par une renormalisation adéquate, on élimine les oscillations de la chaîne pour ne garder que son axe principal: la fonction h .

Le lien entre stabilité du modèle fluide et stabilité des chaînes de Markov nécessite bien sûr un peu plus de précisions que ce qui vient d'être fait. Nous donnons donc maintenant une preuve rigoureuse de cette intuition.

Définition du modèle fluide

Soit une chaîne de Markov $\{X_k, k \geq 0\}$ sur un espace euclidien X de dimension d , muni de sa tribu borélienne. On note $|\cdot|$ la norme euclidienne.

Soit $C(\mathbb{R}^+, X)$ l'espace des fonctions continues de \mathbb{R}_+ dans X , espace que l'on munit de la topologie locale uniforme. Soit $D(\mathbb{R}^+, X)$ l'espace des fonctions continues à droite et limitées à gauche (cadlag) de \mathbb{R}^+ dans X ; on munit cet espace de la topologie de Skorokhod. Pour $0 < T < +\infty$, soit $C([0, T], X)$ (resp. $D([0, T], X)$) l'espace des fonctions continues (resp. cadlag) définies sur $[0, T]$ à valeur dans X ; ces espaces sont munis de la topologie uniforme (resp. topologie de Skorokhod).

On note $\mathbb{Q}_{r;x}^\alpha$ la probabilité sur $D(\mathbb{R}^+, X)$, image de \mathbb{P}_x - la loi de la chaîne de Markov de loi initiale δ_x - par l'application $\eta_r^\alpha(\cdot; x)$ définie par (2.6).

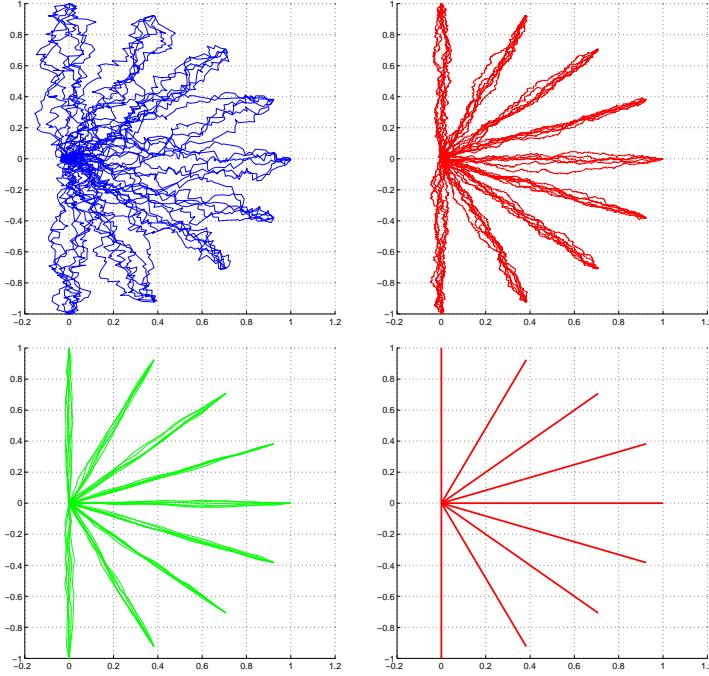


Fig. 2.1: Réalisations du processus $t \mapsto \eta_r^\alpha(t; x)$ pour différentes valeurs du point initial x sur la sphère unité de $[-\pi/2, \pi/2]$ et pour différents facteurs d'échelle r : $r \in \{100, 1000, 5000, +\infty\}$ (gauche à droite, et haut en bas)

Définition 8 (α -limite fluide). Soit $\alpha \geq 0$ et $x \in X$. Une probabilité \mathbb{Q}_x^α sur $D(\mathbb{R}^+, X)$ est une α -limite fluide si il existe une suite de facteurs d'échelle $\{r_n\} \subset \mathbb{R}_+$ et une suite de points initiaux $\{x_n\} \subset X$ vérifiant $\lim_{n \rightarrow \infty} r_n = +\infty$ et $\lim_{n \rightarrow \infty} x_n = x$ tels que $\{\mathbb{Q}_{r_n; x_n}^\alpha, n \geq 0\}$ converge faiblement vers \mathbb{Q}_x^α sur $D(\mathbb{R}^+, X)$ (ce que l'on notera $\mathbb{Q}_{r_n; x_n}^\alpha \Rightarrow \mathbb{Q}_x^\alpha$).

L'ensemble $\{\mathbb{Q}_x^\alpha, x \in X\}$ de ces limites fluides constitue le (α) -modèle fluide.

Sur la Figure 2.1 on trace quelques réalisations du processus $t \mapsto \eta_r^\alpha(t; x)$ à valeur dans \mathbb{R}^2 , pour différentes valeurs du point initial x sur la sphère unité de $[-\pi/2, \pi/2]$ et pour différents facteurs d'échelles. Ces processus sont produits à partir d'une chaîne de Hastings-Metropolis à marche aléatoire symétrique, de loi de proposition $q \sim \mathcal{N}_2(0, 4 \text{ Id})$ et de densité cible donnée par $\pi(x_1, x_2) \propto (1 + x_1^2 + x_2^2 + x_1^8 x_2^2) \exp(-(x_1^2 + x_2^2))$ (cf. section 3.1.1)

Existence du modèle fluide

Nous montrons l'existence d'un modèle fluide pour les chaînes dont la fonction Δ et l'incrément de martingale ϵ_k définis par (2.7) vérifient les conditions

H1 pour tout $x \in X$, $\mathbb{E}_x[|X_1|] < +\infty$.

H2 il existe $p > 1$ tel que $\lim_{K \rightarrow +\infty} \sup_{x \in X} \mathbb{E}_x [| \epsilon_1 |^p \mathbb{I}_{\{|\epsilon_1| \geq K\}}] = 0$.

H3 il existe $\beta \in [0, 1 \wedge (p - 1)]$ tel que $\sup_{x \in X} \{1 + |x|^\beta\} |\Delta(x)| < +\infty$.

La condition H3 implique que l'incrément moyen $\mathbb{E}_x[X_1 - x]$ est uniformément borné: nous ne définissons la limite fluide que pour de telles chaînes (dites *skip-free* dans la littérature).

Théorème 9. [16, Théorème 1.2] Supposons H1 à H3. Soit $x \in \mathbb{X}$.

- (i) Pour tout $\alpha \in [0, \beta]$, toutes suites $\{r_n\} \subset \mathbb{R}^+$ et $\{x_n\} \subset \mathbb{X}$ telles que $\lim_n r_n = +\infty$ et $\lim_n x_n = x$, il existe une mesure de probabilité \mathbb{Q}_x^α sur $C(\mathbb{R}_+, \mathbb{X})$, des sous-suites $\{r_{n_j}\}$ et $\{x_{n_j}\}$ de $\{r_n\}$ et $\{x_n\}$ telles que $\mathbb{Q}_{r_{n_j}; x_{n_j}}^\alpha \Rightarrow \mathbb{Q}_x^\alpha$.
- (ii) Pour tout $\alpha \in [0, \beta[$, les α -limites fluides sont triviales : $\mathbb{Q}_x^\alpha = \delta_g$ pour $g(t) \equiv x$.

La conclusion (ii) du Théorème 9 montre que si la chaîne n'est pas renormalisée en temps à la "bonne" vitesse, la limite fluide n'apporte aucune information et reste gelée en son point initial x . La "bonne" vitesse de normalisation est donnée par la vitesse à laquelle $\Delta(x)$ tend vers zero (cf. H3).

Stabilité du modèle fluide et ergodicité des chaînes de Markov

Le modèle fluide est dit stable si il existe $T > 0$ et $\rho \in]0, 1[$ tels que pour tout $x \in \mathbb{X}$, $|x| = 1$,

$$\mathbb{Q}_x^\alpha \left(\eta \in D(\mathbb{R}_+, \mathbb{X}), \inf_{0 \leq t \leq T} |\eta(t)| \leq \rho \right) = 1. \quad (2.8)$$

La stabilité du modèle fluide entraîne un contrôle du temps d'entrée de la chaîne dans une boule de rayon $\rho|X_0|$; on établit ([16, Proposition 4.8])

$$\limsup_{|x| \rightarrow +\infty} \mathbb{P}_x \left(\sigma > \lceil T|X_0|^{1+\beta} \rceil \right) = 0, \quad \text{où } \sigma := \inf \{k \geq 0, |X_k| < \rho|X_0|\}.$$

On en déduit une condition de drift géométrique pour la chaîne échantillonnée à l'instant aléatoire $\tau := \sigma \wedge \lfloor T|X_0|^{1+\beta} \rfloor$: $\sup_{x \in \mathbb{X}} |x|^{-p} \mathbb{E}_x [|X_\tau|^p] < 1$. On applique enfin des résultats sur l'ergodicité des chaînes quand un itéré du noyau $P^{n(x)}(x, \cdot)$ - l'indice d'itération dépendant de l'état courant de la chaîne - vérifie une condition de drift de type géométrique (cf. par exemple, section 2.1.2). On obtient

Théorème 10. [16, Théorème 1.4] Soit $\{X_k, k \geq 0\}$ une chaîne de Markov psi-irréductible, apériodique, telle que les compacts sont small. Supposons H1 à H3, et que le β -modèle fluide est stable. Alors pour tout $q \in [1, p/(1 + \beta)]$, la chaîne est V_q -ergodique à la vitesse polynomiale $(n + 1)^{q-1}$ où $V_q(x) = 1 + |x|^{p-q(1+\beta)}$.

Conditions suffisantes pour la stabilité du modèle fluide

Pour déterminer la stabilité du modèle fluide, il faut identifier les limites fluides ce qui nécessite de préciser le comportement de la fonction Δ

H4 il existe un cône ouvert $O \subseteq \mathbb{X} \setminus \{0\}$ et une fonction continue $h : O \rightarrow \mathbb{X}$ tels que pour tout compact $H \subseteq O$,

$$\lim_{r \rightarrow +\infty} \sup_{x \in H} |x|^\beta \left| r^\beta \Delta(rx) - h(x) \right| = 0.$$

Sous les conditions H1 à H4, nous montrons que les limites fluides restreintes à O ont une dynamique (déterministe) qui suit la trajectoire d'une EDO de champ moyen h ([16, Proposition 1.5]). En particulier, si

H5 pour tout $x \in O$, il existe $T_x > 0$ tel que l'EDO $\dot{\mu} = h(\mu)$ de condition initiale x a une unique solution sur $[0, T_x]$

alors la trajectoire de la limite fluide issue de $x \in O$ coïncide avec la solution de l'EDO de condition initiale x , notée $\mu(\cdot; x)$. Par suite, sous les conditions H1 à H5,

pour tout $x \in O$ il existe $T_x > 0$ tel que $\mathbb{Q}_x^\beta = \delta_{\mu(\cdot; x)}$ sur $D([0, T_x], X)$ ([16, Théorème 1.6]).

Dans le cas où le cône O est égal à $X \setminus \{0\}$, la stabilité du modèle fluide est impliquée par la stabilité des solutions de l'EDO, que l'on définit par l'existence de $\rho \in]0, 1[$ et $T < +\infty$ tel que pour tout $x \in X$ vérifiant $|x| = 1$, $\inf_{[0, T \wedge T_x]} |\mu(\cdot; x)| \leq \rho$. Dans le cas où $O \subsetneq X \setminus \{0\}$, la situation est plus complexe car certaines limites fluides ne sont pas caractérisées par les solutions de l'EDO; néanmoins, la stabilité des solutions de l'EDO peut encore entraîner celle du modèle fluide dans le cas où les limites fluides “rencontrent” le cône O . C'est ce qu'exprime la condition 2 ci-dessous.

Théorème 11. [16, Théorèmes 1.8 et 1.9] *Supposons H1 à H5. Si de plus l'une des deux conditions est vérifiée*

1. $O = X \setminus \{0\}$ et il existe $\rho \in]0, 1[$ et $T < +\infty$ tel que pour tout $x \in X$ vérifiant $|x| = 1$, la solution de l'EDO issue de x est telle que $\inf_{[0, T \wedge T_x]} |\mu(\cdot; x)| \leq \rho$;
2. $O \subsetneq X \setminus \{0\}$ et
 - (a) il existe $T_0 > 0$ tel que pour tout x , $|x| = 1$, et toute β -limite fluide \mathbb{Q}_x^β ,

$$\mathbb{Q}_x^\beta(\eta : \eta([0, T_0]) \cap O \neq \emptyset) = 1 ;$$

- (b) pour tout $K > 0$, il existe $T_K > 0$ et $0 < \rho_K < 1$ tels que pour tout $x \in O$, $|x| \leq K$, $\inf_{[0, T_K \wedge T_x]} |\mu(\cdot; x)| \leq \rho_K$;
- (c) pour tout compact $H \subset O$ et tout K , l'ensemble $\{\mu(t; x) : x \in H, t \in [0, T_x \wedge T_K]\}$ est un compact de O ;

alors le modèle fluide est stable.

Conclusion

Nous avons montré comment établir la V -ergodicité polynomiale d'une chaîne de Markov à partir de l'étude des limites fluides.

Vérifier que les limites fluides sont stables est certainement l'étape la plus difficile, particulièrement dans le cas où la condition H4 est vérifiée avec $O \subsetneq X \setminus \{0\}$. Etablir la condition 2a du Théorème 11 peut reposer sur des constructions particulièrement techniques, et très liées au problème considéré. C'est ce que nous avons observé en étudiant l'ergodicité de chaînes de Markov associées à quelques échantillonneurs MCMC; ces résultats seront présentés au Chapitre 3.

Comme nous l'avons vu dans les sections 2.1.1 et 2.1.2, une alternative à ces méthodes pour l'étude de l'ergodicité des chaînes est une approche par condition de drift. Néanmoins, l'avantage de l'approche “limite fluide/EDO” sur l'approche par condition de drift est que l'EDO porte une information très précise sur la dynamique de la chaîne, que le drift ne donne pas. Nous illustrerons ce point Chapitre 3, section 3.1, en montrant comment les paramètres d'implémentation de certains échantillonneurs MCMC apparaissent dans la définition du champ moyen h .

2.1.4 Condition de drift sur le générateur d'un processus de Markov fort

L'ergodicité d'un processus de Markov fort $\{X_t, t \in \mathbb{R}_+\}$ de transitions $\{P^t, t \in \mathbb{R}_+\}$ est aussi liée à la récurrence d'un ensemble \mathcal{C} petite. Il existe dans la littérature des critères d'ergodicité sous la forme de conditions de moments du temps de retour dans \mathcal{C} ; et sous la forme de conditions de drift vérifiée par le générateur du processus. Les résultats existant concernent la caractérisation de la Harris-récurrence, de la positive Harris-récurrence, de l'ergodicité, de la V -ergodicité et de la V -ergodicité géométrique ([86, 87, 54]).

Nous avons étendu ces résultats à la caractérisation de la V -ergodicité sous-géométrique. Soit X un espace métrique séparable et localement compact muni de sa tribu borélienne \mathcal{X} . On suppose que $\{X_t, t \in \mathbb{R}_+\}$ est un processus droit borélien, ce qui entraîne que c'est un processus de Markov fort, homogène dont les trajectoires sont continues à droite. Le premier résultat est formellement semblable à celui énoncé dans le Théorème 3 pour le cas des chaînes de Markov: il relie le contrôle de moments (V, r) -modulés du temps de retour dans un ensemble petite à la V -ergodicité sous-géométrique.

Un ensemble \mathcal{C} mesurable est dit *petite* si il existe une mesure de probabilité b sur \mathbb{R}_+ et une mesure de probabilité ν sur X , une constante $\epsilon > 0$ telles que

$$\int_0^{+\infty} P^t(x, \cdot) b(dt) \geq \mathbb{I}_{\mathcal{C}}(x) \epsilon \nu(\cdot).$$

On définit les taux sous-géométriques $r : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ de façon analogue à la Définition 1 - la seule différence est l'ensemble de définition de r -. Pour $\delta > 0$, on définit le temps de retour dans \mathcal{C} retardé de δ

$$\tau_{\mathcal{C}}(\delta) := \inf\{t \in [\delta, +\infty[, X_t \in \mathcal{C}\}.$$

Théorème 12. [22, Théorème 1] Soit $V : X \rightarrow [1, +\infty[$ une fonction mesurable et r_\star un taux sous-géométrique. Supposons que

(i) le processus est Harris-récurrent de mesure invariante π .

(ii) il existe un ensemble \mathcal{C} fermé et petite, et $\delta > 0$ tels que

$$\sup_{x \in \mathcal{C}} \mathbb{E}_x \left[\int_0^{\tau_{\mathcal{C}}(\delta)} V(X_s) ds \right] + \sup_{x \in \mathcal{C}} \mathbb{E}_x \left[\int_0^{\tau_{\mathcal{C}}(\delta)} r(s) ds \right] < +\infty. \quad (2.9)$$

(iii) il existe $m > 0$ tel que la chaîne de Markov à temps discret de noyau de transition P^m est psi-irréductible; et $\sup_{x \in X} \sup_{t \leq m} [P^t V(x)]/V(x) < +\infty$.

Alors π est une probabilité invariante telle que $\pi(V) < +\infty$ et pour π -presque tout x ,

$$\lim_{t \rightarrow +\infty} \|P^t(x, \cdot) - \pi\|_V + \lim_{t \rightarrow +\infty} r_\star(t) \|P^t(x, \cdot) - \pi\|_{TV} = 0. \quad (2.10)$$

De plus pour toute paire de fonctions $(\Psi_1, \Psi_2) \in \mathcal{Y}$ telle que $\Psi_1 \circ r_\star$ est un taux sous-géométrique, et pour π -presque tout x ,

$$\lim_{t \rightarrow +\infty} \Psi_1 \circ r_\star(t) \|P^t(x, \cdot) - \pi\|_{\Psi_2 \circ V} = 0. \quad (2.11)$$

La preuve de (2.10) consiste à montrer que la chaîne discrète de noyau P^m vérifie les hypothèses du Théorème 3 avec $(r, V) \leftarrow (r_*, 1)$; puis avec $(r, V) \leftarrow (1, V)$. On en déduit l'ergodicité sous-géométrique et la V -ergodicité de la chaîne de Markov; puis, en utilisant la relation $\sup_{x \in \mathbb{X}} \sup_{t \leq m} P^t V(x)/V(x) < +\infty$, on en déduit la même stabilité pour le processus de Markov à temps continu.

Le résultat (2.11) permet d'avoir des vitesses de convergence en \tilde{V} -norme pour une famille de fonctions \tilde{V} croissant au plus comme V . L'obtention de ce second résultat est semblable à ce qui a été fait dans le cas des chaînes de Markov à temps discret, section 2.1.1.

Nous avons ensuite énoncé des conditions en termes de contraction du générateur du processus, pour établir des contrôles de moments de la forme (2.9). Soit \mathbb{P}_x la probabilité canonique associée au processus de Markov de loi initiale δ_x . Notons $\mathcal{D}(\mathcal{A})$ l'ensemble des fonctions mesurables $f : \mathbb{X} \rightarrow \mathbb{R}$ telles que : il existe une fonction mesurable $h : \mathbb{X} \rightarrow \mathbb{R}$ telle que $t \mapsto h(X_t)$ est intégrable \mathbb{P}_x -p.s. et

$$t \mapsto f(X_t) - f(X_0) - \int_0^t h(X_s) \, ds,$$

est une \mathbb{P}_x -martingale locale pour tout x . Si $f \in \mathcal{D}(\mathcal{A})$, on note $\mathcal{A}f$ la fonction h associée. Le Théorème 13 est, pour les processus de Markov à temps continu, l'analogue de la Proposition 4 pour les chaînes de Markov à temps discret.

Théorème 13. [5, Théorèmes 3.4 et 4.1] *Soit $V : \mathbb{X} \rightarrow [1, +\infty[$ une fonction continue telle que $V \in \mathcal{D}(\mathcal{A})$; et $\phi : [1, +\infty[\rightarrow]0, \infty[$ une fonction concave, croissante, différentiable. Supposons qu'il existe un ensemble fermé et petite \mathcal{C} et une constante $b < +\infty$ tels que*

$$\mathcal{A}V \leq -\phi \circ V + b\mathbb{1}_{\mathcal{C}}.$$

Alors pour tout $\delta > 0$, il existe une constante $C < +\infty$ telle que pour tout $x \in \mathbb{X}$

$$\mathbb{E}_x \left[\int_0^{\tau_{\mathcal{C}}(\delta)} \phi \circ V(X_s) \, ds \right] \leq V(x) + b\delta, \quad \mathbb{E}_x \left[\int_0^{\tau_{\mathcal{C}}(\delta)} r_\phi(s) \, ds \right] \leq V(x) + \frac{b}{\phi(1)} \int_0^\delta r_\phi(s) \, ds,$$

où le taux r_ϕ est définie en Section 2.1.1.

Le Théorème 13 s'inscrit dans la continuité des résultats existant dans la littérature: [88] considèrent le cas où ϕ est constante (resp. $\phi(t) \sim t$) et établissent l'ergodicité (resp. l'ergodicité géométrique) du processus.

A notre connaissance, les seuls résultats relatifs à l'ergodicité sous-géométrique de processus de Markov fort sont dûs à [50] : ils montrent comment p conditions de drift emboitées entraînent la V_k -ergodicité polynomiale à la vitesse $r_k(t) \sim t^{p-k}$, pour $V_k(x) \sim 1 + \int_0^{+\infty} \int P^t(x, dy) |y|^k \exp(-t) dt$. Les résultats que nous avons obtenu étendent ces travaux relatifs à la V -ergodicité polynomiale au cas général de la V -ergodicité sous-géométrique. Ils permettent aussi d'avoir une expression plus explicite de la norme de convergence dans le cas de l'ergodicité polynomiale.

2.2 Contrôle explicite de V -ergodicité

Le contrôle explicite de V -ergodicité consiste à donner l'expression d'un taux $r = \{r(n), n \geq 0\}$ maximal et d'une fonction $B : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}^+$ minimale tels que

$$r(n) \|P^n(x, \cdot) - P^n(x', \cdot)\|_V \leq B(x, x'). \quad (2.12)$$

Lorsque B est $\delta_x \otimes \pi$ -intégrable pour une mesure de probabilité π invariante pour le noyau de transition P , l'inégalité (2.12) entraîne

$$r(n) \|P^n(x, \cdot) - \pi\|_V \leq \int \pi(dx') B(x, x') . \quad (2.13)$$

Dans le contexte de la simulation à dynamique markovienne, ces contrôles peuvent être utilisés comme critère d'arrêt des échantillonneurs, par exemple pour déterminer un nombre d'itérations au bout duquel la loi de la chaîne approche π avec une précision ϵ donnée. Une seconde application est d'exploiter la dépendance du taux de convergence et du contrôle B en les paramètres d'implémentation de l'échantillonneur, pour optimiser le choix de ces paramètres. Néanmoins, les différentes fonctions de contrôles B proposées dans la littérature ne sont pas suffisamment minimales pour donner des critères d'arrêt intéressants. La dépendance de r, B en les paramètres d'implémentation se révèle trop complexe pour être exploitable.

L'étude du comportement asymptotique des méthodes adaptatives à dynamique markovienne (par exemple, les échantillonneurs *Adaptive Metropolis* et *Equi-Energy*, sections 1.2.1 et 1.2.4) est en revanche une utilisation beaucoup plus crédible de ces bornes. En effet, nous verrons Chapitre 3 qu'un argument fondamental pour la convergence de ces méthodes est que les différents noyaux de transition $\{P_\theta, \theta \in \Theta\}$ utilisés dans la mise en œuvre de ces algorithmes aient un comportement ergodique comparable. Dans ce contexte, maîtriser la dépendance en le point initial x et la dépendance en θ dans une majoration de la forme (2.13) est nécessaire.

Une autre forme de contrôle explicite d'ergodicité est la construction d'un majorant de séries de la forme

$$\sum_{n \geq 0} r(n) \|P^n(x, \cdot) - P^n(x', \cdot)\|_V . \quad (2.14)$$

Lorsque ces majorants sont $\delta_x \otimes \pi$ -intégrables, pour une mesure de probabilité π invariante pour P , ces bornes permettent en particulier le contrôle des solutions des équations de Poisson (cf. [84, Chapitre 17] par exemple) associées aux fonctions f de V -norme finie.

2.2.1 Contrôle explicite par couplage

Nous avons proposé un contrôle explicite de la forme (2.2), (2.12) et (2.14). Cette activité de recherche est la suite de travaux effectués durant ma thèse sur le contrôle explicite d'ergodicité géométrique, polynomiale et plus généralement, sous-géométrique ([10, 19, 18]). Suite aux travaux sur les critères de V -ergodicité sous-géométrique décrits en section 2.1 et plus précisément, l'obtention d'une caractérisation par une seule condition de drift (cf. section 2.1.1), nous avons amélioré les contrôles d'ergodicité sous-géométrique présentés dans [10].

La construction de ces bornes est basée sur le **couplage** de chaînes: on définit un processus markovien $\{(X_n, X'_n, d_n), n \geq 0\}$ à valeur dans $\mathbb{X} \times \mathbb{X} \times \{0, 1\}$ sur un espace de probabilité muni de la probabilité $\bar{\mathbb{P}}$ tel que $P^n(x, \cdot) = \bar{\mathbb{P}}_{x,x',0}(X_n \in \cdot)$, $P^n(x', \cdot) = \bar{\mathbb{P}}_{x,x',0}(X'_n \in \cdot)$, et

$$\bar{\mathbb{P}}_{x,x',0}(X_n = X'_n, \forall n \geq T) = 1 , \quad \text{où} \quad T := \inf\{n \geq 0, d_n = 1\} . \quad (2.15)$$

On a alors l'inégalité de couplage

$$\|P^n(x, \cdot) - P^n(x', \cdot)\|_V \leq \sup_{|f|_V \leq 1} \bar{\mathbb{E}}_{x,x',0} [\{f(X_n) - f(X'_n)\} \mathbb{1}_{T > n}] ;$$

les bornes que nous proposons sont donc liées aux moments du temps de couplage T . La construction de ce processus Markovien $\{(X_n, X'_n, d_n), n \geq 0\}$ est possible sous une hypothèse de minoration du noyau de transition

A1 il existe un ensemble mesurable $\Delta \subset \mathbb{X} \times \mathbb{X}$, une mesure de probabilité ν sur \mathbb{X} et une constante $\varepsilon \in]0, 1[$ tels que

$$P(x, \cdot) \wedge P(x', \cdot) \geq \mathbb{I}_\Delta(x, x') \varepsilon \nu ;$$

une condition qui peut être affaiblie et remplacée par une condition de Doeblin locale (cf. [11] par exemple). Etant donnée (X_n, X'_n, d_n) , la transition du processus est décrite par

1. si $d_n = 1$, simuler $X_{n+1} \sim P(X_n, \cdot)$ et poser $X'_{n+1} = X_{n+1}$, $d_{n+1} = 1$.
2. si $d_n = 0$ et $(X_n, X'_n) \in \Delta$
 - (i) avec probabilité ε , simuler $X_{n+1} \sim \nu$ et poser $X'_{n+1} = X_{n+1}$ et $d_{n+1} = 1$.
 - (ii) avec probabilité $1 - \varepsilon$, poser $d_{n+1} = 0$ et simuler $(X_{n+1}, X'_{n+1}) \sim R(X_n, \cdot) \otimes R(X'_n, \cdot)$ où R désigne le noyau résiduel $R(x, \cdot) := [1 - \varepsilon]^{-1}\{P(x, \cdot) - \varepsilon\nu\}$.

Le temps de couplage T est donc lié aux passages successifs dans Δ .

Nous avons donc obtenu deux formes de bornes d'ergodicité. Dans la première, nous exprimons ces bornes à l'aide de moments du temps d'atteinte de Δ . Les moments de temps d'atteinte sont liés aux solutions de conditions de drift (cf. section 2.1.1, Proposition 4 par exemple); nous avons montré comment obtenir un majorant de ces moments à partir d'une condition de drift sur le noyau P , dans le cas où Δ est un ensemble symétrique $(\mathcal{C} \times \mathcal{D}) \cup (\mathcal{D} \times \mathcal{C})$ pour deux ensembles small $\mathcal{C} \subseteq \mathcal{D}$. Nous en avons déduit une seconde expression des bornes d'ergodicité.

2.2.2 Contrôle explicite d'ergodicité géométrique

Nous nous sommes intéressés aux contrôles de quantités de la forme (2.12) et (2.14). Nous présentons ci-dessous les résultats relatifs à l'ergodicité au sens (2.12); les autres résultats sont détaillés dans [11].

Soit $\beta_\star > 1$; on définit

$$U_\star(x, x') := \overline{\mathbb{E}}_{x, x', 0} [\beta_\star^\sigma] , \quad \text{où} \quad \sigma := \inf\{n \geq 0, (X_n, X'_n) \in \Delta\} ,$$

et

$$RU_\star(x, x') := (1 - \varepsilon)^{-2} \int_{\mathbb{X}^2} \{P(x, dy) - \varepsilon\nu(dy)\} \{P(x', dy') - \varepsilon\nu(dy')\} U_\star(y, y')$$

Théorème 14. [11, Théorème 1] Supposons A1 et qu'il existe $\beta_\star > 1$ tel que $\sup_\Delta RU_\star < +\infty$. Alors pour tout $x, x' \in \mathbb{X}^2$

$$\frac{1}{2} \|P^n(x, \cdot) - P^n(x', \cdot)\|_{\text{TV}} \leq 1 \wedge \{\beta_\star^{-n} U_\star(x, x')\}^\alpha ,$$

où $\alpha := \inf\{1; -\ln[1 - \varepsilon]/\ln[\beta_\star \sup_\Delta RU_\star]\}$.

Dans le cas particulier où l'espace X est small, la condition A1 est vérifiée avec $\Delta = \mathsf{X}^2$. Dans ce cas, $\sigma = 0$, $U_\star(x, x') = RU_\star(x, x') = 1$ de sorte que le Théorème 14 s'applique avec $\beta_\star = (1 - \varepsilon)^{-1}$ et $\alpha = 1$. On retrouve alors un résultat bien connu d'uniforme ergodicité, prouvé habituellement par un argument de contraction du noyau P au sens de Dobrushin (cf. par ex. [84, Chapitre 16]).

Sous les mêmes hypothèses, nous avons proposé des bornes de V -ergodicité pour des fonctions V telles que $V(x) + V(x') \leq U_\star(x, x')$. Ces résultats peuvent être consultés dans [11].

L'existence de $\beta_\star > 1$ tel que $\sup_\Delta RU_\star < +\infty$ et $U_\star(x, x') < +\infty$ pour au moins un couple $(x, x') \notin \Delta$ peut s'établir à partir d'une condition de drift sur le noyau produit $P \otimes P$; et sur le noyau P pour certaines formes de l'ensemble Δ . Sous l'hypothèse

A2 Il existe $V : \mathsf{X} \rightarrow [1, +\infty)$, des ensembles $(1, \nu, \varepsilon)$ -small $\mathcal{C} \subseteq \mathcal{D}$, des constantes positives $b < +\infty$ et $\lambda \in]0, 1[$ tels que $\sup_{\mathcal{D}} PV < +\infty$ et

$$PV \leq \lambda V + b \mathbb{1}_{\mathcal{C}}, \quad (2.16)$$

$$\beta_\star^{-1} := \lambda + b \left\{ \inf_{\mathcal{C}} V + \inf_{\mathcal{D}^c} V \right\}^{-1} < 1. \quad (2.17)$$

nous montrons que pour β_\star défini par (2.17) et pour Δ égal à $[\mathcal{C} \times \mathcal{D}] \cup [\mathcal{D} \times \mathcal{C}]$, le moment U_\star est majoré par $\{V(x) + V(x')\}/2$ ([11, Proposition 2]). Nous en déduisons donc le théorème

Théorème 15. [11, Corollaire 4] Supposons A2. Pour tout $(x, x') \in \mathsf{X} \times \mathsf{X}$,

$$\frac{1}{2} \|P^n(x, \cdot) - P^n(x', \cdot)\|_{\text{TV}} \leq 1 \wedge (\beta_\star^{-n} \{V(x) + V(x')\}/2)^{\hat{\alpha}}$$

et

$$\|P^n(x, \cdot) - P^n(x', \cdot)\|_V \leq \beta_\star^{-n} \{V(x) + V(x')\} \left(1 + C \beta_\star^{-n(1-\hat{\alpha})} \{V(x) + V(x')\}^{\hat{\alpha}} \right),$$

où $\hat{\alpha} := 1 \wedge (-\ln[1 - \varepsilon]/[\ln \beta_\star J])$ et $J := (1 - \varepsilon)^{-2} \sup_{\mathcal{C} \times \mathcal{D}} \{V(x) + V(x')\}/2$.

Pour établir A2, on commence par vérifier la condition de drift (2.16) pour un ensemble \mathcal{C} small tel que $\sup_{\mathcal{C}} V < +\infty$. On en déduit ensuite que les ensembles de niveau de V sont small; par suite, si $\sup_{\mathsf{X}} V = +\infty$, la condition (2.17) est vérifiée pour un ensemble $\mathcal{D} \supseteq \mathcal{C}$ assez grand. Néanmoins, il peut être difficile de démontrer que \mathcal{D} (et donc \mathcal{C}) est (m, ν, ε) -small avec $m = 1$. Dans [11], nous avons établi des bornes d'ergodicité dans le cas où $m > 1$; pour simplifier l'énoncé des résultats, nous ne présentons ici que le cas $m = 1$.

Ce travail s'inscrit dans une littérature très fournie sur le contrôle explicite de convergence dans le cas géométrique: [33, 73, 77, 80, 89, 104, 106, 107] pour la convergence en variation totale et [85] pour la convergence en norme V . L'originalité de notre travail a été (i) d'expliciter ces bornes dans le cas où la condition de minoration et la condition de drift ne portent pas sur le même itéré du noyau; et (ii) de proposer des bornes dans le cas de l'ergodicité en norme V par une approche différente de celle proposée par [85]. D'autres contributions sur ce sujet ont été développées dans le même temps [52] ou postérieurement [32].

2.2.3 Contrôle explicite d'ergodicité sous-géométrique

Nous nous sommes intéressés aux contrôles de quantités de la forme (2.12) et (2.14) pour des taux r sous-géométriques. Nous présentons ci-dessous les résultats relatifs à l'ergodicité au sens (2.12) pour les taux polynomiaux; les autres résultats sont détaillés dans [1].

Soit $a > 0$ et la suite à croissance polynomiale $r = \{(n+1)^a, n \geq 0\}$. On définit

$$U_\star(x, x') = \overline{\mathbb{E}}_{x, x', 0} \left[\sum_{k=0}^{\sigma} r(k) \right] \quad \text{où } \sigma := \inf\{n \geq 0, (X_n, X'_n) \in \Delta\}.$$

et

$$RU_\star := (1 - \varepsilon)^{-2} \int_{\mathsf{X}^2} \{P(x, dy) - \varepsilon\nu(dy)\} \{P(x', dy') - \varepsilon\nu(dy')\} U_\star(y, y')$$

Théorème 16. [1, Lemme 4.2] Supposons A1 et qu'il existe $a > 0$ tel que $\sup_{\Delta} RU_\star < +\infty$. Pour tout $(x, x') \in \mathsf{X} \times \mathsf{X}$,

$$\frac{1}{2} \|P^n(x, \cdot) - P^n(x', \cdot)\|_{\text{TV}} \leq (n+1)^{-a} (U_\star(x, x') + C)^{a/(1+a)},$$

pour une constante C qui dépend de ε, a et $\sup_{\Delta} RU_\star$.

L'expression de C en fonction de ε, a et $\sup_{\Delta} RU_\star$ est donnée dans [1, Lemme 4.2].

L'existence de $a > 0$ tel que $\sup_{\Delta} RU_\star < +\infty$ et $U_\star(x, x') < +\infty$ pour au moins un couple $(x, x') \notin \Delta$ peut s'établir à partir d'une condition de drift sur le noyau produit $P \otimes P$; ou simplement sur le noyau P pour certaines formes de l'ensemble Δ . Sous les hypothèses

A3 Il existe un ensemble \mathcal{C} , une fonction $V : \mathsf{X} \rightarrow [1, +\infty)$, des constantes $b, c > 0, \alpha \in]0, 1[$ tels que $\sup_{\mathcal{C}} V < +\infty$ et

$$PV(x) \leq V(x) - c V^{1-\alpha}(x) + b \mathbb{I}_{\mathcal{C}}(x),$$

A4 Il existe un ensemble \mathcal{D} $(1, \nu, \varepsilon)$ -small contenant \mathcal{C} tel que $\sup_{\mathcal{D}} V < +\infty$ et pour tout $(x, x') \in \mathcal{C} \times \mathcal{D}$,

$$\{V(x) + V(x')\}^{1-\alpha} - V^{1-\alpha}(x) - V^{1-\alpha}(x') + b \leq 0.$$

nous montrons que pour $a = 1/\alpha - 1$ et $\Delta = [\mathcal{C} \times \mathcal{D}] \cup [\mathcal{D} \times \mathcal{C}]$, le moment U_\star est majoré par $c^{1/\alpha} \{V(x) + V(x')\}$. La condition A4 est une condition sur la croissance de V à l'extérieur de \mathcal{D} : c'est donc l'analogue de la condition (2.17) pour le cas géométrique. Nous en déduisons

Théorème 17. Supposons A3 et A4. Pour tout $(x, x') \in \mathsf{X} \times \mathsf{X}$,

$$\frac{1}{2} \|P^n(x, \cdot) - P^n(x', \cdot)\|_{\text{TV}} \leq (n+1)^{-a} C (V(x) + V(x'))^{1-\alpha},$$

et pour tout $1 \leq l \leq \kappa \leq 1/\alpha$

$$\|P^n(x, \cdot) - P^n(x', \cdot)\|_{V^{1-\kappa\alpha}} \leq C (n+1)^{\kappa-l} (V^{1-l\alpha}(x) + V^{1-l\alpha}(x')),$$

pour une constante C qui dépend de ε, α et de $\sup_{\mathcal{C} \times \mathcal{D}} \{PV(x) + PV(x')\}$.

Lorsque la condition de drift A3 est relative à P mais que l'ensemble \mathcal{D} dans A4 est (m, ε, ν) -small pour $m > 1$, on peut encore obtenir des bornes d'ergodicité (cf. [1]).

A notre connaissance, en dehors de ces résultats et de ceux obtenus par [53] dans un travail contemporain, il n'existe pas d'autres contrôles explicites d'ergodicité sous-géométrique. Il est vrai que dans le cas sous-géométrique, les constantes exactes deviennent rapidement très complexes.

Conclusion

Les bornes d'ergodicité que nous avons obtenues ont un point commun avec celles que l'on trouve par ailleurs dans la littérature: elles se détériorent de façon drastique lorsque la valeur de la constante de minoration ε dans A1 tend vers zero. Dans la littérature, la construction des bornes explicites d'ergodicité est

(i) soit basée sur le **couplage** de deux composantes d'un processus markovien multidimensionnel (cf. par exemple [52]: le processus $\{(X_n, X'_n, d_n), n \geq 0\}$ est défini de façon plus générale que ce que nous avons présenté, une des propriétés essentielles étant que ses marginales vérifient la relation (2.15)).

(ii) soit basée sur le **renouvellement**: cette approche repose sur la construction d'une chaîne auxiliaire qui possède un atome, et dont l'étude de la stabilité (irréductibilité, apériodicité, existence d'une probabilité invariante, ergodicité, théorème central limite, ...) est équivalente à l'étude de la stabilité de la chaîne initiale. La construction des bornes explicites repose alors sur le temps d'atteinte de l'atome, qui lui-même s'exprime à l'aide des temps de retour dans un ensemble Δ . Plus précisément, chaque passage dans Δ est un passage dans l'atome, avec probabilité ϵ .

Ces deux approches sont donc très étroitement liées à la probabilité de succès ε dans un tirage de Bernoulli: le temps de couplage comme le temps d'atteinte de l'atome sont d'autant plus longs que ε est petit. Vues les valeurs très faibles de cette constante dans des applications non triviales, ces bornes sont assez peu adaptées à un contrôle fin de l'ergodicité. Dans une situation où l'enjeu est de comparer le comportement ergodique de plusieurs noyaux, elles peuvent néanmoins être utiles : nous les utiliserons Chapitre 3 dans ce contexte, pour établir la convergence de méthodes de simulation adaptatives.

2.3 Applications : vitesse de convergence d'algorithmes MCMC

2.3.1 Ergodicité sous-géométrique de l'algorithme Hastings-Metropolis à marche aléatoire symétrique

L'algorithme de Hastings-Metropolis à marche aléatoire symétrique, ci-après désigné par SRWM, est un algorithme de simulation d'une chaîne de Markov $\{X_n, n \geq 0\}$ de probabilité invariante $\pi d\lambda$; le noyau de transition est donné par

$$P(x, A) = \int_A \alpha(x, y) q(y - x) d\lambda(y) + \delta_x(A) \int (1 - \alpha(x, y)) q(y - x) d\lambda(y)$$

où $\alpha(x, y) = 1 \wedge [\pi(y)/\pi(x)]$ et $q d\lambda$ est une loi de proposition symétrique.

Dans la suite, nous supposons que π, q sont des densités sur \mathbb{R}^d et que λ est la mesure de Lebesgue sur \mathbb{R}^d .

Le ratio d'acceptation-rejet α est choisi de sorte que, quelle que soit la loi de proposition, la chaîne soit réversible. Cela entraîne que $\pi P = \pi$ et donc la chaîne possède par construction une probabilité invariante. Pour vérifier une condition de drift sur P , un choix (assez) naturel est de prendre V comme une fonction décroissante de π . En effet, on attend de l'algorithme qu'il dirige la chaîne vers des zones où π est grand et donc qu'en moyenne, en une itération, des fonctions décroissantes de π diminuent.

Avec ce choix de fonction V , l'existence d'une condition de drift du noyau est liée à la forme de la région d'acceptation $\mathcal{A}(x)$ définie comme l'ensemble des points y tels que $\alpha(x, y) = 1$

$$\mathcal{A}(x) := \{y \in \mathbb{X}, \pi(y) \geq \pi(x)\}.$$

On a en effet

$$\frac{PV(x)}{V(x)} = 1 + \int_{\mathcal{A}(x)-x} \left[\frac{V(x+y)}{V(x)} - 1 \right] q(y)d\lambda(y) + \int_{[\mathbb{R}^d \setminus \mathcal{A}(x)]-x} \frac{\pi(x+y)}{\pi(x)} \left[\frac{V(x+y)}{V(x)} - 1 \right] q(y)d\lambda(y)$$

où $A - x := \{y, x+y \in A\}$; de sorte que, puisque V est une fonction décroissante de π , le seul terme qui peut être négatif et peut traduire la contraction du noyau est

$$\mathcal{T}(x) := \int_{\mathcal{A}(x)-x} \left[\frac{V(x+y)}{V(x)} - 1 \right] q(y)d\lambda(y).$$

On trouve dans la littérature deux formes de fonction de drift V : $V \propto \pi^{-s}$, $s \in]0, 1[$ ou $V \propto [-\ln \pi]^s$, $s > 0$ ([17, 19, 103, 68, 70] et Théorème 18 ci-dessous)

L'étude de la vitesse d'ergodicité du SRWM est donc généralement menée sous deux groupes d'hypothèses. Le premier garantit l'irréductibilité et l'apériodicité de la chaîne. Le second permet de déterminer l'ordre de grandeur en x du terme de rappel $\mathcal{T}(x)$ et de vérifier que $[PV/V](x) - 1$ est négatif, au moins pour $|x|$ grand.

Durant ma thèse, je me suis intéressée à l'ergodicité d'échantillonneurs SRWM lorsque π est à décroissance polynomiale et plus généralement à variation régulière dans les queues; dans ce cas, le SRWM est V -ergodique à une vitesse polynomiale [10, 17, 19]. Nous avons poursuivi cette recherche en établissant des conditions sur (π, q) impliquant l'ergodicité à des taux sous-géométriques. Nous avons étudié cet échantillonneur sous les hypothèses suivantes

- B1** (i) la loi de proposition q est symétrique et il existe $\delta_q, \varepsilon_q > 0$ tels que $|x| \leq \delta_q \Rightarrow q(x) \geq \varepsilon_q$.
(ii) la loi de proposition est à support compact

- B2** π est continue et positive sur \mathbb{R}^d .

Sous les hypothèses B1(i) et B2 la chaîne est psi-irréductible, apériodique et les ensembles compact sont small ([103, Théorème 2.2]). L'hypothèse de compacité du support de la loi de proposition q peut être relâchée; elle n'est faite que pour simplifier l'étude et pourrait être substituée par des conditions de moments.

Dans le cas $d = 1$, [80] montre que l'algorithme SRWM est V -ergodique à une vitesse géométrique quand la loi de proposition q satisfait des conditions de moment et que la densité cible π est continue, positive et log-concave dans les queues. Cette condition est nécessaire dans le sens où, si la chaîne est V -géométriquement ergodique alors il existe $s > 0$ tel que $\int \exp(s|z|) \pi(z)d\lambda(z) < +\infty$. Ces résultats ont été étendus au cas multi-dimensionnel par [103] et [68]. Nous considérons des densités cibles π qui n'ont pas de moments géométriques, et décroissent sous-exponentiellement vite dans les queues. La classe de densité décrite ci-après contient en particulier, les densités de Weibull sur \mathbb{R} (pour des exemples de densités sur \mathbb{R}^d , cf. par exemple [17]).

B3 il existe $m \in]0, 1[$, $r \in]0, 1[$ et des constantes strictement positives $d_i, D_i, i \in \{0, 1, 2\}$ telles que pour tout x assez grand, π est deux fois continûment différentiable et

$$\begin{aligned} \left\langle \frac{\nabla \pi(x)}{|\nabla \pi(x)|}, \frac{x}{|x|} \right\rangle &\leq -r, \\ d_0|x|^m &\leq -\ln \pi(x) \leq D_0|x|^m, \\ d_1|x|^{m-1} &\leq |\nabla \log \pi(x)| \leq D_1|x|^{m-1}, \\ d_2|x|^{m-2} &\leq |\nabla^2 \log \pi(x)| \leq D_2|x|^{m-2}. \end{aligned}$$

Théorème 18. [7, Théorème 3.1] Supposons B1, B2 et B3. Alors la chaîne est psi-irréductible apériodique et il existe une boule compacte et small \mathcal{C} et des réels $s, b, c > 0$ tels que $PV \leq V - \phi \circ V + b\mathbb{I}_{\mathcal{C}}$ avec

$$V = \pi^{-s}, \quad \phi(x) = cx(1 + \ln x)^{-2(1-m)/m}.$$

Par application des résultats présentés en Section 2.1.1, nous en déduisons qu'il existe $s, c > 0$ tels que le SRWM est ergodique en norme π^{-s} à la vitesse sous-géométrique $\exp(cn^{m/(2-m)})$.

2.3.2 Ergodicité sous-géométrique des diffusions de Langevin - Algorithmes de simulation MALA

Les diffusions de Langevin sont des diffusions solution d'une équation différentielle stochastique de la forme $dX_t = b(X_t) dt + \sigma(X_t) dW_t$ où étant donnée une loi de densité π sur \mathbb{R}^d et une matrice $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$, le vecteur de drift $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ vérifie

$$b_i(x) = \frac{1}{2} \sum_{j=1}^d a_{i,j}(x) \partial_{x_j} \ln \pi(x) + \frac{1}{2} \sum_{j=1}^d \partial_{x_j} a_{i,j}(x), \quad \text{avec } a(x) := \sigma(x) \sigma^T(x).$$

Ces diffusions, quand existent et sont ergodiques, admettent π pour mesure invariante. Notons que la définition de cette diffusion ne demande la connaissance de π qu'à une constante multiplicative près.

Les algorithmes de Hastings-Metropolis de type *Metropolis-Adjusted Langevin Algorithm* (MALA) sont basés sur le choix d'un noyau de proposition q qui mime la dynamique de cette diffusion

$$q(x, \cdot) = \mathcal{N}\left(x + \frac{l^2}{2} \nabla \log \pi(x), l^2 \text{Id}_d\right),$$

le facteur d'échelle l étant un paramètre d'implémentation laissé au libre-arbitre de l'utilisateur, mais dont dépend bien sûr l'efficacité de l'algorithme. L'échantillonneur MALA est donc formellement assez proche d'un algorithme de Hastings-Metropolis à marche aléatoire symétrique dans lequel la loi de proposition serait $q(x, \cdot) = \mathcal{N}(x, l^2 \text{Id}_d)$: il diffère par l'introduction d'un 'biais' dans la loi de proposition, qui a pour effet de pousser la chaîne dans la direction $\nabla \ln \pi(X_n)$ et donc vers des valeurs plus grandes de π . On s'attend donc à de meilleures propriétés de convergence pour l'algorithme MALA.

La loi de proposition d'un algorithme MALA peut être vue comme une discrétisation de la diffusion de Langevin en prenant $\sigma = l \text{Id}_d$. Certaines discrétisations héritent des mêmes propriétés de convergence vers leur loi stationnaire que celles de la diffusion vers π [101, 102, 109]. Néanmoins, la chaîne obtenue par discrétisation du processus à temps continu n'a pas π pour mesure invariante,

ce qui justifie d'introduire les réalisations du processus discréétisé dans un schéma de correction type Hastings-Metropolis afin de produire une chaîne de Markov de mesure stationnaire π . Des travaux récents ont montré que le noyau de Hastings-Metropolis pouvait hériter des propriétés de convergence du noyau de proposition q [110].

Ce lien entre propriétés de convergence de la diffusion vers π et propriétés de convergence de l'algorithme MALA, explique l'intérêt pour accélérer la vitesse de convergence des diffusions par un choix astucieux de la matrice σ . L'idée est de choisir σ de sorte que σ est faible quand le processus est proche des modes de π et grand quand il est loin des modes.

Pour ces raisons, nous nous sommes intéressés au cas où $\sigma(x)$ est égal à $\pi^{-s}(x) \text{Id}_d$, $s > 0$, ou $\sigma(x) = |\ln \pi(x)|^s \text{Id}_d$, $s > 0$, et à l'influence de s sur la vitesse de convergence de la diffusion. Les premiers travaux dans cette direction sont dûs à [109, 101] et traitent de diffusion sur \mathbb{R} . Nous avons considéré le cas multidimensionnel, tout d'abord lorsque π est à décroissance polynomiale dans les queues ([22]) puis lorsque π est à décroissance sous-exponentielle dans les queues ([5]). Dans ces deux situations, nous avons montré que en deçà d'une certaine “température critique” s_* , la diffusion ne pouvait pas être géométriquement ergodique, et qu'elle était sous-géométriquement ergodique. Au-delà de cette température critique s_* , la diffusion est géométriquement ergodique. Nous détaillons maintenant plus précisément sous quelles conditions sur π ces résultats sont obtenus.

C1 π est une densité (connue à une constante multiplicative près) sur \mathbb{R}^d , strictement positive et deux fois continûment différentiable et $\nabla_{x_i x_j}^2 \log \pi$ est localement uniformément Hölder-continue.

C2 il existe $\beta \in]0, 1/d[$ tel que

$$\begin{aligned} 0 < \liminf_{|x| \rightarrow +\infty} \frac{|\nabla \log \pi(x)|}{\pi^\beta(x)} &\leq \limsup_{|x| \rightarrow +\infty} \frac{|\nabla \log \pi(x)|}{\pi^\beta(x)} < +\infty, \\ 2\beta - 1 < \liminf_{|x| \rightarrow +\infty} \frac{\text{Tr}[\nabla^2 \log \pi(x)]}{|\nabla \log \pi(x)|^2} &\leq \limsup_{|x| \rightarrow +\infty} \frac{\text{Tr}[\nabla^2 \log \pi(x)]}{|\nabla \log \pi(x)|^2} < +\infty. \end{aligned}$$

On pose $\gamma := \liminf_{|x| \rightarrow +\infty} \text{Tr}(\nabla^2 \log \pi(x)) |\nabla \log \pi(x)|^{-2}$.

Cette famille décrit notamment des densités à décroissance polynomiale; elle contient par exemple les densités de la forme $\pi(x) \sim |x|^{-1/\beta}$ pour $|x|$ grand; dans ce cas, $\gamma = \beta(2 - d)$.

Théorème 19. [22, Théorème 16] Soit π vérifiant C1 et C2. Soit la diffusion de Langevin sur \mathbb{R}^d de densité cible π , dont la matrice de diffusion est $\sigma(x) = \pi^{-s}(x)$ pour $s \geq 0$ tel que le processus est régulier.

(i) Si $s \in [0, \beta[$, le processus n'est pas géométriquement ergodique. Pour tout $\kappa \in [0, 1 + \gamma - 2\beta[$,

$$\lim_{t \rightarrow +\infty} (t+1)^p \|P^t(x, \cdot) - \pi\|_{1+\pi^{-\kappa}} = 0, \quad p < \frac{1 + \gamma - 2\beta - \kappa}{2(\beta - s)}$$

(ii) Si $\beta \leq s < (1 + \gamma)/2$ alors pour tout $\kappa \in]0, 1 + \gamma - 2s[$, la diffusion est V -géométriquement ergodique avec $V = 1 + \pi^{-\kappa}$.

(iii) Si $s > \beta$, alors la diffusion est uniformément ergodique.

La régularité du processus signifie que le temps d'explosion défini par $\lim_n \zeta_n$ où ζ_n est le temps de sortie des boules de rayon n , est infini presque-sûrement. Nous donnons des conditions suffisantes pour la régularité du processus (cf. [22, Paragraph 3.2]); voir aussi [66]. Par exemple, en dimension $d = 1$, il suffit que $\int \pi^{2s-1}(x)dx < +\infty$.

Dans la démonstration de ce théorème, on caractérise la vitesse de convergence vers π en établissant une condition de contraction sur le générateur, pour une fonction V adéquate: nous prenons $V(x) = 1 + \text{sign}(\rho)\pi^{-\rho}$ à l'extérieur d'un compact et $V = 1$ sinon. En particulier, les vitesses polynomiales s'obtiennent en appliquant les caractérisations présentées en Section 2.1.4.

Le Théorème 19 étend des résultats connus dans le cas $d = 1$ au cas multidimensionnel; et dans le cas de diffusions “froides” (i.e. s petit), il établit la V -ergodicité polynomiale. Dans le cas $s = 0$ et $d = 1$, [59] montrent que la diffusion est ergodique (en variation totale) à la vitesse polynomiale sans expliciter le taux de convergence. Le Théorème 19 traduit l'influence de la température s sur la vitesse de convergence; il fait apparaître une température critique $s_* = \beta$: cette température critique coïncide avec celle obtenue par [110] dans le cas unidimensionnel $d = 1$. Nous montrons que pour une norme π^{-p} donnée, la vitesse minimale de convergence est obtenue pour $s = 0$; dans ce cas, la diffusion de Langevin a la même vitesse de convergence que l'algorithme de Hastings-Metropolis à marche aléatoire symétrique dans le cas de loi de proposition ayant un moment d'ordre 2 [19, 70].

C3 il existe $0 < \beta < 1$ tel que pour tout $|x|$ grand : $|x|^{1-\beta} \langle \partial \ln \pi(x), \frac{x}{|x|} \rangle < 0$ et

$$0 < \liminf_{x \rightarrow \infty} |\partial \ln \pi(x)| |\ln \pi(x)|^{1/\beta-1} \leq \limsup_{x \rightarrow \infty} |\partial \ln \pi(x)| |\ln \pi(x)|^{1/\beta-1} < \infty,$$

$$\limsup_{x \rightarrow \infty} \text{Tr}(\partial^2 \ln \pi(x)) |\partial \ln \pi(x)|^{-2} = 0.$$

Cette famille décrit des densités à décroissance sous-exponentielle; elle contient par exemple les densités sur \mathbb{R} de type Weibull $\pi(x) \sim |x|^{\beta-1} \exp(-\alpha|x|^\beta)$ pour $\beta < 1$ et $\alpha > 0$.

Théorème 20. [5, Théorème 5.5] Soit π vérifiant C1 et C3. Soit la diffusion de Langevin sur \mathbb{R}^d de densité cible π et telle que $\sigma(x) = |\ln \pi|^s(x)$ pour $s \geq 0$.

(i) Si $0 \leq s < 1/\beta - 1$, le processus n'est pas géométriquement ergodique.

(ii) Si $0 \leq s < 1/\beta - 1$, le processus est V -ergodique à une vitesse sous-géométrique: en particulier, pour tout $\kappa \in]0, 1[$, il existe des constantes $b, c_\kappa > 0$ et un ensemble fermé petite \mathcal{C} tels que $\mathcal{A}V \leq -\phi \circ V + b\mathbb{I}_{\mathcal{C}}$ avec

$$V(x) = 1 + \pi^{-\kappa}(x), \quad \phi(x) = c_\kappa x(1 + \ln x)^{-2(1/\beta-1-s)}.$$

(iii) Si $s \geq 1/\beta - 1$, alors pour tout $0 < \kappa < 1$, la diffusion est V -géométriquement ergodique avec $V(x) = 1 + \pi^{-\kappa}(x)$.

(iv) Si $s > 1/\beta - (1/2)$, la diffusion est uniformément ergodique.

Dans la démonstration de ce théorème, on caractérise la vitesse de convergence vers π en établissant une condition de contraction sur le générateur, pour la fonction $V(x) = 1 + \pi^{-\kappa}(x)$ à l'extérieur d'un compact et $V = 1$ sinon. En particulier, les vitesses polynomiales s'obtiennent en appliquant les caractérisations présentées en Section 2.1.4.

Comme le Théorème 19, le Théorème 20 montre l'influence de la température s dans la vitesse de convergence de la diffusion vers le régime stationnaire. En particulier, pour une norme donnée, la vitesse est minimale lorsque $s = 0$ et dans ce cas, la diffusion converge (à une vitesse sous-géométrique) vers π à la même vitesse qu'un algorithme de Hastings-Metropolis à marche aléatoire symétrique (cf. Théorème 18 et l'item (ii) du Théorème 20 avec $s = 0$, par exemple dans le cas particulier où π est une loi de Weibull sur \mathbb{R}).

Chapitre 3

Optimisation des méthodes de simulation

Les résultats que nous présentons dans ce chapitre sont issus des articles

- [2] Y. Atchadé and G. Fort. Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. *Bernoulli*, 2009. Accepted for publication.
- [8] P. Etoré, G. Fort, B. Jourdain, and E. Moulines. On adaptive stratification. *Annals of Operations Research*, 2009. Accepted for publication.
- [16] G. Fort, S. Meyn, E. Moulines, and P. Priouret. The ODE method for the stability of skip-free Markov Chains with applications to MCMC. *Ann. Appl. Probab.*, 18(2):664–707, 2008.

et des rapports de recherche

- [3] Y. Atchadé, G. Fort, E. Moulines, and P. Priouret. Adaptive Markov chain Monte Carlo : Theory and Methods. Technical report, 2009.
- [13] G. Fort. Fluid limit-based tuning of some hybrid MCMC samplers. Technical report, 2007.

Motivation. Les méthodes de simulation dépendent de paramètres d’implémentation: par exemple, la loi instrumentale des échantillonneurs d’importance; ou la loi de proposition des échantillonneurs de Hastings-Metropolis. Le choix de ces paramètres n’est pas sans conséquences sur l’efficacité des échantillonneurs. Par suite, l’utilisateur doit avoir suffisamment d’expertise pour détecter des problèmes de convergence des algorithmes, et surtout, pour savoir corriger les paramètres d’implémentation pour résoudre ces problèmes.

Une alternative à cette expertise, est de mettre en oeuvre des algorithmes de Monte Carlo qui apprennent automatiquement la “bonne” valeur des paramètres d’implémentation. Pour définir de tels algorithmes, il faut savoir mesurer l’impact des paramètres d’implémentation sur la convergence des échantillonneurs. Il faut aussi savoir définir ce que sont les paramètres optimaux pour le problème considéré, afin de décrire une procédure qui adapte les paramètres d’implémentation pour les guider

vers ces valeurs optimales. Il faut enfin s'assurer que l'ajout d'une étape d'adaptation des paramètres d'implémentation ne détruit pas la convergence des échantillonneurs vers la loi cible.

Points forts de notre contribution. Nous présentons dans ce chapitre les travaux relatifs à l'optimisation des méthodes de simulation.

En section 3.1, nous montrons que le modèle fluide d'échantillonneurs MCMC porte une information très précise sur le rôle des paramètres d'implémentation dans la phase transiente de ces échantillonneurs. L'originalité de notre travail réside dans l'approche retenue pour mesurer l'impact des paramètres d'implémentation: nous étudions le modèle fluide, qui n'est rien d'autre que l'observation de la chaîne dans une certaine asymptotique.

En section 3.2, nous étudions la convergence des échantillonneurs adaptatifs à dynamique markovienne. Nous présentons tout d'abord en section 3.2.1 des conditions entraînant l'ergodicité de ces échantillonneurs i.e. la convergence de la loi du processus simulé vers la loi cible. Nous avons ensuite établi une Loi forte des Grands Nombres (section 3.2.2), établissant par là-même le bien-fondé de méthodes de Monte Carlo adaptatives. Ces résultats sont assez généraux: ils peuvent s'appliquer à l'étude d'échantillonneurs à adaptation interne (i.e. adaptation guidée par le seul passé du processus d'intérêt), et adaptation externe (i.e. adaptation guidée aussi par des processus auxiliaires). Ces résultats de convergence sont appliqués à l'étude de deux échantillonneurs adaptatifs, le *Adaptive Metropolis* et l'*Equi-Energy* qui sont resp. des exemples d'échantillonneurs à adaptation interne et externe (cf. sections 3.2.3 et 3.2.4). L'originalité de nos résultats est de proposer une unique théorie pour ces deux schémas d'adaptation (interne / externe), théorie a priori assez générale pour permettre l'étude de nombreux échantillonneurs adaptatifs.

En section 3.3, nous nous intéressons à l'optimisation d'un algorithme de réduction de variance: l'algorithme de stratification. En section 3.3.1, nous établissons tout d'abord la dépendance d'un critère d'efficacité de l'échantillonneur en les paramètres d'implémentation; le critère retenu ici est celui de la variance asymptotique de l'estimateur stratifié. Puis, en section 3.3.2, nous proposons une procédure adaptative pour le choix de ces paramètres d'implémentation, l'objectif de cette adaptation étant de guider les paramètres vers l'ensemble des minimiseurs de la variance asymptotique. Enfin, en section 3.3.3, nous validons l'intérêt de la procédure adaptative en considérant une application à l'ingénierie financière.

3.1 Etude de la phase transiente d'algorithmes MCMC

L'efficacité des échantillonneurs MCMC dépend du choix des paramètres d'implémentation. Par exemple, pour les algorithmes de Hastings-Metropolis, il faut choisir la variance de la loi de proposition, problème connu dans la littérature MCMC sous le nom de *scaling*.

La comparaison de différents algorithmes MCMC repose sur un critère d'efficacité. Pour une chaîne de Markov $\{X_n, n \geq 0\}$ de probabilité stationnaire π , et pour une fonction ϕ telle que $\pi(\phi^2) < +\infty$, on définit le temps d'autocorrélation intégré τ_ϕ par

$$\tau_\phi := 1 + 2 \sum_{k \geq 1} \text{Corr}(\phi(X_0), \phi(X_k)) , \quad X_0 \sim \pi .$$

Si un Théorème Central Limite est vérifié, la variance de l'estimateur $n^{-1} \sum_{k=1}^n \phi(X_k)$ de $\pi(\phi)$ est approximativement donnée par $n^{-1} \text{Var}(\phi(X_0)) \tau_\phi$. Cela suggère de comparer les algorithmes selon le temps d'autocorrélation intégré τ_ϕ . Néanmoins ce critère, et donc, le résultat de la comparaison, sont largement dépendants de la fonction ϕ d'intérêt.

Roberts et al. ont montré que lorsque la dimension de l'espace de simulation d tend vers $+\infty$, et pour certains algorithmes qui dans cette asymptotique se comportent comme des processus de diffusion, ces mesures d'efficacité sont indépendantes de ϕ [98]. Ces dix dernières années, cette remarque a suscité de nombreux travaux sur l'étude de la dépendance de la variance en la dimension de l'espace de simulation [34, 99, 90, 36]. Par exemple, Roberts et al. ont établi que, pour l'algorithme de Hastings-Metropolis à marche aléatoire symétrique (SRWM) de densité cible π sur \mathbb{R}^d de la forme $\prod_{k=1}^d f(x_k)$ et de loi de proposition $q = \mathcal{N}_d(0, l^2 d^{-\beta} \text{Id}_d)$, il faut prendre $\beta = 1$ pour que l'algorithme se comporte asymptotiquement comme une diffusion; le choix optimal du facteur d'échelle est $l = 2.38$ et cela entraîne un taux d'acceptation-rejet optimal α_* égal à 0.234; pour un algorithme de type Langevin, il faut prendre $\beta = 1/3$ et le taux d'acceptation-rejet optimal est $\alpha_* = 0.574$ [98].

Néanmoins, cette approche a quelques limites.

Ces résultats sont obtenus “à la limite”, lorsque $d \rightarrow +\infty$. Par suite, cela restreint l'ensemble des densités cibles π pour lesquelles on sait étudier les échantillonneurs; et cela donne des paramètres d'implémentation assez ‘optimaux’ dans le cas de la grande dimension mais moins intéressants sinon.

De plus, ces résultats sont établis dans le cas où $X_0 \sim \pi$: les propriétés de mélange de la chaîne sont donc étudiées dans sa phase stationnaire uniquement, et il est bien connu (empiriquement) que les trajectoires d'une chaîne dans sa phase “stationnaire” ne sont pas du même type que celles dans sa phase “transiente”. Dans le cas stationnaire, elles ressemblent à des processus de diffusion, tandis que dans la phase transiente, elles ont des trajectoires qui présentent peu de fluctuations autour d'un drift déterministe (cf. Figure 3.1 où le régime transient est de l'ordre de 100 itérations (gauche) et 4000 itérations (droite))

A notre connaissance, très peu de résultats existent relatifs à la phase transiente des algorithmes MCMC. Christensen et al. [48] ont étudié le comportement en phase transiente d'un SRWM et d'un algorithme de Langevin, lorsque la loi cible est $\pi \sim \mathcal{N}_d(0, \text{Id}_d)$ et dans l'asymptotique $d \rightarrow +\infty$.

Nous avons proposé une étude de la phase transiente des échantillonneurs à dynamique markovienne, en comparant les chaînes à un processus à temps continu, dans une asymptotique qui est indépendante de la dimension de l'espace de simulation. Cette approche est basée sur le modèle fluide associé à une chaîne de Markov dont nous avons montré que la stabilité entraînait l'ergodicité de la chaîne (cf. section 2.1.3 et article [16]). Nous avons mené cette étude pour des algorithmes de Hastings-Metropolis à marche aléatoire symétrique et, dans une moindre mesure, pour des algorithmes Metropolis-within-Gibbs. Nous étudions tout d'abord l'existence de ce modèle fluide, sa caractérisation par une équation différentielle ordinaire (EDO) et enfin la stabilité du modèle fluide. Les discussions qui suivent sont donc guidées par la vérification des hypothèses des Théorèmes 9, 10 et 11 présentés en section 2.1.3.

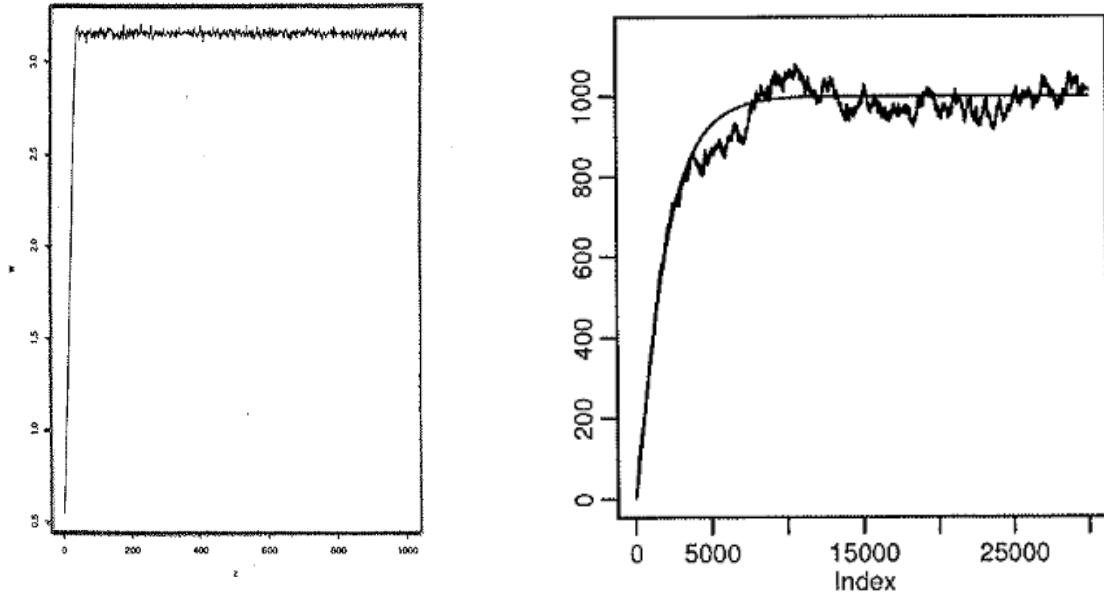


Fig. 3.1: Figures extraits de [48]. [Gauche] Trajectoire d'un SRWM lorsque π est une loi a posteriori pour l'estimation de paramètres d'une diffusion partiellement observée. [Droite] Evolution de $k \mapsto \|X_k\|^2$ pour une chaîne SRWM de loi cible $\pi \sim \mathcal{N}_{1000}(0, \text{Id})$ issue de l'origine ($X_0 = 0$)

3.1.1 Caractérisation des limites fluides d'un algorithme de Hastings-Metropolis

Nous étudions l'existence du modèle fluide de l'algorithme SRWM (cf. section 1.2.1) de densité cible π sur \mathbb{R}^d par rapport à la mesure de Lebesgue λ .

L'expression du noyau de transition des algorithmes SRWM entraîne que les conditions d'existence des limites fluides (cf. conditions H1 à H3, Section 2.1.3) sont vérifiées dès que la loi de proposition $q(y) d\lambda(y)$ possède un moment d'ordre p , $p > 1$. L'accroissement moyen $\Delta(x) = \mathbb{E}_x[X_1 - x]$ est en particulier donné par

$$\Delta(x) = \int_{\{y, \pi(x+y) \leq \pi(x)\}} y \left(\frac{\pi(x+y)}{\pi(x)} - 1 \right) q(y) \lambda(dy),$$

ce qui montre que Δ est toujours bornée dès que $\int |y| q(y) d\lambda(dy) < +\infty$.

La caractérisation des limites fluides nécessite de s'intéresser à la limite radiale de Δ : montrer l'existence d'un cône \mathcal{O} de $\mathbb{R}^d \setminus \{0\}$, d'une fonction $h : \mathcal{O} \rightarrow \mathbb{R}^d$ et d'un réel $\beta \in [0, 1 \wedge (p - 1)]$ tels que pour tout compact \mathcal{H} de \mathcal{O} on ait

$$\lim_{r \rightarrow +\infty} \sup_{x \in \mathcal{H}} |x|^\beta |r^\beta \Delta(rx) - h(x)| = 0; \quad (3.1)$$

cf. Section 2.1.3. Nous étudions l'existence de cette limite pour deux familles de densités cibles π

(i) des densités sur-exponentielles: densités strictement positives, dont les dérivées partielles d'ordre 1 sont continues et telles que

$$\lim_{|x| \rightarrow +\infty} \left\langle \frac{x}{|x|}, \nabla \log \pi(x) \right\rangle = -\infty.$$

(ii) des densités sous-exponentielles: densités strictement positives, dont les dérivées partielles d'ordre 1 sont continues et pour lesquelles il existe R tel que

$$\lim_{|x| \rightarrow +\infty} |\nabla \log \pi(x)| = 0, \quad \left\langle \frac{x}{|x|}, \frac{\nabla \log \pi(x)}{|\nabla \log \pi(x)|} \right\rangle < 0, \quad |x| \geq R.$$

La première famille est un exemple de densités pour lesquelles $\beta = 0$; pour la seconde famille en revanche, les β -limites fluides avec $\beta = 0$ sont triviales (cf. Théorème 9, Section 2.1.3) et nous considérerons donc des β -limites fluides pour $\beta > 0$. Dans les deux cas, l'existence de la limite (3.1) dépend du comportement radial des zones de rejet, et plus précisément de la mesure sous $q d\lambda$ de la différence symétrique entre la limite radiale dans la direction x des zones de rejet, et un hyperplan (limite). Pour simplifier cette étude, nous supposons que la loi de proposition est de la forme

$$q(x) = \det^{-1/2}(\Sigma) q_0(\Sigma^{-1/2}x), \quad (3.2)$$

où Σ est une matrice de variance-covariance inversible, et q_0 est une densité symétrique invariante par rotation (i.e. pour toute matrice unitaire U , $q_0(Ux) = q_0(x)$) vérifiant $\int_{\mathbb{R}^d} y_1^2 q_0(y) d\lambda(y) < +\infty$, $y = (y_1, \dots, y_d)$.

Dans la suite, nous noterons $A \ominus B$ la différence symétrique de deux ensembles A et B . R_x, A_x et C_x désigneront resp. les zones de rejet, d'acceptation et les lignes de niveau (cf. figure 3.2) translatées de x , et définies par

$$R_x := \{y, \pi(x+y) \leq \pi(x)\}, \quad \mathsf{A}_x := \mathbb{R}^d \setminus R_x, \quad \mathsf{C}_x := \{y, \pi(x+y) = \pi(x)\}.$$

Limites fluides pour des densités sur-exponentielles

- **Caractérisation des limites fluides** Nous avons montré ([16, Proposition 2.6]) que si
 - A. la densité cible π est sur-exponentielle dans les queues; la loi de proposition q vérifie (3.2).
 - B. il existe un cône ouvert \mathcal{O} de $\mathbb{R}^d \setminus \{0\}$ et une famille d'ensembles $\{R_{\infty,x}, x \in \mathcal{O}\}$ tels que pour tout compact H de \mathcal{O} $\lim_{r \rightarrow +\infty} \sup_{x \in \mathsf{H}} \mathbf{Q}(R_{rx} \ominus R_{\infty,x}) = 0$ ¹.
 - C. il existe une fonction continue $\ell_\infty : \mathbb{R}^d \setminus \{0\} \rightarrow \mathbb{R}^d$ such that, for all $x \in \mathcal{O}$,

$$\mathbf{Q}(R_{\infty,x} \ominus \{y \in \mathcal{X}, \langle y, \ell_\infty(x) \rangle < 0\}) = 0. \quad (3.3)$$

alors (3.1) est vérifiée avec $\beta = 0$ et

$$h(x) = m_1 \frac{\Sigma \ell_\infty(x)}{|\sqrt{\Sigma} \ell_\infty(x)|}, \quad m_1 := \int_{\mathbb{R}^d} y_1 \mathbf{1}_{\{y_1 \geq 0\}} q_0(y) \lambda(dy) > 0.$$

La condition B exprime que la famille de zones de rejet $\{R_{rx}, r \geq 0\}$ possède une q -limite radiale. La condition C exprime que cette q -limite radiale converge vers un hyperplan. Lorsque $\Sigma = \text{Id}$ et $\ell_\infty(x) = \lim_{r \rightarrow \infty} \{\nabla \ln \pi(rx) / |\nabla \ln \pi(rx)|\}$, l'EDO est une version d'un algorithme de gradient de plus grande pente pour maximiser $\log \pi$. Par exemple, lorsque la courbure des lignes de niveau C_{rx} tend vers 0 quand $r \rightarrow +\infty$, $\ell_\infty(x) = \lim_{r \rightarrow \infty} \{\nabla \ln \pi(rx) / |\nabla \ln \pi(rx)|\}$.

¹pour un ensemble A , on note $\mathbf{Q}(A)$ la mesure de A sous $qd\lambda$

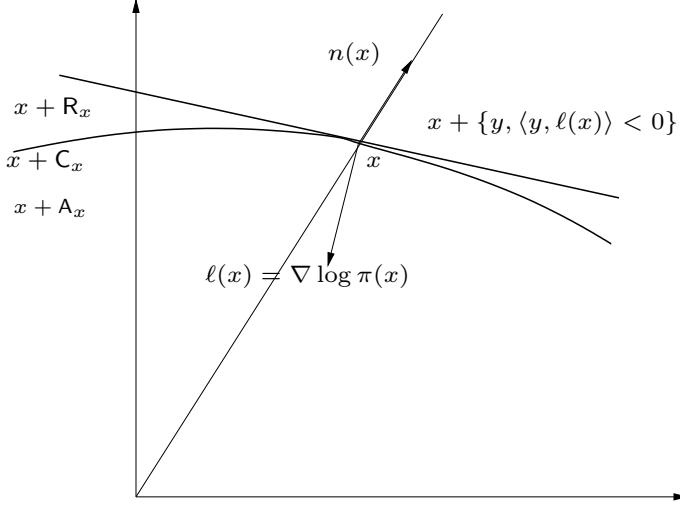


Fig. 3.2: Ligne de niveau $x + C_x$ délimitant la zone d'acceptation $x + A_x$ et la zone de rejet $x + R_x$. Comparaison de la zone de rejet et du demi-espace défini par l'hyperplan de vecteur normal $\ell(x) := \nabla \log \pi(x)$. $n(x)$ désigne le vecteur normé $x/|x|$.

► **Exemple 1: cas régulier** Considérons la densité cible dans \mathbb{R}^2

$$\pi(x_1, x_2) \propto (1 + x_1^2 + x_2^2 + x_1^8 x_2^2) \exp(-(x_1^2 + x_2^2)) , \quad (3.4)$$

et une loi de proposition gaussienne $q \sim \mathcal{N}_2(0, \sigma^2 \text{Id}_2)$. Les courbes de niveau de π sont représentées Figure 3.3[gauche] : elles sont presque circulaires, mais présentent des “pincentments” en les points $(x, 0)$. Néanmoins, cette densité vérifie les conditions A à C avec $O = \mathbb{R}^d \setminus \{0\}$ et $\ell_\infty = -x/|x|$. Par suite, $h(x) = -\sigma x / [\sqrt{2\pi} |x|]$. La fonction Δ et la limite h sont représentées sur la Figure 3.4[gauche]; du fait de la symétrie du problème par rapport à l'axe des ordonnées, nous représentons ces fonctions sur au moins un demi-espace.

La solution de l'EDO $\dot{\mu} = h(\mu)$ de condition initiale x est donnée par $\mu(t; x) = (|x| - \sigma t / \sqrt{2\pi}) \mathbb{I}\{\sigma t \leq \sqrt{2\pi} |x|\} x / |x|$. Cette solution est unique et est stable. Sur la Figure 3.4[droite], nous traçons quelques trajectoires de processus interpolé $t \mapsto \eta_r^0(t; x)$ (cf. Eq. 2.6), issues de points x sur la sphère unité de $[0, \pi/2]$ ainsi que la solution de l'EDO issue de ces points.

Par application des Théorèmes 10 et 11 (cf. Section 2.1.3), cet échantillonneur est ergodique à n'importe quelle vitesse polynomiale. Cet exemple a été introduit par [68] comme un exemple pour lequel les conditions suffisantes d'ergodicité qu'ils proposent, ne s'appliquent pas. Nous avons donc pu, par une approche limite fluide, prouver l'ergodicité d'un échantillonneur pour lequel des critères d'ergodicité plus classiques (basés sur les conditions de drift [103, 68]) ne s'appliquent pas.

Cette densité est un exemple de densités de la famille exponentielle de la forme $g(x) \exp(-p(x))$ deux fois continûment différentiables, où (a) g est une fonction strictement positive, à variation lente à l'infini, i.e. pour tout $K > 0$, $\limsup_{|x| \rightarrow \infty} \inf_{|y| \leq K} \frac{g(x+y)}{g(x)} = \limsup_{|x| \rightarrow \infty} \sup_{|y| \leq K} \frac{g(x+y)}{g(x)} = 1$; (b)

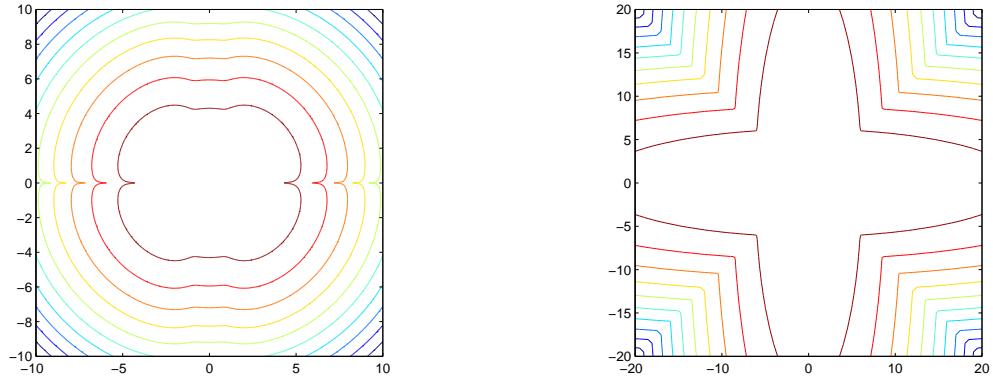


Fig. 3.3: [Gauche] Courbes de niveau de la densité (3.4) et [Droite] de la densité (3.5) lorsque $\delta = 0.4$

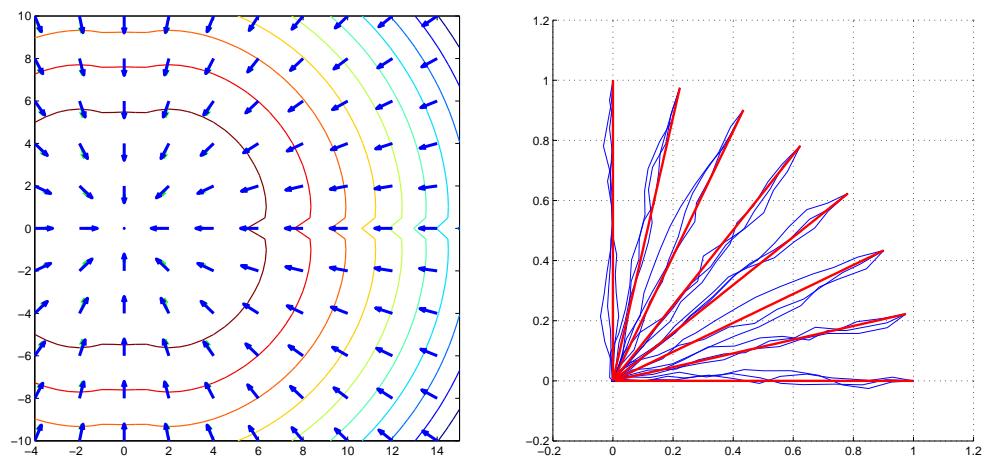


Fig. 3.4: Cas de la densité (3.4) [Gauche]: Fonctions Δ en vert et h en bleu. [Droite]: En bleu, trajectoires du processus (2.6) issues d'un ensemble de points sur la sphère unité de $[0, \pi/2]$; en rouge, les trajectoires de l'EDO.

p est une fonction polynomiale d'ordre m strictement positive et $\lim_{|x| \rightarrow \infty} p_m(x) = +\infty$, p_m désignant le polynôme constitué des termes d'ordre m .

Nous montrons que la conclusion obtenue pour la densité (3.4) reste vraie pour toute densité de cette famille : $\Omega = \mathbb{R}^d \setminus \{0\}$, $\ell_\infty(x) \propto -[\nabla p_m(x/|x|)]$ et la solution de l'EDO $\dot{\mu} = h(\mu)$ de condition initiale x , pour $x \in \Omega$, est unique sur $[0, T_x]$ et vérifie $\lim_{t \rightarrow T_x^-} \mu(t; x) = 0$ ([16, Proposition 2.7]).

► **Exemple 2: cas irrégulier** Considérons la densité cible sur \mathbb{R}^2 donnée par

$$\pi(x) \propto \alpha \exp(-(1/2)x'\Gamma_1^{-1}x) + (1-\alpha) \exp(-(1/2)x'\Gamma_2^{-1}x), \quad (3.5)$$

où $\alpha \in]0, 1[$, $\Gamma_1^{-1} := \text{diag}(a^2, 1)$ et $\Gamma_2^{-1} := \text{diag}(1, a^2)$ pour $a^2 > 1$; et une loi de proposition gaussienne $q \sim \mathcal{N}_2(0, \text{Id}_2)$. Les courbes de niveau sont représentées Figure 3.3[droite]: elles ont des points anguleux le long de la diagonale qui ne disparaissent pas à la limite (quand $r \rightarrow +\infty$) - la courbure des lignes de niveau $\{\mathbf{C}_{rx}, r \geq 0\}$ pour x sur la diagonale tend vers $+\infty$.

Cette densité vérifie les conditions A à C sur le cône $\Omega = \mathbb{R}^d \setminus \{x = (x_1, x_2) \in \mathbb{R}^2, |x_1| = |x_2|\}$ et

$$\ell_\infty(x) = \begin{cases} \Gamma_1^{-1}x & \text{si } |x_1| < |x_2|, \\ \Gamma_2^{-1}x & \text{si } |x_1| > |x_2|. \end{cases}$$

Par suite, $h(x) = -c_q \Gamma_i^{-1}x / [\Gamma_i^{-1}x]$ où $i = 1$ si $|x_1| < |x_2|$ et $i = 2$ si $|x_1| > |x_2|$, et c_q est une constante qui dépend de la loi de proposition q ([16, Lemme 2.9]). La fonction Δ et la limite h sont représentées sur la figure 3.5[gauche]; du fait de la symétrie du problème, nous ne représentons ces fonctions que sur l'orthant supérieur droit.

Dans cet exemple, le cône Ω n'est qu'un sous-ensemble de $\mathbb{R}^d \setminus \{0\}$. Pour $x \in \Omega$, il existe une unique solution à l'EDO $\dot{\mu} = h(\mu)$ de condition initiale x et ces solutions sont stables. En revanche, il n'y a pas unicité lorsque x est un point de la diagonale. Sur la Figure 3.5[droite], nous traçons quelques trajectoires de processus interpolé $t \mapsto \eta_r^0(t; x)$ (cf. Equation 2.6) issues de points x de la sphère unité sur $[0, \pi/2]$ ainsi que la/les solutions de l'EDO issue de ces points. Puisque $\Omega \neq \mathbb{R}^d \setminus \{0\}$, le modèle fluide n'est pas déterministe: pour tout point initial sur la diagonale, le support de la limite fluide est composé de deux trajectoires qui sont toutes les deux solution de l'EDO

Pour établir la stabilité du modèle fluide, nous vérifions la condition 2 du Théorème 11, Section 2.1.3; la vérification de cette condition est très technique et assez spécifique au problème considéré ([16, Proposition 2.10]). Cette difficulté est un point faible de l'approche par limites fluides pour étudier l'ergodicité des échantillonneurs.

Nous déduisons ensuite du Théorème 10 la V -ergodicité de cet échantillonneur à n'importe quelle vitesse polynomiale, pour toute fonction $V(x) = 1 + |x|^u$, $u \geq 0$.

Limites fluides pour des densités sous-exponentielles

Nous avons montré ([16, Proposition 2.13]) que si

- A. la densité cible π est sous-exponentielle dans les queues; la loi de proposition q vérifie (3.2) et possède des moments d'ordre $p \geq 2$.

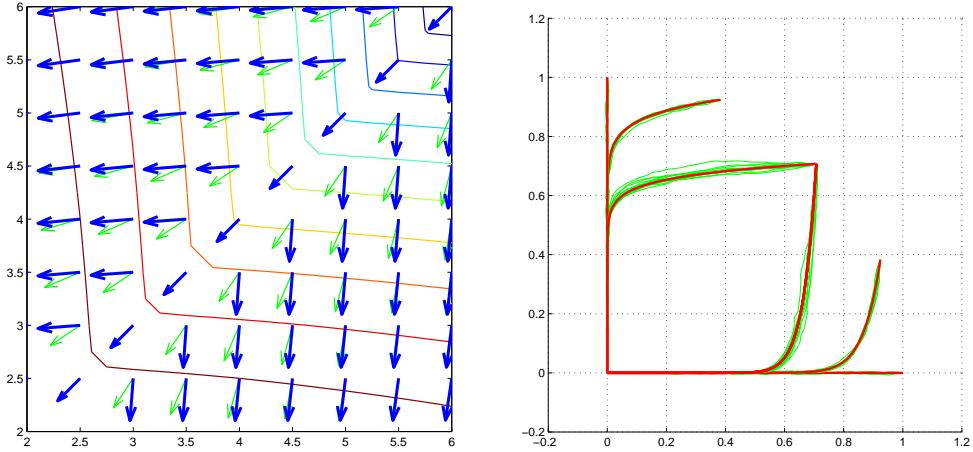


Fig. 3.5: Cas de la densité (3.5) [Gauche]: Fonctions Δ en vert et h en bleu. [Droite]: En vert, trajectoires du processus (2.6) issues d'un ensemble de points sur la sphère unité de $[0, \pi/2]$; en rouge, les trajectoires de l'EDO.

B. il existe un cône ouvert \mathcal{O} de $\mathbb{R}^d \setminus \{0\}$, une fonction continue $\ell_\infty : \mathcal{O} \rightarrow \mathbb{R}^d$ et $\beta \in]0, 1[$ tels que pour tout compact H du cône \mathcal{O} et tout $K > 0$,

$$\lim_{r \rightarrow \infty} \sup_{x \in H} \int_{R_{rx} \cap \{y, |y| \leq K\}} \left| r^\beta |x|^\beta \left\{ \frac{\pi(rx + y)}{\pi(rx)} - 1 \right\} - \langle \ell_\infty(x), y \rangle \right| q(y) \lambda(dy) = 0 ,$$

$$\lim_{r \rightarrow \infty} \sup_{x \in H} Q(R_{rx} \ominus \{y, \langle \ell_\infty(x), y \rangle \geq 0\}) = 0 .$$

alors (3.1) est vérifiée avec

$$h(x) = |x|^{-\beta} \left(\int_{\mathbb{R}^d} y_1^2 \mathbb{I}_{\{y_1 \geq 0\}} q_0(y) \lambda(dy) \right) \Sigma \ell_\infty(x) .$$

► **Exemple 1: cas régulier** Considérons la densité cible dans \mathbb{R}^2

$$\pi(x_1, x_2) \propto (1 + x_1^2 + x_2^2 + x_1^8 x_2^2)^\delta \exp(-(x_1^2 + x_2^2)^\delta) . \quad (3.6)$$

et une loi de proposition gaussienne $q \sim \mathcal{N}_2(0, \sigma^2 \text{Id}_2)$. Les courbes de niveau de π sont représentées Figure 3.6[gauche]. Cette densité vérifie les conditions A et B avec $\mathcal{O} = \mathbb{R}^d \setminus \{0\}$, $\beta = 1 - 2\delta$ et $\ell_\infty = -2\delta x/|x|$. Par suite, $h(x) = -\sigma^2 \delta x/|x|$. La fonction Δ et la limite h sont représentées sur la Figure 3.7[gauche].

La solution de l'EDO $\dot{\mu} = h(\mu)$ de condition initiale x est donnée par $\mu(t; x) = [|x|^{2(1-\delta)} - 2\sigma^2 \delta (1-\delta)t]^{0.5(1-\delta)^{-1}} (x/|x|) \mathbb{I}_{|x|^{2(1-\delta)} - 2\sigma^2 \delta (1-\delta)t \geq 0}$. Cette solution est unique et est stable. Sur la Figure 3.7[droite], nous traçons quelques trajectoires de processus interpolé $t \mapsto \eta_r^\beta(t; x)$ (cf. Equation 2.6), issues de points x sur la sphère unité de $[0, \pi/2]$ ainsi que la solution de l'EDO issue de ces points.

Par application des Théorèmes 10 et 11 (cf. Section 2.1.3), cet échantillonneur est V_u -ergodique à la vitesse polynomiale $(n+1)^{u-1}$, avec $V_u(x) = 1 + |x|^{p-u(2-m\delta)}$ pour tout $u \in [1, p/(2-m\delta)]$. Pour cet exemple, les conditions suffisantes d'ergodicité proposées dans [19, 7] - basées sur des conditions de drift - ne s'appliquent pas.

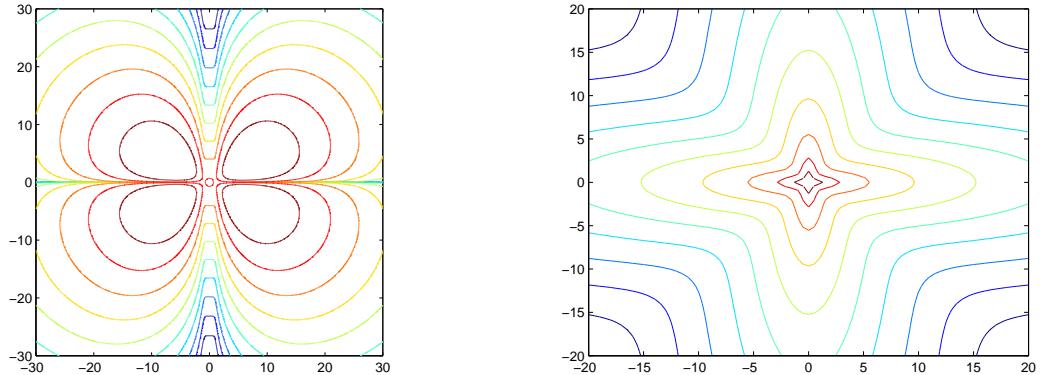


Fig. 3.6: [Gauche] Courbes de niveau de la densité (3.6) avec $\delta = 0.4$ et [Droite] de la densité (3.7) avec $\delta = 0.4$

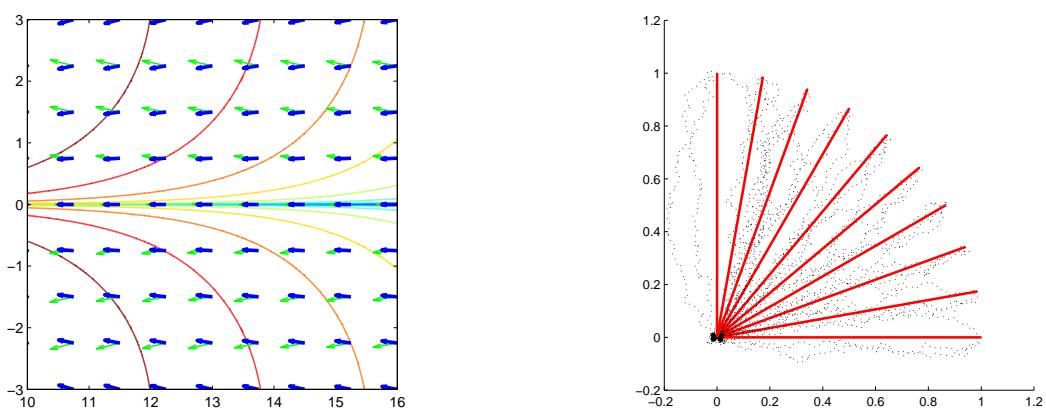


Fig. 3.7: Cas de la densité (3.6) avec $\delta = 0.4$ [Gauche]: Fonctions Δ en vert et h en bleu. [Droite]: En pointillés, trajectoires du processus (2.6) issues d'un ensemble de point sur la sphère unité de $[0, \pi/2]$; en trait rouge plein, les trajectoires de l'EDO.

Cette densité est un exemple de densités de la famille exponentielle de la forme $g(x) \exp(-p^\delta(x))$, deux fois continûment différentiables, où (a) g est une fonction strictement positive, à variation lente à l'infini; (b) p est une fonction polynomiale d'ordre m strictement positive et $\lim_{|x| \rightarrow \infty} p_m(x) = +\infty$, p_m désignant le polynôme constitué des termes d'ordre m ; (c) $\delta \in]0, 1[$.

Nous montrons que la conclusion obtenue pour la densité (3.6) reste vraie pour toute densité de cette famille : $\mathcal{O} = \mathbb{R}^d \setminus \{0\}$, $\beta = 1 - m\delta$, $\ell_\infty(x) \propto -\delta p_m^{\delta-1}(x/|x|) \nabla p_m(x/|x|)$ et la solution de l'EDO $\dot{\mu} = h(\mu)$ de condition initiale x , pour $x \in \mathcal{O}$, est unique sur $[0, T_x]$ et vérifie $\lim_{t \rightarrow T_x^-} \mu(t; x) = 0$ ([16, Proposition 2.14]).

► **Exemple 2: cas irrégulier** Considérons la densité cible sur \mathbb{R}^2 donnée par

$$\pi(x) \propto \alpha(x' \Gamma_1^{-1} x)^{\delta-1} \exp(-(1/2)(x' \Gamma_1^{-1} x)^\delta) + (1-\alpha)(x' \Gamma_2^{-1} x)^{\delta-1} \exp(-(1/2)(x' \Gamma_2^{-1} x)^\delta), \quad (3.7)$$

où $\alpha \in]0, 1[$, $\Gamma_1^{-1} := \text{diag}(a^2, 1)$ et $\Gamma_2^{-1} := \text{diag}(1, a^2)$ pour $a^2 > 1$; et une loi de proposition gaussienne $q \sim \mathcal{N}_2(0, \text{Id}_2)$. Les courbes de niveau sont représentées Figure 3.6[droite]: elles ont des points singuliers le long de la diagonale qui ne disparaissent pas à la limite (quand $r \rightarrow +\infty$) - la courbure des lignes de niveau $\{C_{rx}, r \geq 0\}$ pour x sur la diagonale tend vers $+\infty$.

Cette densité vérifie les conditions A à C sur le cône ouvert $\mathcal{O} = \mathbb{R}^d \setminus \{x = (x_1, x_2) \in \mathbb{R}^2, |x_1| = |x_2|\}$ et on a

$$h(x) = \begin{cases} -m_2 \delta (x' \Gamma_2^{-1} x)^{\delta-1} \Sigma \Gamma_2^{-1} x & \text{si } |x_1| > |x_2| \\ -m_2 \delta (x' \Gamma_1^{-1} x)^{\delta-1} \Sigma \Gamma_1^{-1} x & \text{si } |x_1| < |x_2| \end{cases} \quad m_2 := \int_{\mathbb{R}^d} y_1^2 \mathbf{1}_{y_1 \geq 0} q_0(y) \lambda(dy) > 0.$$

La fonction Δ et la limite h sont représentées sur la figure 3.8[gauche]; du fait de la symétrie du problème, nous ne représentons ces fonctions que sur l'orthant supérieur droit. Sur la Figure 3.8[droite], nous traçons quelques trajectoires de processus interpolé $t \mapsto \eta_r^\beta(t; x)$ (cf. Equation 2.6), issue d'un point x sur la diagonale, ainsi que la/les solutions de l'EDO issue de ce point. Puisque $\mathcal{O} \neq \mathbb{R}^d \setminus \{0\}$, le modèle fluide n'est pas déterministe: pour tout point initial sur la diagonale, le support de la limite fluide est composé de deux trajectoires qui sont toutes les deux solutions de l'EDO.

L'étude de la V -ergodicité de cet échantillonneur est menée de façon analogue à celle de l'échantillonneur de densité cible (3.5).

3.1.2 Limites fluides d'un algorithme Metropolis-within-Gibbs

Nous avons étudié les limites fluides des algorithmes Metropolis-within-Gibbs (décrit section 1.2.2) afin d'étudier l'influence des poids de sélection $\{\omega_i, i \leq d\}$ sur la phase transiente de ces échantillonneurs. Nous avons considéré le cas d'une densité cible π sur $\mathcal{X} \subseteq \mathbb{R}^d$; les étapes de Hastings-Metropolis sont des algorithmes à marche aléatoire symétrique de loi de proposition directionnelle $\mathcal{N}(0, \sigma_i^2)$.

Les Figures 3.9, 3.10 et 3.11 montrent l'allure de ces limites fluides

(i) pour différentes densités cibles sur \mathbb{R}^2 : $\pi^{(1)} = \mathcal{N}_2(0, \Gamma_1)$, $\pi^{(2)} = \mathcal{N}_2(0, \Gamma_2)$ et $\pi^{(3)} = 0.5(\pi^{(1)} + \pi^{(2)})$ avec

$$\Gamma_1 = \begin{bmatrix} 0.5 & 0 \\ 0 & 4.5 \end{bmatrix} \quad \Gamma_2 = \begin{bmatrix} 1 & 1.3 \\ 1.3 & 4 \end{bmatrix}$$

(ii) pour différentes valeurs des poids de sélection: $\omega^{(1)} = (0.25; 0.75)$, $\omega^{(2)} = (0.5; 0.5)$.

Ces exemples illustrent le cadre général suivant

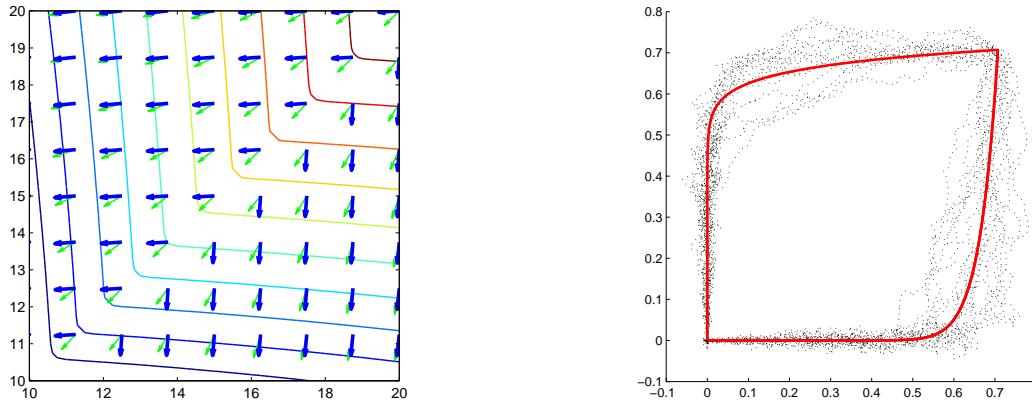


Fig. 3.8: Cas de la densité (3.7) avec $\delta = 0.4$. [Gauche]: Fonctions Δ en vert et h en bleu. [Droite]: En pointillés, trajectoires du processus (2.6) issues du point de coordonnées $(\sqrt{2}/2, \sqrt{2}/2)$; en trait plein, les trajectoires de l'EDO.

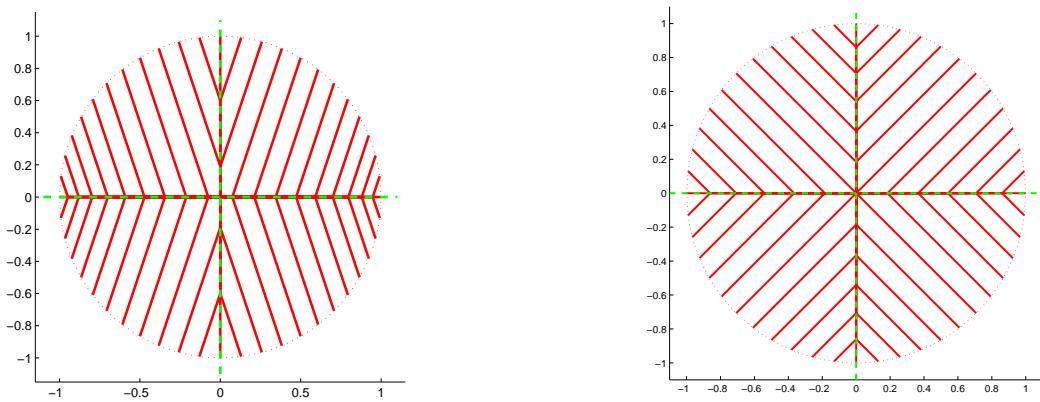


Fig. 3.9: En rouge, limites fluides issues de la sphère unité, de l'algorithme Metropolis-within-Gibbs dans le cas [gauche] $\pi^{(1)}$ et $\omega^{(1)}$; et [droite] $\pi^{(1)}$ et $\omega^{(2)}$

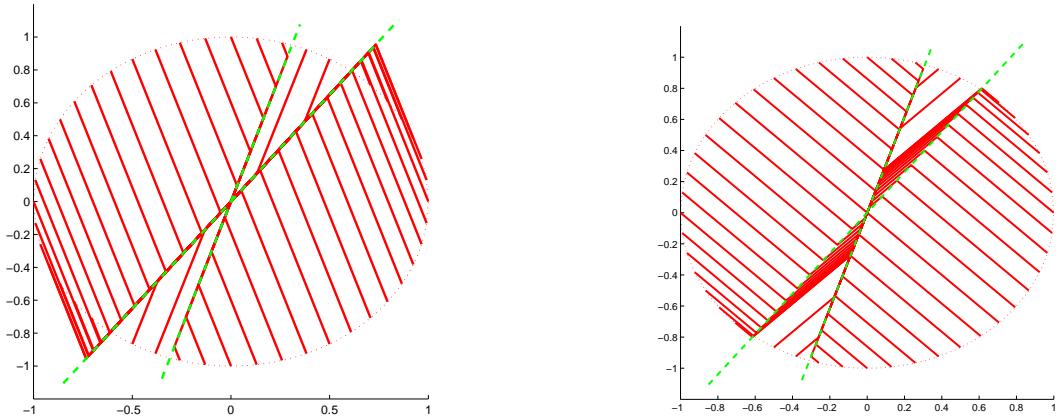


Fig. 3.10: En rouge, limites fluides issues de la sphère unité, de l'algorithme Metropolis-within-Gibbs dans le cas [gauche] $\pi^{(2)}$ et $\omega^{(1)}$; et [droite] $\pi^{(2)}$ et $\omega^{(2)}$

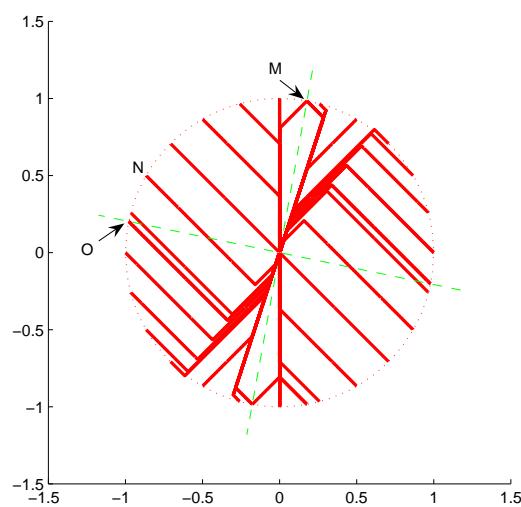


Fig. 3.11: En rouge, limites fluides issues de la sphère unité de l'algorithme Metropolis-within-Gibbs dans le cas $\pi^{(3)}$ et $\omega^{(1)}$

la chaîne de Markov est à valeur dans \mathbb{Y} , sous-ensemble fermé et non borné de \mathbb{R}^d , tel que $\mathbb{X} := \{x \in \mathbb{R}^d, \exists(r, y) \in \mathbb{R}_+ \times \mathbb{Y}, x = ry\}$ est fermé.

il existe une partition de la sphère unité sur \mathbb{X} de la forme $\mathcal{F} \cup \bigcup_{i=1}^a O_i$ où F est un fermé et O_i est un ouvert, tels que $x \mapsto h(x) := \lim_{r \rightarrow +\infty} \Delta(rx)$ est constante sur tous les ouverts O_i .

Nous avons mené une étude des limites fluides dans ce cadre général et montré que la limite fluide issue de la sphère unité sur O_i était une masse de Dirac en une fonction continue g_i , qui est linéaire jusqu'au premier instant où elle rencontre le fermé $\mathbf{r}\mathcal{F} := \{x \in \mathbb{X}_*, |x|^{-1}x \in \mathcal{F}\}$; puis, selon le comportement de h dans un voisinage de $\mathcal{F} \cap \overline{O_i}$, les limites fluides restent dans \mathcal{F} ou pas.

La stabilité de ces limites fluides est assez difficile à caractériser en toute généralité dans \mathbb{R}^d . Nous avons donc considéré le cas $d = 2$ et donné des conditions suffisantes de stabilité en termes de l'orientation des champs h au voisinage de $\mathbf{r}\mathcal{F}$.

Sur les Figures 3.9 et 3.10, l'ensemble $\mathbf{r}\mathcal{F}$ est représenté en vert. Les situations représentées sur la figure 3.9 ou 3.10[gauche] sont des exemples où $\mathbf{r}\mathcal{F}$ est absorbant : la limite fluide, quand elle l'atteint, reste dans $\mathbf{r}\mathcal{F}$. En revanche, la situation représentée figure 3.10[droite] est un cas où $\mathbf{r}\mathcal{F}$ est l'union de deux droites, dont l'une est absorbante et l'autre répulsive: la limite, quand elle l'atteint, traverse cette composante pour être absorbée par la seconde. La situation représentée figure 3.11 est un exemple où le modèle fluide n'est pas déterministe: la limite fluide issue du point O ou du point N est une masse de Dirac en une fonction continue, mais celle issue de M ne l'est pas.

Nous avons établi que pour certaines densités cibles dont les densités $\pi^{(i)}$ sont des exemples, la limite fluide - quand elle est déterministe - est solution de l'EDO $\dot{\mu} = h(\mu)$ où le champ $h = (h_1, \dots, h_d)$ est donné par

$$h_i(x) = \frac{\omega_i \sigma_i}{\sqrt{2\pi}} \lim_{r \rightarrow +\infty} \frac{\nabla_i \ln \pi(rx)}{|\nabla \ln \pi(rx)|}.$$

La chaîne suit donc une dynamique d'algorithme de gradient. On peut alors appliquer les résultats présentés en section 2.1.3 pour établir l'ergodicité de cet échantillonneur.

Cette étude, menée en partie pour des chaînes à valeur dans \mathbb{R}^2 , montre une des faiblesses de l'approche par limites fluides pour caractériser l'ergodicité: la complexité de la caractérisation du modèle fluide et par suite, la difficulté d'établir sa stabilité. Pour se convaincre de cette difficulté, il suffit d'imaginer les différentes formes de la partition en ouverts/fermés de la sphère unité de \mathbb{R}^d dans le cas $d > 2$; et il suffit d'observer dans le cas $d = 2$ pour des densités cibles très simples ($\pi \sim \mathcal{N}_2$) les différents comportements des limites fluides selon le point initial x , la valeur des paramètres qui définissent la loi cible π et selon la valeur des paramètres d'implémentation (variances des lois de proposition, poids de sélections).

En revanche, cette approche donne une vision très précise sur le rôle des différents paramètres d'implémentation de l'échantillonneur sur sa dynamique en phase transiente. Ces travaux sont détaillés dans [13].

3.1.3 Conclusion

Nous avons montré que les limites fluides permettaient d'établir l'ergodicité de certains échantillonneurs, là où d'autres approches basées sur les conditions de drift échouent (cf. l'exemple du SRWM lorsque la

loi cible π est donnée par (3.4) ou (3.6)). Néanmoins, dans certains cas, cette approche peut s'avérer très fastidieuse; c'est par exemple le cas lorsque l'accroissement moyen Δ possède des limites radiales uniformes dans un cône ouvert strictement inclu dans $\mathbb{X} \setminus \{0\}$. (cf. l'exemple du SRWM lorsque la loi cible π est donnée par (3.5) ou (3.7); dans ce cas, vérifier les conditions du Théorème 11 est très technique).

Les limites fluides portent une information très précise sur l'échantillonneur: l'expression du champ h dépend de la loi cible π et des paramètres d'implémentation. La limite fluide traduisant la dynamique de la chaîne de Markov, on peut vouloir choisir les paramètres d'implémentation des échantillonneurs de façon à optimiser le comportement du modèle fluide. Une idée naturelle - conséquence de la relation entre la stabilité de la limite fluide et la récurrence de la chaîne (cf. section 2.1.3) - est de choisir les paramètres d'implémentation de façon à minimiser le temps d'entrée dans une boule de rayon ρ ($\rho < 1$) des EDO issues de la sphère unité. Cette idée appliquée par exemple au SRWM pour la densité cible (3.4) montre qu'il faut choisir la variance σ^2 de la loi de proposition arbitrairement grande ! ce résultat n'est pas conforme à l'expérience que l'on a des algorithmes SRWM, dont l'efficacité se détériore lorsque σ^2 est trop grand (cf. figure 1.1).

Le fait est qu'optimiser le comportement du "squelette" de l'algorithme ne garantit pas que l'on optimise le comportement de l'algorithme lui-même, car on ne maîtrise pas les fluctuations de l'algorithme autour de son modèle fluide. L'étude des limites fluides que nous avons menée doit donc, pour être appliquée à l'optimisation des échantillonneurs à dynamique markovienne, être complétée par une étude de ces fluctuations: une étude de "limites diffusives".

L'approche limite fluide comme l'approche présentée en début de section 3.1.1 repose sur la construction d'un processus à temps continu obtenu par une renormalisation de la chaîne en temps et en espace. Ces deux constructions n'utilisent pas la même asymptotique; il est donc difficile de les comparer.

3.2 Convergence d'algorithmes adaptatifs à dynamique markovienne

Les algorithmes de simulation MCMC sont des exemples d'algorithmes à dynamique markovienne: il s'agit d'obtenir la réalisation d'une chaîne de Markov $\{X_n, n \geq 0\}$ de noyau de transition P construit pour que la mesure cible π soit l'unique mesure invariante (à une constante multiplicative près). Les procédures adaptatives *Adaptive Metropolis* et *Equi-Energy* présentées Chapitre 1 sont des exemples du cadre général suivant:

(AA) obtenir la réalisation d'un processus $\{X_n, n \geq 0\}$, adapté par rapport à une filtration $\{\mathcal{F}_n, n \geq 0\}$ tel que la loi conditionnelle vérifie

$$\mathbb{P}(X_{n+1} \in \cdot | \mathcal{F}_n) = P_{\theta_n}(X_n, \cdot)$$

où $\{\theta_n, n \geq 0\}$ est une famille de v.a. \mathcal{F}_n -adaptée et $\{P_\theta, \theta \in \Theta\}$ est une famille de noyaux de transition sur \mathbb{X} .

Dans l'algorithme *Adaptive Metropolis* décrit section 1.2.1, on peut définir la filtration \mathcal{F}_n comme la filtration naturelle du processus $\{X_n, n \geq 0\}$; le processus $\{\theta_n, n \geq 0\}$ est défini à l'aide du passé du processus d'intérêt $\{X_n, n \geq 0\}$: l'adaptation est **interne**. C'est de plus un cas où tous les noyaux $\{P_\theta, \theta \in \Theta\}$ ont même probabilité invariante π . (cf. [27, 105] pour d'autres exemples d'algorithmes à adaptation interne).

Dans l'algorithme *Equi-Energy* décrit section 1.2.4, on peut définir la filtration \mathcal{F}_n comme la filtration naturelle des processus $\{X_n, Y_n, n \geq 0\}$; le processus $\{\theta_n, n \geq 0\}$ est défini à l'aide du passé de processus d'intérêt $\{X_n, n \geq 0\}$ et d'un processus auxiliaire $\{Y_n, n \geq 0\}$: l'adaptation est **externe**. Dans cet exemple en revanche, les noyaux $\{P_\theta, \theta \in \Theta\}$ n'ont pas nécessairement la même probabilité invariante; néanmoins, ils sont tels qu'il existe θ_* tel que $\pi P_{\theta_*} = \pi$ et le processus $\{\theta_n, n \geq 0\}$ est construit de façon à converger (au sens d'une limite à préciser) vers θ_* (cf. [25, 43, 44] pour d'autres exemples d'algorithmes à adaptation externe).

La convergence des algorithmes adaptatifs décrits par **(AA)** n'est pas une conséquence de l'ergodicité de chaque noyau: même si pour tout $\theta \in \Theta$, la chaîne de noyau P_θ est ergodique et converge vers π , il n'est pas vrai que la loi du processus $\{X_n, n \geq 0\}$ décrit par **(AA)** converge vers π . Pour s'en convaincre, considérons l'exemple de noyaux de transition à valeur dans $\{0, 1\}$ donnés par

$$P_\theta = \begin{bmatrix} 1 - \theta & \theta \\ \theta & 1 - \theta \end{bmatrix}$$

pour $\theta \in]0, 1[$, dont la mesure invariante est $\pi \propto [1, 1]$; pour $t_0 \neq t_1 \in]0, 1[$, on définit l'adaptation par $\theta_k = \Xi(X_k)$ où $\Xi(i) = t_i$. Alors le processus $\{X_n, n \geq 0\}$ est une chaîne de Markov de matrice de transition

$$\begin{bmatrix} 1 - t_0 & t_0 \\ t_1 & 1 - t_1 \end{bmatrix}$$

et la mesure invariante de cette chaîne est proportionnelle à $[t_1, t_0]$. L'adaptation peut donc détruire la convergence vers la loi d'intérêt π !

Cela justifie l'étude de la convergence des échantillonneurs adaptatifs: quelles conditions sur les noyaux de transition P_θ et sur la procédure d'adaptation, garantissent la convergence de la loi du processus $\{X_n, n \geq 0\}$ décrit par **(AA)** ?

Nous avons étudié la stabilité des algorithmes adaptatifs de type **(AA)**, stabilité au sens de

(i) leur ergodicité, i.e. la convergence de la loi de X_n vers la loi cible π .

(ii) l'existence d'une loi forte des grands nombres i.e. la convergence presque-sûre de $n^{-1} \sum_{k=1}^n f(X_k)$ vers $\pi(f)$ pour des fonctions f qui ne sont pas nécessairement bornées.

En particulier, pour caractériser l'ensemble des fonctions f pour lesquelles une loi des grands nombres existe, nous avons dû supposer que chaque noyau P_θ était ergodique à une certaine vitesse; nous avons considéré le cas géométrique [3], et sous-géométrique [2].

3.2.1 Ergodicité de l'échantillonneur adaptatif

Pour $\theta, \theta' \in \Theta$, on définit

$$D_{\text{TV}}(\theta, \theta') := \sup_{x \in \mathbb{X}} \|P_\theta(x, \cdot) - P_{\theta'}(x, \cdot)\|_{\text{TV}} , \quad (3.8)$$

et pour $x \in \mathsf{X}, \theta \in \Theta, \epsilon > 0$, $M_\epsilon(x, \theta) := \inf\{n \geq 0, \|P_\theta^n(x, \cdot) - \pi_\theta\|_{\text{TV}} \leq \epsilon\}$. On note $\mathbf{b}\mathcal{X}$ l'espace des fonctions mesurables de X dans \mathbb{R} bornées par 1.

L'ergodicité des processus $\{X_n, n \geq 0\}$ satisfaisant **(AA)** est établie sous les hypothèses

E1 Pour tout $\theta \in \Theta$, P_θ est un noyau de transition sur X qui possède une unique probabilité invariante π_θ .

E2 (*Diminishing adaptation*) La suite $\{D_{\text{TV}}(\theta_n, \theta_{n-1}), n \geq 1\}$ converge en probabilité vers 0.

E3 (*Containment condition*) Pour tout $\epsilon > 0$, la suite $\{M_\epsilon(X_n, \theta_n), n \geq 0\}$ est bornée en probabilité.

E4 il existe un sous-ensemble F de $\mathbf{b}\mathcal{X}$, et une probabilité π sur X telle que pour tout $\epsilon > 0$, $\lim_{n \rightarrow \infty} \sup_{f \in \mathsf{F}} \mathbb{P}(|\pi_{\theta_n}(f) - \pi(f)| \geq \epsilon) = 0$.

La condition E2, introduite par [100], est une condition sur la façon d'adapter: le changement doit être de plus en plus faible, au sens de l'effet induit sur les noyaux de transition. La condition E3, qui généralise la *containment condition* de [100], est une condition sur l'ergodicité de chaque noyau: celle-ci doit être uniforme en x, θ au sens où le temps nécessaire pour que $P_\theta^n(x, \cdot)$ soit dans un ϵ -voisinage de π_θ est contrôlable en (x, θ) .

Théorème 21. [3, Théorème 1.3.3] Supposons E1 à E4. Alors $\lim_n \sup_{f \in \mathsf{F}} |\mathbb{E}[f(X_n)] - \pi(f)| = 0$.

Sous la condition E2, on montre que pour tout N , la loi conditionnelle $\mathbb{P}(X_{n+N} \in \cdot | \mathcal{F}_n)$ - la filtration \mathcal{F}_n contenant la filtration naturelle des processus θ_k, X_k - est proche de $P_{\theta_n}^N(X_n, \cdot)$; ce noyau itéré est lui-même proche de π_{θ_n} puis de π sous les conditions E3 et E4.

Nous donnons ci-après des conditions suffisantes pour vérifier E3. Elles reposent sur des conditions de V -ergodicité géométrique, uniforme en θ , des noyaux de transition $\{P_\theta, \theta \in \Theta\}$. On peut affaiblir cette condition de vitesse géométrique, en une condition de vitesse sous-géométrique: cf.[2, Proposition 2.4].

E5 (a) Pour tout $\theta \in \Theta$, P_θ est psi-irréductible.

(b) Il existe une fonction mesurable $V : \mathsf{X} \rightarrow [1, +\infty)$, des constantes $b \in (0, +\infty)$, $\lambda \in (0, 1)$ et un ensemble $\mathcal{C} \in \mathcal{X}$ tels que pour tout $\theta \in \Theta$, $P_\theta V \leq \lambda V + b \mathbb{1}_{\mathcal{C}}$.

(c) Les ensembles de niveau de V sont 1-small uniformément en θ i.e. pour tout $v \geq 1$, il existe une constante $\varepsilon > 0$ et une mesure de probabilité ν telle que pour tout $\theta \in \Theta$, $P_\theta(x, \cdot) \geq \varepsilon \mathbb{1}_{\{V \leq v\}}(x) \nu$.

Ces conditions entraînent que P_θ est apériodique; par application des résultats sur le contrôle explicite d'ergodicité présentés en Section 2.2, on déduit de E5 l'existence de constantes C et $\beta > 1$ telles que $\sup_\theta \|P_\theta^n(x, \cdot) - \pi_\theta\|_V \leq C \beta^n V(x)$.

La condition E4 est assez difficile à vérifier car, très souvent, on ne dispose pas de l'expression explicite de la mesure invariante π_θ (c'est le cas dans l'étude de la stabilité de l'algorithme *Equi-Energy* présentée en section 3.2.4). On propose une condition suffisante exprimée en terme de convergence “presque-sûre” des noyaux de transition.

E6 Il existe $\theta_* \in \Theta$ et un ensemble $A \in \mathcal{A}$ tel que $\mathbb{P}(A) = 1$ et pour tout $\omega \in A$, $x \in \mathsf{X}$, et $B \in \mathcal{X}$,

$$\lim_n P_{\theta_n(\omega)}(x, B) = P_{\theta_*}(x, B).$$

On établit alors

Théorème 22. [3, Théorème 1.3.7] Supposons E2, E5 et E6. Alors pour toute fonction bornée f , $\lim_{n \rightarrow \infty} |\mathbb{E}[f(X_n)] - \pi_{\theta_*}(f)| = 0$.

3.2.2 Loi forte des grands nombres

En utilisant les solutions de l'équation de Poisson, nous décomposons la quantité d'intérêt en une martingale et des termes "de reste". La loi des grands nombres est alors une conséquence d'une loi des grands nombres pour des martingales dans $L^p, p > 1$.

Supposons que pour tout θ , il existe une fonction \hat{f}_θ sur X solution de l'équation de Poisson $f - \pi_\theta(f) = \hat{f}_\theta - P_\theta \hat{f}_\theta$. On écrit

$$\frac{1}{n} \sum_{k=1}^n \{f(X_k) - \pi(f)\} = \frac{1}{n} \sum_{k=1}^n \{f(X_k) - \pi_{\theta_{k-1}}(f)\} + \frac{1}{n} \sum_{k=1}^n \{\pi_{\theta_{k-1}}(f) - \pi(f)\}$$

puis

$$\frac{1}{n} \sum_{k=1}^n \{f(X_k) - \pi_{\theta_{k-1}}(f)\} = n^{-1} M_n(f) + \sum_{i=1}^2 R_{n,i}(f)$$

avec

$$\begin{aligned} M_n(f) &:= \sum_{k=1}^n \{\hat{f}_{\theta_{k-1}}(X_k) - P_{\theta_{k-1}} \hat{f}_{\theta_{k-1}}(X_{k-1})\}, \\ R_{n,1}(f) &:= n^{-1} \left(P_{\theta_0} \hat{f}_{\theta_0}(X_0) - P_{\theta_{n-1}} \hat{f}_{\theta_{n-1}}(X_n) \right), \\ R_{n,2}(f) &:= n^{-1} \sum_{k=1}^{n-1} \{P_{\theta_k} \hat{f}_{\theta_k}(X_k) - P_{\theta_{k-1}} \hat{f}_{\theta_{k-1}}(X_k)\}. \end{aligned}$$

Pour tout $\alpha \in]0, 1[$ et toute fonction f telle que $|f|_{V^\alpha} < +\infty$, la convergence presque-sûre de $\{\pi_{\theta_k}(f), k \geq 0\}$ vers $\pi_*(f)$ est garantie sous les conditions E5 et E6. La loi des grands nombres pour le terme martingale $M_n(f)$ est entraînée par des conditions de moments sur les incrément de martingales; ces conditions sont vérifiées sous E5. La convergence p.s. de $R_{n,1}(f)$ vers 0 est établie par une condition de moments, conséquence de E5. Enfin, le terme $R_{n,2}(f)$ est spécifique à l'adaptation: il est nul si $\theta_k = \theta_{k+1}$; on établit sa convergence p.s. en étudiant la régularité en θ des solutions de Poisson $\theta \mapsto \hat{f}_\theta$ et des mesures invariantes $\theta \mapsto \pi_\theta$. Le contrôle de ce dernier terme $R_{n,2}(f)$ nécessite de renforcer la condition d'adaption décroissante E2: on définit pour tout $\theta, \theta' \in \Theta$,

$$D_{V^\alpha}(\theta, \theta') := \sup_{x \in \mathsf{X}} \frac{\|P_\theta(x, \cdot) - P_{\theta'}(x, \cdot)\|_{V^\alpha}}{V^\alpha(x)};$$

et on suppose

E7 Pour tout $\alpha \in (0, 1)$, $\sum_{k \geq 1} k^{-1} D_{V^\alpha}(\theta_k, \theta_{k-1}) V^\alpha(X_k) < +\infty$ presque-sûrement.

On établit alors

Théorème 23. [3, Théorèmes 1.4.3. et 1.4.4] Supposons E5, E6, E7 et $\mathbb{E}[V(X_0)] < +\infty$. Alors pour tout $\alpha \in]0, 1[$ et toute fonction f telle que $|f|_{V^\alpha} < +\infty$, $n^{-1} \sum_{k=1}^n f(X_k) - \pi_{\theta_*}(f) \rightarrow 0$ presque-sûrement.

Les premiers résultats de Loi forte des Grands Nombres pour des algorithmes adaptatifs à dynamique markovienne, sont dûs à haario et al. : ils établissent un résultat de consistance pour l'algorithme *Adaptive Metropolis*, et pour les fonctions bornées [65]. Atchade & Rosenthal ont ensuite étendu ce résultat au cas des fonctions non bornées [31]. Andrieu & Moulines ont prouvé une Loi forte des Grands Nombres (et un Théorème Central Limite) pour des algorithmes de type **(AA)** à adaptation interne, pour lesquels l'adaptation est guidée par une procédure d'approximation stochastique [26]. Enfin, Roberts & Rosenthal ont établi une Loi faible des Grands Nombres pour des fonctions bornées, dans le cadre des algorithmes de type **(AA)** [100].

Ergodicité des noyaux uniforme en θ

La condition E3 est ici vérifiée par une condition d'ergodicité uniforme en θ ; on trouve les mêmes conditions dans [100]. Cette hypothèse peut s'avérer invérifiable dans certaines applications : par exemple, dans l'exemple de l'algorithme *Adaptive Metropolis* en dimension $d = 1$ où θ est une variance, donc $\Theta = \mathbb{R}_*^+$, cette condition n'est pas vérifiable uniformément sur Θ . Il est en revanche plus réaliste de supposer que la condition est vraie uniformément sur Θ_n où $\{\Theta_n, n \geq 0\}$ est une famille de sous-ensembles de Θ vérifiant quelques "bonnes" propriétés pour que ces uniformités locales suffisent pour établir E3. Dans l'exemple de l'algorithme *Adaptive Metropolis*, $\{\Theta_n, n \geq 0\}$ est une famille croissante de compacts de \mathbb{R}_*^+ .

Nous n'avons pas écrit en détail cette extension : remplacer l'hypothèse d'uniformité globale par des hypothèses d'uniformité locale. Néanmoins, une façon de prouver cette extension, est par exemple d'utiliser un argument analogue à ce qui est fait dans [18, 26], basé sur une méthode de troncations aléatoires proposée par [46, 47]. Cette approche est indiquée dans les situations où la construction de $\{\theta_n, n \geq 0\}$ n'est pas homogène dans le temps, ce qui est par exemple le cas dans les situations où l'adaptation est basée sur une procédure d'approximation stochastique.

Cas des noyaux sous-géométriquement ergodiques

La condition E3 est aussi impliquée par une condition d'ergodicité sous-géométrique (uniforme en θ) des noyaux de transition. Nous avons établi une Loi forte des Grands Nombres, dans le cas où la condition de drift dans E5 est remplacée par une condition de la forme $P_\theta V \leq V - cV^{1-\alpha_*} + b\mathbb{I}_{\mathcal{C}}$ (cf. Section 2.1.1 pour l'étude de la stabilité des noyaux vérifiant cette inégalité).

Cette étude a été menée dans le cas où les noyaux ont tous même probabilité invariante ($\pi_\theta = \pi$); et le processus $\{(X_n, \theta_n), n \geq 0\}$ est une chaîne de Markov non-homogène.

L'approche retenue est encore basée sur les solutions d'une équation de type Poisson et sur la Loi des Grands Nombres pour les martingales. Néanmoins, dans le cas d'un drift sous-géométrique, l'étude de la dépendance en x et de la régularité en θ des solutions $\hat{f}_\theta(x)$ de l'équation de Poisson $f - \pi_\theta(f) = \hat{f}_\theta - P_\theta \hat{f}_\theta$ est plus difficile que dans le cas géométrique. Nous avons donc préféré utiliser une approche type "noyau résolvant" - adaptée des travaux de [81, 79].

Sous des conditions qui sont essentiellement

- (a) des conditions de minoration du noyau et de drift polynomial, uniformes en θ ;
- (b) une condition d'adaption décroissante;

nous avons établi que pour tout $\beta \in [0, 1 - \alpha_*[$, et tout fonction f telle que $|f|_{V^\beta} < +\infty$, on a : $\lim_n n^{-1} \sum_{k=1}^n f(X_k) - \pi(f) = 0$ presque-sûrement.

Ces travaux sont détaillés dans [2].

3.2.3 Application à la convergence d'un algorithme de Hastings-Metropolis adaptatif

Nous avons étudié la stabilité d'un algorithme proche de l'algorithme *Adaptive Metropolis* dans le cas où la loi cible π , sur \mathbb{R}^d , est à décroissance sous-exponentielle. Le cas de décroissance sur-exponentielle a été étudié par [26].

Il s'agit de mettre en oeuvre un algorithme de Hastings-Metropolis à marche aléatoire symétrique dans \mathbb{R}^d , dont le noyau de proposition $q(x, dy)$ est une loi gaussienne centrée en x dont la matrice de variance-covariance Σ est modifiée à chaque itération. Des études théoriques suggèrent de fixer Σ à $c_* \Sigma_*$, où Σ_* est la variance de la loi cible π et c_* est une réel positif choisi pour que, à stationnarité, le ratio d'acceptation-rejet soit optimal [99]. Typiquement, ce ratio d'acceptation-rejet optimal est de l'ordre de 0.3 – 0.4.

La version adaptative de cet algorithme a pour objectif de déterminer une estimation de Σ_* et du paramètre d'échelle c_* . Plus précisément, nous avons considéré l'algorithme suivant, qui est un exemple d'algorithme à adaptation interne.

Soient des réels κ_l, κ_u , un compact Θ_μ de \mathbb{R}^d et un compact convexe Θ_+ du cône des matrices de variance-covariance $d \times d$, on définit $\Theta := \Theta_\mu \times \Theta_+ \times [\kappa_l, \kappa_u]$. Pour $\theta = (\mu, \Sigma, c) \in \Theta$, on note $q_\theta = \mathcal{N}_d(0, c\Sigma)$. Soit $\bar{\alpha} \in]0, 1[$, un taux d'acceptation-rejet moyen cible; $\alpha(x, y)$ désigne le ratio d'acceptation-rejet $\alpha(x, y) = 1 \wedge [\pi(y)/\pi(x)]$.

Algorithme 1. Initialisation Choisir $X_0 \in \mathsf{X}$, $(\mu_0, \Sigma_0, c_0) \in \Theta$.

Itération Etant donné $(X_n, \mu_n, \Sigma_n, c_n)$:

- a. Simuler $Z_{n+1} \sim q_{\theta_n}$ et poser $Y_{n+1} = X_n + Z_{n+1}$. Avec probabilité $\alpha(X_n, Y_{n+1})$, poser $X_{n+1} = Y_{n+1}$ et avec probabilité $1 - \alpha(X_n, Y_{n+1})$, poser $X_{n+1} = X_n$.

- b. Mise à jour des paramètres:

$$\mu = \mu_n + (n+1)^{-1} (X_{n+1} - \mu_n), \quad (3.9)$$

$$\Sigma = \Sigma_n + (n+1)^{-1} \left[(X_{n+1} - \mu_n) (X_{n+1} - \mu_n)^T - \Sigma_n \right], \quad (3.10)$$

$$c = c_n + \frac{1}{n+1} (\alpha(X_n, Y_{n+1}) - \bar{\alpha}). \quad (3.11)$$

- c. Si $(\mu, \Sigma, c) \in \Theta$, poser $\mu_{n+1} = \mu$, $\Sigma_{n+1} = \Sigma$ and $c_{n+1} = c$. Sinon, poser $\mu_{n+1} = \mu_n$, $\Sigma_{n+1} = \Sigma_n$ et $c_{n+1} = c_n$.

Cet algorithme a été proposé par [31] et étudié sous l'hypothèse que π est à décroissance sur-exponentielle. Nous l'étudions sous les hypothèses

F1 π est strictement positive et continue sur \mathbb{R}^d , deux fois continûment différentiable pour $|x|$ grand.

F2 il existe $m \in (0, 1)$, des constantes strictement positives $d_i < D_i$, $i = 0, 1, 2$ et $r > 0$ telles que pour tout x grand

- (i) $\langle \frac{\nabla \pi(x)}{|\nabla \pi(x)|}, \frac{x}{|x|} \rangle \leq -r$.
- (ii) $d_0|x|^m \leq -\log \pi(x) \leq D_0|x|^m$,
- (iii) $d_1|x|^{m-1} \leq |\nabla \log \pi(x)| \leq D_1|x|^{m-1}$,
- (iv) $d_2|x|^{m-2} \leq |\nabla^2 \log \pi(x)| \leq D_2|x|^{m-2}$.

F3 il existe $s_* > 0$, $0 < v < 1 - m$ et $0 < \eta < 1$ tels que quand $|x| \rightarrow +\infty$,

$$\sup_{\theta \in \Theta} \int_{\{z, |z| \geq \eta|x|^v\}} \left(1 \vee \frac{\pi(x)}{\pi(x+z)}\right)^{s_*} q_\theta(z) \lambda(dz) = o(|x|^{2(m-1)}) .$$

Les conditions F1 et F2 définissent une classe de densités cibles pour lesquelles on sait établir l'ergodicité sous-géométrique de l'échantillonneur de Hastings-Metropolis à marche aléatoire symétrique lorsque la loi de proposition est à support compact ([17, 7] et Section 2.3.1). La condition F3 permet de relâcher cette condition de compacité du support de q et d'établir l'ergodicité sous-géométrique des noyaux dans le cas où q est une loi gaussienne.

Théorème 24. [2, Proposition 3.5] Supposons F1 à F3. Soit le processus $\{X_n, n \geq 0\}$ décrit par l'algorithme 1.

(i) Si $\mathbb{E}[|\ln \pi|^{2/m}(X_0)] < +\infty$,

$$\sup_{\{f, |f|_1 \leq 1\}} |\mathbb{E}[f(X_n)] - \pi(f)| \rightarrow 0 .$$

(ii) Il existe $0 < s \leq s_*$ telle que si $\mathbb{E}[\pi^{-s}(X_0)] < +\infty$, alors pour tout $r \in [0, s[$ et pour toute fonction f telle que $|f|_{1+\pi^{-r}} < +\infty$

$$n^{-1} \sum_{k=1}^n f(X_k) \rightarrow \pi(f) , \quad \mathbb{P} - p.s.$$

Nous obtenons des résultats de convergence pour des fonctions croissant comme V pour V de la forme $V = 1 + \pi^{-s}$ car nous avons montré que cette fonction vérifiait une condition de drift sous-géométrique $P_\theta V \leq V - \phi \circ V + b \mathbb{I}_{\mathcal{C}}$. En modifiant F3 et les conditions sur la loi de X_0 , on peut montrer que des fonctions de la forme $V(x) = 1 = |x|^r$ ou $V(x) = (-\ln \pi)^s$ sont solutions d'une condition de drift sous-géométrique, et en déduire des résultats de convergence pour d'autres familles de fonctions (cf. [2]).

3.2.4 Application à la convergence de l'échantillonneur *Equi-Energy*

L'échantillonneur *Equi-Energy* décrit en Section 1.2.4 est un exemple d'adaptation externe. Si le processus auxiliaire $\{Y_n, n \geq 0\}$ est tel que $n^{-1} \sum_{k=1}^n f(Y_k)$ converge vers $\tilde{\pi}(f)$ presque-sûrement pour toute fonction bornée f , la mesure empirique $\theta_n = n^{-1} \sum_{k=1}^n \delta_{Y_k}$ converge faiblement vers $\tilde{\pi}$. Par suite,

asymptotiquement, la dynamique de X_n est donnée par $P_{\tilde{\pi}}$. Puisque par construction des noyaux P_θ on a $\pi P_{\tilde{\pi}} = \pi$, il est légitime de penser que π gouverne la loi de $\{X_n, n \geq 0\}$ asymptotiquement.

Les résultats que nous avons obtenu nous permettent de démontrer rigoureusement cette intuition. Nous supposons que $\mathsf{X} = \mathbb{R}^d$ par simplicité; mais cette condition pourrait être affaiblie en supposant que X est un espace Polonais.

G1 (a) π est strictement positive, continue sur X et $\sup_{\mathsf{X}} \pi < +\infty$.

(b) Il existe $\beta \in]0, 1[$ tel que $\tilde{\pi} = \pi^\beta$.

G2 P est un noyau de transition psi-irréductible tel que $\pi P = \pi$. Les ensembles de niveau $\{x \in \mathsf{X}, \pi(x) \geq p\}$ sont 1-small pour P , pour tout $p > 0$.

G3 (a) Il existe $\tau \in]0, 1]$, $\lambda \in]0, 1[$, $b < +\infty$ et un ensemble $\mathcal{C} \in \mathcal{X}$ tels que $PV \leq \lambda V + b \mathbb{1}_{\mathcal{C}}$ avec $V(x) := (\pi(x)/\sup_{\mathsf{X}} \pi)^{-\tau(1-\beta)}$.

(b) La probabilité d'interaction ϵ est bornée $0 \leq \epsilon < (1 - \lambda) [1 - \lambda + \tau(1 - \tau)^{\frac{1-\tau}{\tau}}]^{-1}$.

G4 Pour toute fonction bornée $f : \mathsf{X} \rightarrow \mathbb{R}$, $\lim_n \theta_n(f) = \tilde{\pi}(f)$ \mathbb{P} -p.s.

Pour vérifier G2, on peut par exemple prendre pour P le noyau d'un algorithme de Hastings-Metropolis à marche aléatoire symétrique, avec une loi de proposition gaussienne. L'existence d'une fonction V vérifiant le drift géométrique G3 est alors liée à la décroissance de π dans les queues [103]. Une alternative est de considérer un échantillonneur hybride de type "Metropolis-within-Gibbs", pour lesquels on trouve des conditions sur la loi cible π entraînant l'ergodicité géométrique [20]. Enfin, la condition G4 est vérifiée par exemple dès que le processus $\{Y_n, n \geq 0\}$ est i.i.d. ou dès que c'est une chaîne de Markov ergodique de probabilité invariante $\tilde{\pi}$.

Les conditions G1 et G3 nous permettent d'établir une condition de drift géométrique uniforme en θ . Combinées avec G2, nous prouvons E2 et E5. Le fait que X soit polonais, que π soit continue, et que le processus auxiliaire $\{Y_n, n \geq 0\}$ vérifie G4 permet d'établir E6. Par application du Théorème 22, nous avons

Théorème 25. [3, Théorème 1.5.4] Supposons G1 à G4. Alors pour toute fonction bornée f , $\lim_n |\mathbb{E}[f(X_n)] - \pi(f)| = 0$.

Pour établir la Loi des Grands Nombres, nous supposons que le processus auxiliaire est stable en un sens plus général que celui donné par G4.

G5 $\sup_n \mathbb{E}[V(Y_n)] < +\infty$ et pour tout $\alpha \in (0, 1)$, $\lim_n \theta_n(V^\alpha) = \tilde{\pi}(V^\alpha)$ \mathbb{P} -p.s.

Cette condition est en particulier vraie si $\{Y_n, n \geq 0\}$ est une chaîne de Markov de noyau de transition Q vérifiant les deux conditions

(i) Q est psi-irréductible, apériodique et il existe des constantes $\lambda' \in]0, 1[$, $b' > 0$, et une fonction W telle que $QW \leq \lambda' W + b'$,

(ii) $|V|_W < +\infty$ et $\mathbb{E}[W(Y_0)] < +\infty$.

Théorème 26. [3, Théorème 1.5.6] Supposons G1 à G5, et $\mathbb{E}[V(X_0)] < +\infty$. Pour tout $\alpha \in]0, 1[$ et toute fonction f telle que $|f|_{V^\alpha} < +\infty$, $\lim_n n^{-1} \sum_{k=1}^n f(X_k) = \pi(f)$ \mathbb{P} -p.s.

Sous les hypothèses du Théorème 26, $\sup_n \mathbb{E}[V(X_n)] < \infty$. Par suite, les résultats ci-dessus disent que si le processus auxiliaire $\{Y_n, n \geq 0\}$ possède des V -moments qui sont uniformément bornés et qu'il vérifie une loi des grands nombres pour une classe de fonctions, alors le processus $\{X_n, n \geq 0\}$ hérite des mêmes propriétés. On peut donc répéter ce schéma, et construire un processus $\{Z_n, n \geq 0\}$ à partir du processus auxiliaire $\{X_n, n \geq 0\}$ - comme on a construit $\{X_n, n \geq 0\}$ à partir de $\{Y_n, n \geq 0\}$ - et ce processus $\{Z_n, n \geq 0\}$ héritera des mêmes propriétés. On établit ainsi par récurrence l'ergodicité et la loi des grands nombres pour l'échantillonneur *Equi-Energy* à K niveaux ($K > 2$) (cf. Section 1.2.4).

3.3 Méthode de stratification adaptative

Nous utilisons ci-après les notations introduites section 1.1.1 consacrée à la description de la méthode de stratification.

L'efficacité de la méthode de stratification, mesurée en termes de réduction de variance, dépend de tous les paramètres d'implémentation; les paramètres "optimaux" n'étant pas universels mais liés à la loi de la v.a. d'intérêt $\phi(X)$. Le choix de la variable de stratification est guidé par la possibilité d'obtenir des réalisations de v.a. de loi $\mathbb{P}(X \in \cdot | Y \in \mathcal{A}_i)$; la variable est aussi choisie de sorte que les poids des strates $\mathbb{P}(Y \in \mathcal{A}_i)$ soient calculables. Idéalement, les strates devraient être ajustées aux ensembles de l'espace où $\phi(X)$ est constant. Quand l'espace de simulation est grand et que la fonction à intégrer est complexe, il est illusoire d'espérer déterminer les strates et l'allocation optimales dans une étape préliminaire à la méthode de stratification.

Nous avons proposé une nouvelles méthode qui détermine les strates et l'allocation optimale "en ligne" i.e. au fur et à mesure du déroulement de l'algorithme: l'algorithme que nous proposons combine simulation, et adaptation des strates et de l'allocation.

Pour ce faire, nous avons tout d'abord explicité la forme de la variance asymptotique, l'asymptotique étant à prendre ici au sens où le nombre de simulations n et le nombre de strates I tendent vers $+\infty$.

Nous avons ensuite proposé un algorithme itératif, que l'on peut présenter comme la construction d'une famille d'estimateurs stratifiés qui diffèrent par le choix des strates et de l'allocation; ces deux quantités étant mises à jour à chaque itération en fonction du comportement passé de l'algorithme. La dynamique qui guide cette mise à jour est la recherche des strates et de l'allocation qui minimisent la variance asymptotique.

Nous avons enfin appliqué cette méthode de stratification adaptative à un problème d'ingénierie financière; et validé l'intérêt de la méthode en la comparant (numériquement) à d'autres implémentations de l'estimateur stratifié et à d'autres méthodes de réduction de variance.

Dans les simulations, nous avons observé la convergence de l'algorithme; mais la démonstration rigoureuse de la convergence et l'identification des points limites des paramètres adaptés restent encore un problème ouvert.

Nous présentons ci-après les résultats obtenus. Nous avons considéré le cas où X est à valeur dans \mathbb{R}^d et la variable de stratification Y est de la forme $\mu^T X$ où μ est une matrice orthonormée $d \times m$, pour $m \leq d$. Pour simplifier l'exposé, nous ne présentons dans ce mémoire que les résultats obtenus avec $m = 1$. Le cas général est traité dans [8].

3.3.1 Variance asymptotique de l'estimateur stratifié

Dans toute la suite, on considère des densités par rapport à la mesure de Lebesgue sur \mathbb{R} notée λ . Pour une densité h sur \mathbb{R} , on définit H^{-1} l'inverse généralisée de la fonction de répartition H associée $H^{-1}(t) = \inf\{x \in \{H > 0\} : H(x) \geq t\}$.

Le choix des strates (ici, une partition de \mathbb{R}) est paramétré par une densité de probabilité g sur \mathbb{R} de fonction de répartition G :

$$\mathcal{A}_i := \left] G^{-1}\left(\frac{i-1}{I}\right), G^{-1}\left(\frac{i}{I}\right) \right]$$

Le choix de l'allocation est paramétré par une densité de probabilité χ sur \mathbb{R} :

$$q_i := \int_{\mathcal{A}_i} \chi \, d\lambda.$$

On pose ψ_μ et ζ_μ resp. l'espérance et le moment d'ordre 2 conditionnels

$$\psi_\mu(y) := \mathbb{E} [\phi(X)|\mu^T X = y], \quad \zeta_\mu(y) := \mathbb{E} [\phi^2(X)|\mu^T X = y].$$

Pour un nombre de strates I , un nombre total de simulations n , une direction μ , des strates définies par g et une allocation définie par χ , on note $\varsigma_{I,n}^2(\mu, g, \chi)$ la variance

$$\varsigma_{I,n}^2(\mu, g, \chi) := \sum_{i \in \{1, \dots, I\}, n_i > 0} \frac{1}{n_i} \mathbb{P}(\mu^T X \in \mathcal{A}_i) \operatorname{Var}(\phi(X)|\mu^T X \in \mathcal{A}_i).$$

Nous étudions le comportement asymptotique de cette variance sous les hypothèses suivantes

D1 la v.a. $\mu^T X$ possède une densité par rapport à la mesure de Lebesgue, notée f_μ .

D2 $\int \chi^2/g \, d\lambda < +\infty$ et $\operatorname{essinf}_g d\lambda(\chi/g) > 0$

D3 Pour $h \in \{f_\mu, \zeta_\mu f_\mu, \psi_\mu f_\mu\}$, $\int h^2/g \, d\lambda < +\infty$

D4 $\int f_\mu^4 \{\zeta_\mu - \psi_\mu^2\}^2 / [\chi^2 g] \, d\lambda < +\infty$

Dans le cas de l'allocation proportionnelle ($\chi = f_\mu$) et de strates équiprobables ($g = f_\mu$), les conditions D2 à D4 sont impliquées par l'existence du moment $\mathbb{E}[\phi^4(X)]$.

Théorème 27. [8, Proposition 1] Soit $\mu \in \mathbb{R}^d$ de norme 1, et g, χ des densités de probabilité sur \mathbb{R} vérifiant D1 à D4. Alors

$$\lim_{I \rightarrow +\infty} \lim_{n \rightarrow +\infty} n \, \varsigma_{I,n}^2(\mu, g, \chi) = \int f_\mu^2 \ (\zeta_\mu - \psi_\mu^2)^2 \ \chi^{-1} \, d\lambda.$$

Lorsque le nombre de strates I_n est une fonction du nombre de tirages n , on obtient la même variance asymptotique lorsque $n \rightarrow +\infty$ si, par exemple, $I_n^{-1} + I_n^2/n \rightarrow 0$.

Le Théorème 27 montre que la variance asymptotique ne dépend pas du choix des strates g , et que seules la direction de stratification μ et l'allocation χ apparaissent dans l'expression de la variance limite. Ce résultat justifie la construction des strates qui est utilisée en pratique, et qui consiste à choisir des strates équiprobables (donc prendre $g = f_\mu$).

Etant donnée la direction de stratification μ , l'allocation χ qui minimise la variance asymptotique est donnée par $\chi_\mu^* \propto f_\mu \sqrt{\zeta_\mu - \psi_\mu^2}$ en supposant que la constante de normalisation de cette loi soit finie. La variance asymptotique évaluée avec $\chi = \chi_\mu^*$ est alors égale à

$$\left\{ \mathbb{E} \left[\sqrt{\text{Var}(\phi(X)|\mu^T X)} \right] \right\}^2. \quad (3.12)$$

Ce résultat est compatible avec l'expression de la variance asymptotique établie par [62] dans le cas $m = 1$, avec des strates équiprobables ($g = f_\mu$) et pour l'allocation proportionnelle ($\chi = f_\mu$); la variance asymptotique est alors $\mathbb{E} [\text{Var}(\phi(X)|\mu^T X)]$ ([62, Lemme 4.1]). La variance que nous obtenons avec l'allocation $\chi = \chi_\mu^*$ est donc strictement inférieure à celle obtenue avec l'allocation proportionnelle (sauf dans le cas où la variance conditionnelle est constante).

L'allocation optimale χ_μ^* n'est pas calculable en pratique. Le résultat suivant montre que la quantité $\int_{\mathcal{A}_i} \chi_\mu^* d\lambda$ peut être approchée par l'allocation optimale Q^* donnée par (1.2) - que nous noterons dorénavant $Q^*(\mu)$ pour insister sur sa dépendance en la direction de stratification; et que la variance asymptotique optimale (3.12) est la variance asymptotique de l'estimateur stratifié à allocation optimale.

Théorème 28. [8, Proposition 2] Soient $\mu \in \mathbb{R}^d$ une direction de stratification de norme 1, et g une densité de probabilité sur \mathbb{R} vérifiant D1 et D3. Alors

$$\lim_{I \rightarrow +\infty} \sum_{i=1}^I \left| q_i^*(\mu) - \int_{\mathcal{A}_i} \chi_\mu^* \right| = 0.$$

Si de plus, $I_n^{-1} + I_n/n \rightarrow 0$ quand $n \rightarrow +\infty$ alors

$$\lim_n n \varsigma_{I_n, n}^2(\mu, g, Q^*(\mu)) = \left\{ \mathbb{E} \left[\sqrt{\text{Var}(\phi(X)|\mu^T X)} \right] \right\}^2.$$

3.3.2 Algorithme de stratification adaptative

Les résultats présentés en Section 3.3.1 suggèrent d'appliquer l'algorithme de stratification à allocation optimale, et d'optimiser le choix de la direction de stratification μ plutôt que le choix des strates. Néanmoins, la dépendance de la variance asymptotique en μ est complexe: il n'y a pas de solution explicite du fait que les moments conditionnels ψ_μ et ζ_μ ne sont pas connus en pratique.

Vu le Théorème 27, nous choisissons les strates définies par $g = f_\mu$. Nous utilisons la caractérisation de la variance asymptotique (3.12) donnée par le Théorème 28, et proposons d'appliquer l'algorithme stratifié avec l'allocation optimale et pour une direction de stratification μ qui minimise la variance $\varsigma_{I,n}^2(\mu, f_\mu, Q^*(\mu))$.

Cette variance n'est pas explicite mais est une fonction régulière d'espérances. Nous avons montré que la dérivée de ces espérances par rapport à la direction μ s'écrivait sous la forme d'une intégrale [8, Proposition 3]. Nous proposons donc une approche de type gradient stochastique pour déterminer de façon itérative le minimum de cette fonction: à chaque itération, les espérances qui définissent le gradient sont approchées par un estimateur stratifié calculé à partir de l'allocation courante, et de la direction de stratification courante.

A chaque itération t de cet algorithme, nous disposons donc aussi d'un estimateur stratifié $\mathcal{E}^{(t)}$ de la quantité d'intérêt $\mathbb{E}[\phi(X)]$. Nous utilisons cette famille d'estimateurs $\{\mathcal{E}^{(t)}, t \geq 0\}$ pour déterminer un estimateur moyennisé de $\mathbb{E}[\phi(X)]$ de la forme $[\sum_{t=1}^T \alpha^{(t)}]^{-1} \sum_{t=1}^T \alpha^{(t)} \mathcal{E}^{(t)}$ où le poids $\alpha^{(t)}$ est égal à l'inverse de (un estimateur de) la variance de $\mathcal{E}^{(t)}$.

Algorithme AdaptStr. Soit $\{\gamma_t, t \geq 0\}$ une suite de réels strictement positifs.

- 1) Initialisation : se donner une direction initiale $\mu^{(0)}$ et une allocation initiale. Calculer le poids de chaque strate $\mathbb{P}([\mu^{(0)}]^T X \in \mathcal{A}_i)$ pour $i \in \{1, \dots, I\}$.
- 2) A l'itération $(t + 1)$, étant données la direction courant $\mu^{(t)}$, l'allocation courante $q^{(t)}$ et les poids de chaque strate $\mathbb{P}([\mu^{(t)}]^T X \in \mathcal{A}_i)$,
 - a) faire n tirages, alloués à chaque strate selon l'allocation courante; les tirages dans la strate i étant sous la loi conditionnelle $\mathbb{P}(X \in \cdot | [\mu^{(t)}]^T X \in \mathcal{A}_i)$.
 - b) utiliser ces tirages pour estimer le gradient (par rapport à μ) de la variance $\varsigma_{I,n}^2(\mu, f_\mu, q)$ au point $\mu = \mu^{(t)}$ et $q = q^{(t)}$. Cette estimation est notée $\widehat{\nabla V}(\mu^{(t)})$.
 - c) mettre à jour la direction de stratification selon la dynamique $\mu^{(t+1)} \propto \mu^{(t)} - \gamma_t \widehat{\nabla V}(\mu^{(t)})$.
 - d) utiliser ces tirages pour mettre à jour l'allocation selon l'équation (1.2) dans laquelle les variances conditionnelles sont approchées par un estimateur stratifié.
 - e) mettre à jour les poids de chaque strate (calcul exact ou estimation).
 - f) utiliser ces tirages pour calculer un estimateur stratifié de $\mathbb{E}[\phi(X)]$, noté $\mathcal{E}^{(t)}$; et pour calculer un estimateur de sa variance.
- 3) calculer un estimateur moyennisé de la quantité d'intérêt $\mathbb{E}[\phi(X)]$.

Classiquement, on choisit la suite de pas $\{\gamma_t, t \geq 0\}$ de sorte que $\sum_t \gamma_t = +\infty$ et $\sum_t \gamma_t^2 < +\infty$. Mais si le nombre d'itérations est fixé à l'avance, on peut prendre un pas fixe [92].

3.3.3 Application à l'ingénierie financière

Le pricing d'une option consiste à calculer $\mathbb{E}[\Xi(X)]$ pour une fonction $\Xi : \mathbb{R}^d \rightarrow \mathbb{R}^+$, et $X \sim \mathcal{N}_d(0, \text{Id})$; cette structure gaussienne résultant par exemple de la discréétisation d'une diffusion guidée par un mouvement brownien, modélisant l'évolution de l'actif.

La formule de Cameron-Martin entraîne que pour tout $\nu \in \mathbb{R}^d$,

$$\mathbb{E}[\Xi(X)] = \mathbb{E}[\Xi(Z) \exp(-\nu^T Z + 0.5\nu^T \nu)], \text{ avec } Z = X + \nu,$$

L'algorithme de stratification adaptative peut s'appliquer à l'une ou l'autre des espérances. Dans le second cas, on parlera de stratification combinée avec de l'échantillonnage d'importance; le choix du vecteur de drift ν joue un rôle sur la variance de cet estimateur hybride.

Nous avons comparé l'estimateur de stratification adaptative - désigné ci-après par **AdaptStr** - à des estimateurs de stratification

(i) combiné ou pas avec de l'échantillonnage d'importance en prenant ν solution du problème

d'optimisation $\nu^* := \operatorname{argmax}_{\{\nu \in \mathbb{R}^d, \Xi(\nu) > 0\}} (\Xi(\nu) - 0.5\nu^T \nu)$ (choix proposé par [62]). Dans les tableaux 3.1 et 3.2, ces deux cas correspondent resp. à “ ν ” valant “0” puis “ ν_* ”.

(ii) à allocation proportionnelle et à allocation optimale, cette dernière étant déterminée de façon itérative. Dans les tableaux 3.1 et 3.2, ces deux cas correspondent respectivement à “ q_i ” valant “prop” et “opt”.

(iii) dont la direction de stratification μ est fixe et colinéaire au drift ν_* ; ou colinéaire au vecteur de régression linéaire de la fonction $\phi(X)$ sur X ; ou colinéaire à un vecteur μ_l assez intuitif vue l'application considérée. Dans les tableaux 3.1 et 3.2, ces trois cas correspondent respectivement “ μ ” valant “ ν_* , μ_{reg} , μ_l ”.

(iv) tel que proposé par [62]; cet estimateur combine échantillonnage d'importance avec le vecteur de drift égal à ν_* , et stratification à allocation proportionnelle et direction fixée à une certaine valeur μ_g définie dans [62, Section 4.2]. Dans les tableaux 3.1 et 3.2, cela correspond à la colonne “GHS”.

Plusieurs options ont été considérées: options de type **asiatiques**, **barrière**, **basket**, ou **asiatiques** dans un modèle de volatilité stochastique d'Heston. Nous ne reportons ci-dessous que les résultats numériques relatifs aux options de type **asiatiques** et **barrière**.

Options asiatiques

On s'intéresse au pricing d'une option asiatique arithmétique dans un modèle de Black-Scholes; le prix du sous-jacent est décrit par l'EDS $\frac{dS_t}{S_t} = r dt + \sigma dW_t$, $S_0 = s_0$, où $\{W_t, t \geq 0\}$ est un mouvement brownien standard. Le sous-jacent est discrétisé sur une grille régulière de sorte que

$$\Xi(x) = \exp(-rT) \left(\frac{s_0}{d} \sum_{k=1}^d \exp \left((r - 0.5\sigma^2) \frac{kT}{d} + \sigma \sqrt{\frac{T}{d}} \sum_{j=1}^k x_j \right) - K \right)_+, \quad x = (x_1, \dots, x_d) \in \mathbb{R}^d.$$

Dans les applications numériques, $(\nu, K) \in \{(0.1, 45), (0.5, 45), (0.5, 65), (1, 45), (1, 65)\}$, $s_0 = 50$, $r = 0.05$, $T = 1$, et $d = 16$. La direction de stratification μ_l assez intuitive dans cette application, est $\mu_l \propto (d, d-1, \dots, 1)$.

Options barrière

On s'intéresse au pricing d'une option barrière dont le *payoff* est donné par

$$\Xi(x) = \exp(-rT) \left(\frac{s_0}{d} \sum_{k=1}^d \exp \left((r - 0.5\sigma^2) \frac{kT}{d} + \sigma \sqrt{\frac{T}{d}} \sum_{j=1}^k x_j \right) - K \right)_+ \mathbb{1}_{\{S_T(x) \leq B\}},$$

le sous-jacent à terme valant $\ln[S_T(x)/s_0] = (r - 0.5\sigma^2)T + \sigma \sqrt{\frac{T}{d}} \sum_{j=1}^d x_j$. Dans les applications numériques, $s_0 = 50$, $r = 0.05$, $T = 1$, $\sigma = 0.1$, $d = 16$ et $(K, B) \in \{(50, 60), (50, 80)\}$. Dans cet exemple, la direction de stratification assez intuitive est $\mu_l \propto (d, d-1, \dots, 1)$.

Implémentation

Tous les algorithmes de stratification sont implémentés en prenant $I = 100$, et pour un nombre total de simulations égal à $20\,000 \times 200$.

Pour les procédures itératives, ces simulations sont réparties en 20 000 tirages pour chacune des 200 itérations. La direction de stratification est initialisée en $\mu^{(0)} \propto (1, \dots, 1)$, et l'allocation initiale est l'allocation proportionnelle.

Figures

La figure 3.12 illustre la convergence de l'algorithme **AdaptStr**: à gauche, on représente l'évolution des d composantes de la direction de stratification $\mu^{(t)}$, en fonction du nombre d'itérations t . A droite, on trace l'évolution de l'estimation $\mathcal{E}^{(t)}$ et de sa variance, en fonction du nombre d'itérations. On peut ob-

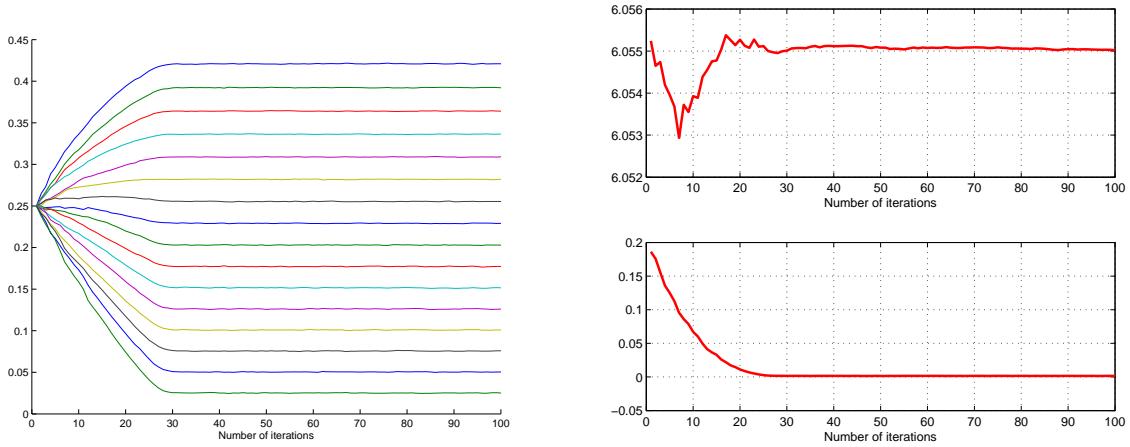


Fig. 3.12: Option Asiatique lorsque $(v, K, \nu) = (0.1, 45, \nu_*)$. [gauche] directions de stratification successives $t \mapsto \mu^{(t)}$. $\mu^{(0)}$ est proportionnelle au vecteur $(1, \dots, 1)$ donc les d courbes sont issues du même point $1/\sqrt{d}$. Par convention, la première composante de $\mu^{(t)}$ est positive. [droite, haut] valeurs successives de l'estimation $t \mapsto \mathcal{E}^{(t)}$ de la quantité d'intérêt $\mathbb{E}[\phi(X)]$. [droite, bas] valeurs successives de l'estimation de la variance de $\mathcal{E}^{(t)}$; la valeur limite est 0.0016.

server en particulier l'influence de la direction de stratification sur la variance de l'estimateur: au long des itérations de **AdaptStr**, la variance décroît de 0.1862 à 0.0016. Ces graphes illustrent la convergence de l'algorithme; nous n'avons pas observé de sensibilité du point limite à différents paramètres d'implémentation comme par exemple l'initialisation de la direction de stratification. Le choix de l'initialisation et des pas $\{\gamma_t, t \geq 0\}$ agit en revanche sur la vitesse de convergence de l'algorithme.

Sur la figure 3.13 on trace différentes directions de stratification pour l'étude des options asiatiques : la direction μ_g proposée par [62]; la direction proportionnelle à ν_* ; la direction à convergence de l'algorithme **AdaptStr** $\mu^{(N)}$; la direction de régression linéaire μ_g et la direction μ_l . On peut en particulier comparer l'influence de l'introduction d'une étape d'échantillonnage d'importance; l'influence est assez faible sur la valeur de la direction de stratification limite $\mu^{(N)}$ obtenue par **AdaptStr** mais elle est notable en termes de réduction de variance (cf. Tableau 3.1). L'efficacité de l'algorithme **AdapStr** est donc sensible au choix du drift ν .

La figure 3.14 compare différentes directions de stratification pour l'étude des options barrière: la direction μ_g proposée par [62]; la direction proportionnelle à ν_* ; la direction à convergence de l'algorithme **AdaptStr** $\mu^{(N)}$; la direction proportionnelle au vecteur de régression linéaire μ_{reg} ; et la

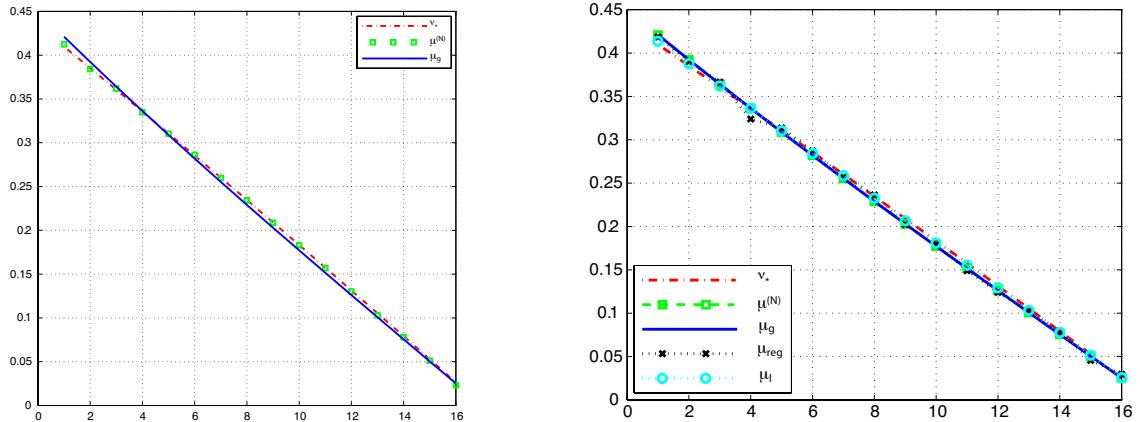


Fig. 3.13: Option asiatique [gauche] dans le cas $(v, K, \nu) = (0.1, 45, 0)$ et [droite] dans le cas $(v, K, \nu) = (0.1, 45, \nu_*)$. Directions de stratification μ_g , $\mu^{(N)}$, ν_* (ν_* est normalisé pour être de norme 1), μ_g et μ_l

direction μ_l .

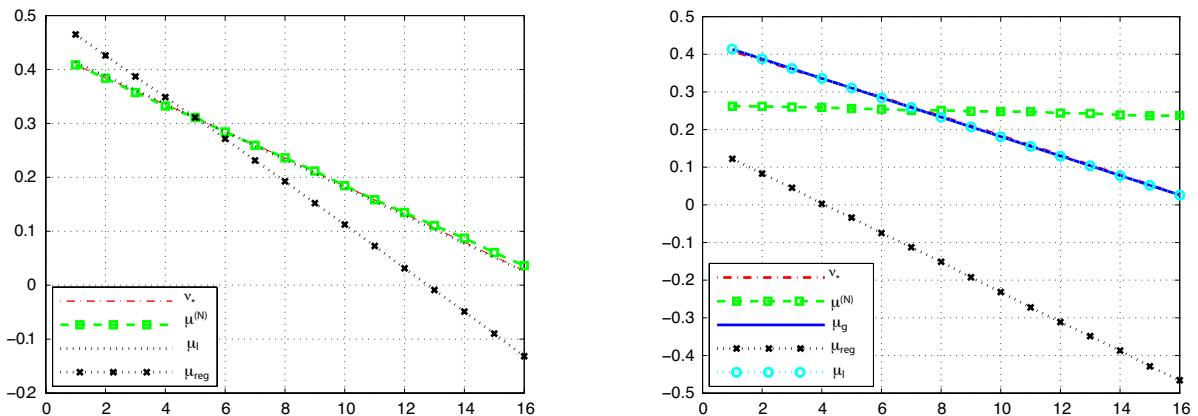


Fig. 3.14: Option Barrière dans le cas $(K, B) = (50, 60)$ et [gauche] $\nu = 0$ ou [droite] $\nu = \nu_*$: directions $\mu^{(N)}$, μ_{reg} , μ_l , ν_* et μ_g . ν_* a été normalisée pour être de norme 1.

Sur cet exemple, les directions de stratification sont plus différenciées: la direction $\mu^{(N)}$ ne coïncide pas avec les directions μ_g , μ_{reg} et μ_l ; $\mu_g \sim \mu_l$ mais la direction de régression μ_{reg} est loin d'être optimale (cf. Tableau 3.2).

Tableaux

Dans les tableaux 3.1 et 3.2, nous présentons la variance des différents estimateurs. Pour comparaison, nous reportons aussi la variance de l'estimateur de Monte Carlo standard (colonne “MC”).

Ces résultats numériques montrent l'intérêt de l'estimateur stratifié par rapport à celui de Monte Carlo standard. Par exemple, pour les options asiatiques, la variance est parfois réduite d'un facteur 2 500.

Les estimateurs stratifiés sont plus efficaces avec l'allocation optimale qu'avec l'allocation proportionnelle : il est donc conseillé d'implémenter un estimateur stratifié itératif, qui estime l'allocation optimale au fur et à mesure des itérations. Etoré & Jourdain ont établi un Théorème Central Lim-

ite pour ces algorithmes adaptatifs (dans lesquels seule l'allocation est adaptée) qui montre que ces algorithmes atteignent la variance de l'estimateur stratifié à allocation optimale [56].

L'estimateur stratifié **AdaptStr** est plus efficace lorsqu'il est combiné avec de l'échantillonnage d'importance avec drift ν_* . Comparé aux autres estimateurs stratifiés, il apparaît comme plus (ou aussi) efficace. Des procédures de stratification plus simples à mettre en oeuvre peuvent se révéler parfois très performantes, ou au contraire bien moins intéressantes que des procédures itératives, selon le modèle et la valeur des paramètres de ces modèles (cf. par exemple le cas $\mu = \mu_{\text{reg}}$). En revanche, quelque soit le modèle et les paramètres de ces modèles, **AdaptStr** donne la meilleure réduction de variance.

Ces résultats numériques justifient donc l'intérêt de procédures adaptatives pour déterminer la direction de stratification et l'allocation optimale.

Model			Price	Variance						
v	K	ν		q_i	MC	AdaptStr	GHS	μ_{reg}	μ_*	μ_l
0.1	45	0	prop	6.05	8.640	-	-	0.017	0.016	0.017
			opt	6.05	8.640	0.004	-	0.005	0.004	0.004
		ν_*	prop	6.05	0.803	-	0.014	0.014	0.008	0.007
			opt	6.05	0.803	0.002	-	0.005	0.002	0.002
0.5	45	0	prop	9.00	158.1	-	-	2.086	2.128	2.243
			opt	9.00	158.1	0.352	-	0.362	0.371	0.390
		ν_*	prop	9.00	14.95	-	0.203	0.225	0.324	0.221
			opt	9.00	14.95	0.147	-	0.162	0.223	0.162
0.5	65	0	prop	2.16	48.41	-	-	1.857	1.859	2.097
			opt	2.16	48.41	0.093	-	0.096	0.096	0.147
		ν_*	prop	2.16	2.32	-	0.039	0.046	0.048	0.049
			opt	2.16	2.32	0.020	-	0.024	0.025	0.026
1	45	0	prop	14.01	852.0	-	-	52.24	54.51	57.72
			opt	14.01	852.0	5.39	-	5.48	5.69	6.02
		ν_*	prop	14.01	42.76	-	3.014	3.185	4.360	3.265
			opt	14.01	42.76	2.270	-	2.400	3.220	2.450
1	65	0	prop	7.79	587	-	-	50.9	50.5	55.8 ;
			opt	7.79	587	3.75	-	3.01	3.08	3.95
		ν_*	prop	7.78	22.34	-	1.55	1.75	2.01	1.56
			opt	7.78	22.34	0.99	-	1.14	1.31	1.00

Tableau 3.1: Option asiatique: estimateurs stratifiés et estimateur de Monte Carlo usuel

Model				Price		Variance						
				ν	q_i	-	MC	AdaptStr	GHS	μ_{reg}	μ_*	μ_l
50	60	0	prop	1.38		2.99		-	-	1.46	1.13	1.14
			opt	1.38		2.99		0.31	-	0.83	0.31	0.31
		ν_*	prop	1.38		1.34		-	0.50	1.15	0.49	0.50
			opt	1.38		1.34		0.17	-	1.12	0.31	0.31
50	80	0	prop	1.92		4.92		-	-	0.016	0.017	0.016
			opt	1.92		4.92		0.002	-	0.002	0.002	0.002
		ν_*	prop	1.92		0.704		-	0.0011	0.0012	0.0013	0.0011
			opt	1.92		0.704		0.0005	-	0.0006	0.0006	0.0005

Tableau 3.2: Option Barrière: estimateurs stratifiés et estimateur de Monte Carlo usuel

Chapitre 4

Application des méthodes de simulation à l'analyse statistique de données complexes

Les résultats que nous présentons dans ce chapitre sont issus des articles

- [9] F. Forbes and G. Fort. Combining Monte Carlo and Mean Field-like methods for inference in Hidden Markov Random Fields. *IEEE Transactions on Image Processing*, 16(3):824–837, 2007.
- [23] D. Wraith, M. Kilbinger, K. Benabed, O. Cappé, J.F. Cardoso, G. Fort, S. Prunet, and C.P. Robert. Estimation of cosmological parameters using adaptive importance sampling. *Phys. Rev. D*, 80(2):18 pages, 2009.

Motivation. Nous présentons deux applications des méthodes de simulation à l'analyse statistique de données complexes.

Il s'agit tout d'abord de résoudre un problème d'inférence et de restauration dans un modèle de champ de Markov caché. Les algorithmes d'inférence habituels ne s'appliquent pas dans cette situation du fait de la complexité du modèle: complexité due à la structure de dépendance des données cachées et à la dimension de ces données. Nous avons donc proposé un algorithme combinant méthode de relaxation basée sur une approche variationnelle de type champ moyen, et méthode de simulations.

Dans la seconde application, il s'agit de résoudre un problème d'inférence (bayésienne) de paramètres cosmologiques. La complexité de ce traitement statistique se traduit ici par un temps de calcul de la vraisemblance prohibitivement long. Par suite, on souhaite que l'exploration de la loi a posteriori soit menée par des échantillonneurs dont on peut paralléliser l'implémentation; les méthodes MCMC par exemple n'ont pas cette souplesse du fait de l'aspect séquentiel de la construction d'une chaîne de Markov. Nous avons étudié les performances d'un échantillonneur d'importance adaptatif pour résoudre ce problème d'inférence.

4.1 Combiner méthodes de Monte Carlo et méthodes de champ moyen pour l'inférence dans des champs de Markov cachés

Les modèles à données manquantes permettent de modéliser la variabilité et l'hétérogénéité des données et de résoudre des problèmes de classification. Du fait de cette structure cachée, les procédures habituelles pour l'inférence et l'estimation de paramètres dans ces modèles ne peuvent être appliquées lorsque l'on pose un modèle d'interactions non triviales entre les données. Cela nécessite donc l'introduction de techniques d'approximation.

Nous avons considéré un problème d'estimation de paramètres dans un modèle de champ de Markov caché. Dans ce contexte, l'objectif est de calculer une quantité d'intérêt caractérisée comme la solution d'un problème d'optimisation. Ce problème n'a pas de solutions explicites du fait de la grande dimension de la structure cachée et de la complexité de la dépendance probabiliste entre les variables (cachées). Cela nécessite de simplifier le problème d'optimisation, en simplifiant l'ensemble des contraintes et/ou la fonction à optimiser. Nous avons considéré la combinaison de deux formes d'approximation : la première, basée sur les méthodes variationnelles, est une technique de relaxation de l'ensemble des contraintes; la seconde, basée sur des méthodes de simulation, est une technique d'approximation de la quantité d'intérêt.

L'algorithme EM [51] permet de résoudre des problèmes d'estimation de paramètres et de restauration de données manquantes dans des modèles à données cachées. Dans le contexte des champs de Markov cachés, cet algorithme nécessite des étapes d'optimisation et d'intégration qui n'ont pas de solutions explicites. L'algorithme *Variational EM* (VEM, [71]) combine EM et approche variationnelle; la convergence du VEM a été étudiée par [35]. L'algorithme VEM n'est pas implantable lorsque le modèle caché est un champ de Markov et différentes approximations du VEM ont été proposées ([117, 40]) dans un contexte d'application à la segmentation des images. Néanmoins, ces approximations ne conservent pas les propriétés de convergence du VEM.

Nous avons proposé un algorithme de type VEM stochastique qui combine VEM et méthodes de Monte Carlo - algorithme *Monte Carlo VEM* (MCVEM) - et montré que sous certaines conditions sur la précision de l'approximation Monte Carlo, cet algorithme a les mêmes points limites que le VEM. La comparaison des points limites du VEM et de ceux de l'EM n'est pas du tout évidente lorsque la structure cachée a, comme dans le cas d'un champ de Markov caché, une structure de dépendance complexe. A défaut de savoir prouver que les points limites du MCVEM sont ceux de l'EM, nous avons fait une étude numérique comparative de différents algorithmes combinant méthodes de relaxation (pas nécessairement variationnelles) et/ou méthodes de Monte Carlo. En particulier, nous avons comparé MCVEM à un algorithme de type EM stochastique dont on peut démontrer, par application des résultats de [18], qu'il a le même comportement asymptotique que l'EM.

Nous résumons maintenant les résultats obtenus. On en trouvera une présentation détaillée dans [9]. La contribution théorique de ces travaux s'appuie sur des résultats relatifs à la convergence des algorithmes stochastiques - et notamment du Monte Carlo EM - établis durant ma thèse [18].

4.1.1 Modèle de champ de Markov caché pour la segmentation d'images

Modèle de champ de Markov caché

Soit S un ensemble de N sites muni d'un système de voisinage. On note \mathcal{C} l'ensemble des cliques. On définit un champ de Markov discret $\mathbf{Z} = \{Z_i, i \in S\}$ à valeur dans $\{1, \dots, K\}$ comme un vecteur aléatoire dont la loi jointe est donnée par $p_Z(\mathbf{z}) = W^{-1} \exp(-H(\mathbf{z}))$, où $H = \sum_{c \in \mathcal{C}} V_c(\mathbf{z}_c)$ est la fonction d'énergie et V_c sont les potentiels de cliques. On modélise les données non observées par un champ de Markov caché \mathbf{Z} dont la loi p_Z dépend d'un paramètre β , $\beta \in \mathcal{B} \subseteq \mathbb{R}$.

On suppose que les observations \mathbf{Y} sont conditionnellement indépendantes étant donné \mathbf{Z} , de loi conditionnelle $p_{Y|Z}$ dépendant d'un vecteur de paramètres $\theta \in \Theta \subseteq \mathbb{R}^{n_\theta}$. La vraisemblance complète (i.e. la vraisemblance de (\mathbf{Y}, \mathbf{Z})) est donnée par

$$p_{(Y,Z)}(\mathbf{y}, \mathbf{z} ; \theta, \beta) = p_{Y|Z}(\mathbf{y} | \mathbf{z}; \theta) p_Z(\mathbf{z}; \beta). \quad (4.1)$$

Complexité du modèle

En segmentation d'images, on cherche à reconstruire l'image $\mathbf{z} \in \{1, \dots, K\}^N$ qui s'interprète comme une classification des pixels en K ensembles. Généralement, cela nécessite l'estimation des paramètres $\psi := (\beta, \theta)$. L'algorithme EM permet d'estimer ces paramètres au sens du maximum de vraisemblance $\hat{\psi} := \text{argmax}_{(\beta, \theta) \in \mathcal{B} \times \Theta} \ln p_Y(\mathbf{y}; \psi)$ où p_Y désigne la vraisemblance des observations; et de restaurer les données manquantes \mathbf{z} à partir de leur loi a posteriori.

L'algorithme EM peut être lu comme un algorithme de maximisation alternée [49, 91] d'une fonction F définie sur $\mathcal{D} \times (\mathcal{B} \times \Theta)$ où \mathcal{D} désigne l'ensemble des probabilités sur $\{1, \dots, K\}^N$:

$$F(q, \psi) = \sum_{\mathbf{z} \in \mathcal{Z}} \ln \left(\frac{p_{(Y,Z)}(\mathbf{y}, \mathbf{z}; \psi)}{q(\mathbf{z})} \right) q(\mathbf{z}). \quad (4.2)$$

Algorithme 2. Algorithme EM

1. *Initialisation* : choisir $(q^{(0)}, \psi^{(0)}) \in \mathcal{D} \times (\mathcal{B} \times \Theta)$.

2. *Iteration t + 1 :*

a. *mise à jour de la probabilité* : $q^{(t+1)} \in \text{argmax}_{q \in \mathcal{D}} F(q, \psi^{(t)})$,

b. *mise à jour des paramètres* : $\psi^{(t+1)} \in \text{argmax}_{\psi \in \Psi} \sum_{\mathbf{z} \in \mathcal{Z}} \ln p_{(Y,Z)}(\mathbf{y}, \mathbf{z}; \psi) q^{(t+1)}(\mathbf{z})$.

La mise à jour de q est explicite et vaut $p_{Z|Y}(\cdot | \mathbf{y}; \psi^{(t)})$; dans la mise à jour de ψ , on reconnaît l'étape d'intégration (sur l'espace \mathcal{Z}) et l'étape de maximisation de l'algorithme EM. Dans ce modèle, la relation (4.1) montre que la mise à jour de β et θ se fait par résolution de deux problèmes d'optimisation indépendants.

Cet algorithme ne peut pas être implémenté dans le cas d'un champ de Markov caché : du fait de la dimension de \mathcal{Z} et de la structure de dépendance des variables, la loi a priori p_Z n'est connue qu'à sa constante de normalisation près qui dépend de β , et l'intégration dans la mise à jour de ψ n'est pas faisable. Ainsi, la quantité à optimiser dans l'étape 2b n'est pas connue explicitement. L'algorithme VEM consiste à restreindre l'ensemble des contraintes dans l'étape 2a à un sous-ensemble $\tilde{\mathcal{D}}$ de \mathcal{D} ;

classiquement, on prend \tilde{D} comme l'ensemble des distributions sur \mathcal{Z} de forme produit si bien que la mise à jour de q revient à résoudre une équation de point fixe

$$\forall i \in S, \forall e_k \in V, \quad \ln q_i^{(t+1)}(e_k) = c_i + \sum_{\mathbf{z} \in \mathcal{Z}} \ln p_{Z|Y}(\mathbf{z}|\mathbf{y}; \psi^{(t)}) \{ \delta_{e_k}(z_i) \prod_{j \neq i} q_j^{(t+1)}(z_j) \}; \quad (4.3)$$

du fait de la propriété de Markov, cette équation ne fait intervenir que les distributions q_j , pour les sites j dans le voisinage de i . Par suite, l'intégration dans (4.3) devient une somme sur un espace de petite dimension. Avec ce choix de l'ensemble \tilde{D} , tout se passe comme si l'on supposait que le rôle des pixels voisins est identique à celui d'un **champ moyen** qui agirait sur le site cible; cela revient à négliger les fluctuations des pixels qui interagissent avec le pixel cible, en fixant ces voisins à leur valeur moyenne [42, 96]. Ces valeurs moyennes sont solutions de l'équation au point fixe (4.3).

Malgré cette relaxation, la mise à jour de ψ - et plus particulièrement de β - reste infaisable puisqu'elle nécessite la connaissance de la constante de normalisation $W(\beta)$ de la loi a priori $p_Z(\cdot; \beta)$.

4.1.2 Algorithme Monte Carlo VEM (MCVEM)

L'algorithme MCVEM combine l'approche variationnelle du VEM et les méthodes de Monte Carlo pour l'approximation de la constante de normalisation $W(\beta)$.

Algorithme MCVEM

Soit \mathcal{I} l'ensemble des probabilités de forme produit sur \mathcal{Z} et \mathcal{I}_r le sous-ensemble tel que : $q \in \mathcal{I}_r \Rightarrow [\forall e_k \in V, \sum_{i=1}^N q_i(e_k) \neq 0]$; . Soient $\{\gamma^t, J_t, t \geq 0\}$ une famille de réels strictement positifs tels que $\inf_t \gamma^t > 0$; $\{\pi^t, t \geq 0\}$ une famille de probabilités sur \mathcal{Z} .

Algorithme 3. Algorithme MCVEM

1. *Initialisation* : choisir $(q^{(0)}, \psi^{(0)}) \in \mathcal{I}_r \times (\mathcal{B} \times \Theta)$.

2. *Iteration* ($t + 1$) :

- a. *Mise à jour de la loi q* : $q^{(t+1)} \in \operatorname{argmin}_{q \in \mathcal{I}_r} \sum_{\mathbf{z}} \log \left(\frac{q(\mathbf{z})}{p_{Z|Y}(\mathbf{z}|\mathbf{y}; \psi^{(t)})} \right) q(\mathbf{z})$.
- b. *Mise à jour du paramètre θ* : $\theta^{(t+1)} \in \operatorname{argmax}_{\theta \in \Theta} \sum_{\mathbf{z}} \ln p_{Y|Z}(\mathbf{y}|\mathbf{z}; \theta) q^{(t+1)}(\mathbf{z})$.
- c. *Mise à jour du paramètre β* : simuler une chaîne de Markov $\{Z^{j,t}, 1 \leq j \leq J_t\}$ de distribution invariante π^t et poser $\beta^{(t+1)} \in \operatorname{argmax}_{\beta \in \mathcal{B}^t} - \{\sum_{\mathbf{z} \in \mathcal{Z}} H(\mathbf{z}; \beta) q^{(t+1)}(\mathbf{z}) + \ln \tilde{W}^{J_t, \pi^t}(\beta)\}$ où

$$\tilde{W}^{J_t, \pi^t}(\beta) := \frac{1}{J_t} \sum_{j=1}^{J_t} \exp(-H(\mathbf{z}^{j,t}; \beta) - \ln \pi^t(\mathbf{z}^{j,t})) , \quad \mathcal{B}^t := \{\beta \in \mathcal{B}, |\beta - \beta^t| \leq \gamma^t\} .$$

MCVEM est donc un algorithme stochastique : la première étape pour assurer sa convergence est de vérifier la stabilité des trajectoires $\{(q^{(t)}, \psi^{(t)}), t \geq 0\}$ ce qui n'est pas du tout acquis pour les trajectoires produites par l'algorithme ci-dessus. Il faut donc rajouter une procédure de stabilisation : nous avons considéré l'approche par reprojctions sur compacts croissants [46, 47] - les détails sont présentés [9, Section III] (cf. [18] pour l'application de cette méthode de reprojctions à la stabilisation de l'algorithme Monte Carlo EM).

L'étape 2a est implémentée comme la recherche d'une solution de l'équation au point fixe (4.3). Cette solution n'est pas nécessairement unique et en pratique, est calculée par un algorithme itératif ([112, 118] et erratum [57]).

L'étape 2b peut être explicite selon la forme de l'attache aux données $p_{Y|Z}$ choisie; c'est le cas par exemple lorsque $p_{Y|Z}$ est une loi gaussienne.

L'étape 2c est basée sur une méthode de Monte Carlo : il s'agit de donner une approximation de la fonction $\beta \mapsto W(\beta)$ - à une constante multiplicative près - ce que l'on obtient à partir de réalisations d'une chaîne de Markov de loi invariante $\pi^{(t)}$, produite par un algorithme de Gibbs, ou de Hastings-Metropolis ou de Swendsen-Wang [111] par exemple. Nous pensons que cette approximation est valable pour β dans un voisinage de $\beta^{(t)}$. C'est la raison pour laquelle, dans la mise à jour de β , l'optimisation est seulement locale (sur \mathcal{B}^t et non sur \mathcal{B}). En pratique, nous avons pris $\pi^{(t)} \propto p_Z(\cdot; \beta^{(t)})$.

Etude du comportement asymptotique du MCVEM

L'étude du comportement asymptotique du MCVEM est une (petite) généralisation de travaux que j'avais effectués durant ma thèse [18]. Nous avions montré que si la convergence d'un algorithme itératif modélisé par une application point-à-ensemble T (i.e. construction d'une suite $\{u^t, t \geq 0\}$ telle que $u^{(t+1)} \in T(u^{(t)})$) repose sur l'existence d'une fonction de Lyapunov pour T , alors une suite produite par $v^{(t+1)} = F_t(v^{(t)})$ où $\{F_t, t \geq 0\}$ est une famille d'applications point-à-ensemble "proches" de T , peut avoir le même comportement asymptotique que la suite $\{u^{(t)}, t \geq 0\}$. Pour ce faire, les conditions portent sur la convergence des perturbations $\{F_n, n \geq 0\}$ vers T , en un sens qui traduit le fait que quand $n \rightarrow +\infty$, l'effet induit sur la fonction de Lyapunov par l'action de F_n converge vers celui induit sur la fonction de Lyapunov par l'action de T .

Traduites dans le contexte des algorithmes MCVEM (famille de perturbations aléatoires $\{F_n, n \geq 0\}$) et VEM (application déterministe T), ces conditions s'expriment en termes de contrôle de l'erreur d'approximation dans l'étape de Monte Carlo; et plus précisément de la vitesse de décroissance de cette erreur en fonction du nombre de simulations. Ce sont donc des conditions sur le nombre de simulations à chaque itération et sur l'ergodicité des échantillonneurs; dans le cas présent, ces propriétés d'ergodicité sont vérifiées sans difficultés du fait que l'espace de simulation est fini (même si il est grand!).

Sous des hypothèses qui sont essentiellement

(a) des hypothèses de régularité du modèle statistique garantissant par exemple une solution unique et explicite aux problèmes d'optimisation 2b et 2c;

(b) des hypothèses sur le contrôle de l'erreur d'approximation par la méthode de Monte Carlo; nous avons établi que, avec probabilité 1, les trajectoires du MCVEM convergent vers le même ensemble limite que le VEM (cf. [9, Théorème 2]).

4.1.3 Application du MCVEM à la segmentation d'images

Nous avons appliqué l'algorithme MCVEM à la segmentation d'images en K classes: nous avons posé un modèle de Potts (avec voisinage de premier ordre) pour les données cachées, et un modèle d'attache aux données gaussien.

Nous avons comparé l'algorithme MCVEM à différents algorithmes basés sur différentes approximations de l'EM:

(i) algorithme MC2-EM : algorithme dans lequel toutes les étapes d'intégration de l'EM sont approchées par une méthode de Monte Carlo. En appliquant les résultats de [18], on peut montrer que le MC2-EM converge vers le même ensemble limite que l'EM (avec probabilité 1). Bien que d'un coût de calcul assez élevé, cet algorithme permet donc de situer les autres algorithmes considérés par rapport à l'EM.

(ii) algorithme *Mean Field* proposé par [40]. Il s'agit d'un algorithme déterministe basé sur une relaxation de type variationnelle.

(iii) algorithme *Simulated Field* proposé par [40]. Il s'agit d'un algorithme qui combine méthodes variationnelles et méthodes de Monte Carlo; à la différence du MCVEM, l'étape 2c est basée sur un seul tirage.

(iv) l'algorithme *Gibbsian EM* proposé par [41]: algorithme qui combine approximations de type pseudo-vraisemblance, et méthodes de Monte Carlo.

Nous avons aussi considéré le résultat obtenu en supposant que les variables cachées étaient indépendantes: dans ce cas, la segmentation d'images s'obtient par un algorithme EM (pour les modèles de mélange). Dans les résultats numériques, cet algorithme est désigné par "ind-EM".

Le comportement asymptotique des algorithmes *Mean Field* et *Simulated Field* est un problème ouvert; nous avons constaté en particulier que l'algorithme *Mean Field* était très sensible à l'initialisation et que les trajectoires de l'algorithme *Simulated Field* ne semblaient pas converger mais ressemblaient à la réalisation d'un processus stochastique (comportement à rapprocher de celui de l'algorithme *Stochastic EM* proposé par [39] et étudié par [94]).

L'étude comparative porte à la fois sur l'estimation des paramètres et sur la segmentation. La segmentation est faite au sens du maximum des lois a posteriori marginales (MPM).

Nous présentons Figures 4.1 et 4.2 la comparaison de ces algorithmes appliqués à la segmentation d'images synthétiques, 128×128 resp. en 2 et 4 classes. Les résultats d'estimation des paramètres sont donnés dans les Tableaux 4.1 et 4.2. Figure 4.3 est une application à la segmentation d'images satellites (image SPOT 256×256).

	β	μ_1	μ_2	σ_1	σ_2	taux d'erreur
vraies valeurs	-	51	255	130	300	-
ind-EM	-	52	255	128	304	22.69
Mean Field	4.22	53	260	130	306	3.10
Simulated Field	2.15	52	250	128	302	3.42
MCVEM	0.93	50	262	125	305	2.89
MC2-EM	0.91	50	261	125	305	2.89
Gibbsian-EM	1.82	52	251	128	303	2.92

Tableau 4.1: Image Logo: Estimation des paramètres et taux d'erreur de segmentation.

De façon peu surprenante, les résultats de ind-EM sont assez mauvais justifiant l'intérêt de considérer des modèles plus complexes tels que les modèles de champ de Markov cachés.

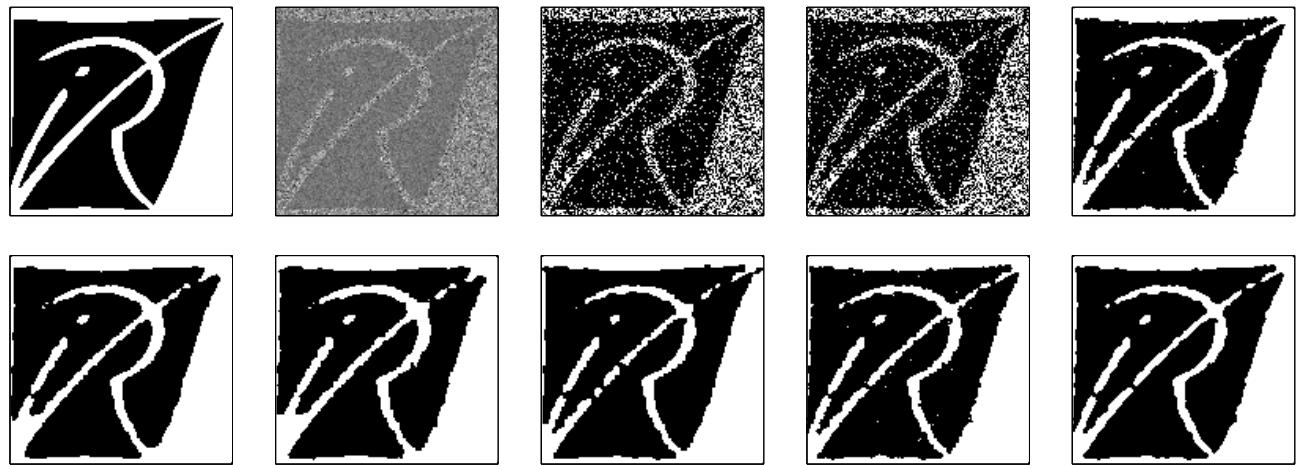


Fig. 4.1: Image Logo : [haut, de gauche à droite] image d'origine, image à segmenter, segmentation initiale obtenue par un algorithme *k-means*, ind-EM, MC2-EM; [bas, gauche à droite] Gibbsian-EM, Simulated Field, Mean Field, MCVEM, MCVEM + Filtre médian

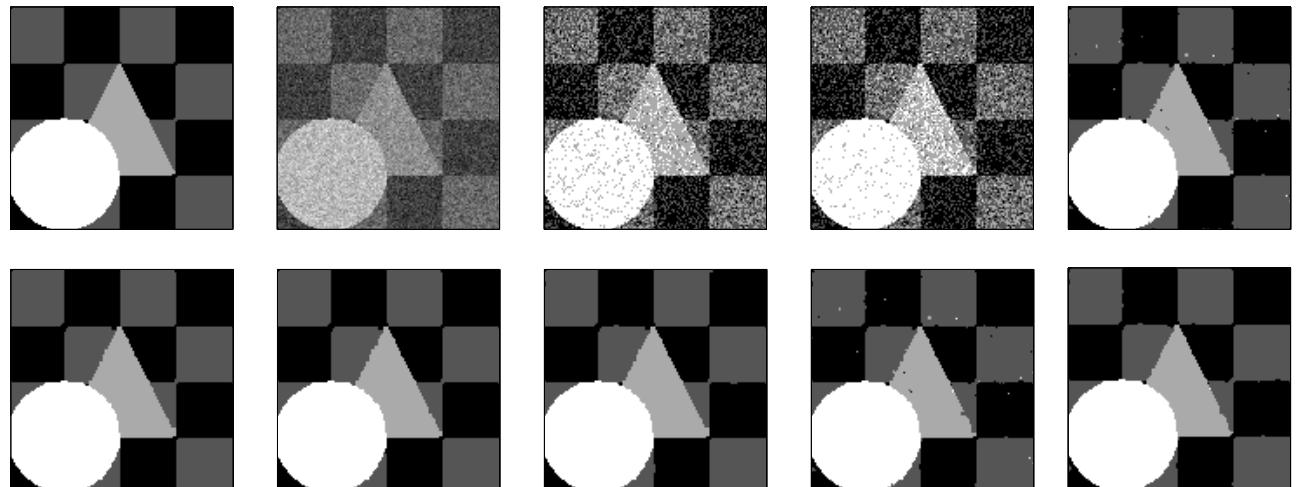
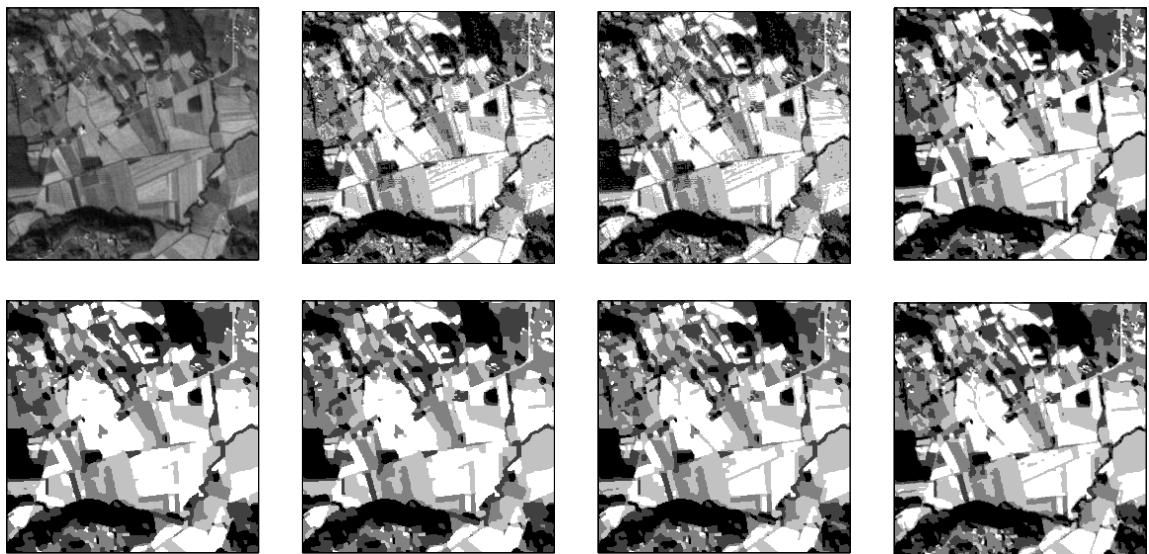


Fig. 4.2: Image Triangle : [haut, de gauche à droite] image d'origine, image à segmenter, segmentation initiale obtenue par un algorithme *k-means*, ind-EM, MC2-EM; [bas, gauche à droite] Gibbsian-EM, Simulated Field, Mean Field, MCVEM, MCVEM + Filtre médian

	β	μ_1	μ_2	μ_3	μ_4	σ_1	σ_2	σ_3	σ_4	taux d'erreur
vraies valeurs	-	1	2	3	4	0.5	0.5	0.5	0.5	-
ind-EM	-	0.85	1.69	2.54	3.93	0.44	0.42	0.46	0.53	29.4
Mean Field	3.99	0.99	2.00	2.98	4.01	0.49	0.50	0.48	0.50	0.44
Simulated Field	3.46	1.00	2.00	2.97	4.01	0.49	0.50	0.48	0.50	0.40
MCVEM	1.27	0.99	2.00	2.98	4.01	0.48	0.48	0.46	0.50	0.80
MC2-EM	1.26	0.99	2.00	2.97	4.01	0.48	0.48	0.46	0.49	0.81
Gibbsian-EM	3.01	1.00	2.00	2.98	4.00	0.49	0.50	0.48	0.50	0.31

Tableau 4.2: Image Triangle: Estimation des paramètres et taux d'erreur de segmentation.

Fig. 4.3: Image SPOT : [haut, de gauche à droite] image à segmenter, segmentation initiale obtenue par un algorithme k -means, ind-EM, MC2-EM; [bas, gauche à droite] Gibbsian-EM, Simulated Field, Mean Field, MCVEM

MCVEM donne de très bons résultats de segmentation. Les estimations du paramètre spatial β sont les plus basses ce qui favorise des pixels isolés: cela entraîne que dans les zones homogènes, certains pixels sont mal estimés. La segmentation par MCVEM peut néanmoins être améliorée par ajout d'un opérateur morphologique; par exemple, dans la segmentation du logo (resp. du triangle), le taux d'erreur peut être ramené à 2.73% (resp. 0.63%). En revanche, MCVEM préserve les structures fines (cf.figure 4.3) et en particulier, mieux que ne le font *Simulated Field* et *Mean Field* qui semblent sur-estimer le paramètre spatial β .

Enfin, MCVEM et MC2-EM ont un comportement comparable en termes d'estimation des paramètres et de segmentation. MCVEM a l'avantage d'être beaucoup moins coûteux en simulations (et donc temps de calcul). De plus, l'étape de segmentation à convergence de MCVEM est plus simple que celle à convergence de MC2-EM, puisqu'à chaque itération de MCVEM, on dispose d'une approximation de la loi a posteriori par une loi produit.

Conclusion

L'objectif principal de ce travail n'était pas de proposer un nouvel algorithme performant pour la segmentation dans des modèles de Markov caché; on peut par exemple objecter que l'algorithme MCVEM est plus coûteux en temps de calcul que d'autres méthodes, pour des améliorations de performance qui ne sont pas si significatives.

C'est surtout sur l'approche méthodologique que nous voulions insister: utiliser les méthodes de simulation pour approcher des algorithmes autrement non-implémentables dans le contexte considéré, de façon à relire ces algorithmes stochastiques comme des perturbations (contrôlables) d'un algorithme déterministe dont le comportement - éventuellement limite - est connu. L'intérêt de cette approche est qu'il devient alors possible de prouver la convergence et caractériser les points limites de l'algorithme stochastique.

4.2 Estimation de paramètres cosmologiques

Les récentes avancées spectaculaires en cosmologie observationnelle donnent accès à des données de plus grande qualité permettant de tester des modèles cosmologiques plus complexes. Ces études reposent sur une approche bayésienne couplée avec des méthodes de simulation MCMC pour l'exploration de la densité a posteriori; les échantillonneurs classiquement utilisés étant les échantillonneurs de Gibbs, de Hastings-Metropolis et l'échantillonner Hamiltonien [55].

Néanmoins, dans ce domaine d'application, l'approche MCMC présente quelques faiblesses. La première est liée au fait que la loi cible est telle que les chaînes de Markov modélisant ces échantillonneurs élémentaires présente une structure de corrélation très forte. Cela entraîne une convergence très lente des échantillonneurs et nuit en particulier à la précision de l'estimation. La seconde est liée à la connaissance que l'on a de la loi cible: on ne dispose pas d'expression de la loi a posteriori, on ne peut que l'évaluer mais au prix d'un temps de calcul relativement long. Par exemple, une évaluation de la loi a posteriori pour le jeu de données WMAP5 ([67], observations pour l'étude des anisotropies du *Cosmic Microwave Background* (CMB)) prend au moins quelques secondes. Ce coût d'évaluation de la loi a posteriori, d'autant plus problématique que l'espace de simulation est grand, oblige à restreindre les modèles que l'on peut étudier en un temps raisonnable.

Lorsque l'on dispose d'une grande force de calcul, et notamment de la possibilité de paralléliser les tâches, on peut essayer de réduire le temps d'exploration de la loi cible π , soit en diminuant le temps d'une évaluation de π en répartissant le calcul de la vraisemblance sur les différents calculateurs; soit en implémentant plusieurs MCMC en parallèle et en recombinant ces différentes trajectoires. La première solution n'est pas toujours faisable ou significativement intéressante. Quant à la seconde, elle nécessite de contrôler la convergence de chaque chaîne et/ou maîtriser les techniques d'interactions entre chaînes parallèles; à notre connaissance, il n'y a pas de résultats vraiment probants sur la parallélisation des MCMC, qui pourraient s'appliquer à ces données de cosmologie.

L'objectif de ce travail, mené en partenariat avec l'Institut d'Astrophysique de Paris (IAP) dans le cadre d'un projet ANR, était de proposer une méthode de simulation réduisant de façon très significative le temps d'exploration de la densité cible en sachant tirer profit de la force de calcul disponible à l'IAP; et ce, afin d'estimer différents paramètres cosmologiques à partir de trois jeux de

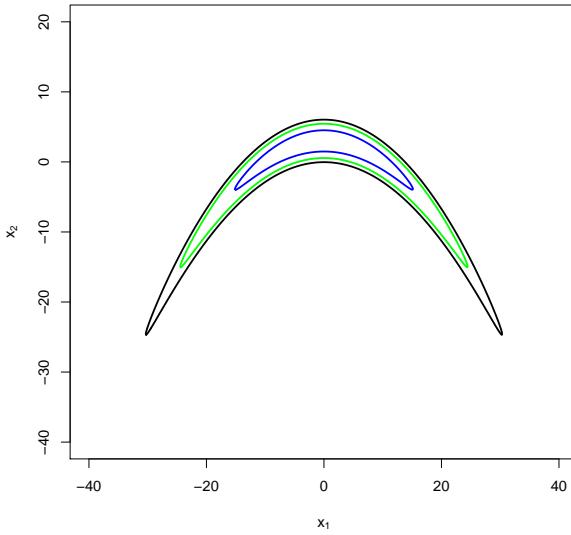


Fig. 4.4: Projection de la densité cible dans le plan (x_1, x_2) plane. Les contours représentent les régions de plus haute densité a posteriori à 68.3% (bleu), 95% (vert) et 99% (noir)

données (WMAP5, SNLS et T0003, resp. issus de l’observation des anisotropies du CMB [67], des supernovae [72] et du lentillage gravitationel faible [58]).

Les méthodes d’échantillonnage d’importance sont basées sur la simulation de variables indépendantes (cf. section 1.1.2) et sont donc des méthodes de simulation parallélisables. Néanmoins, leur efficacité dépend de la loi de proposition, et ce problème est d’autant plus crucial que l’espace de simulation est grand. Nous avons donc proposé d’utiliser une méthode d’échantillonnage d’importance adaptative, et plus précisément, l’algorithme M-PMC (cf. section 1.1.2).

Dans un premier temps, nous avons comparé les performances de M-PMC à celles d’un algorithme de Hastings-Metropolis à marche aléatoire symétrique sur données simulées. Nous avons ensuite appliqué l’algorithme M-PMC à l’estimation bayésienne de paramètres cosmologiques. Nous résumons ci-après ces études; on en trouvera une description détaillée dans [23].

4.2.1 M-PMC sur données simulées

Nous avons implémenté l’algorithme M-PMC lorsque la loi cible π est obtenue par une distorsion d’une densité gaussienne p -dimensionnelle : $(X_1, \dots, X_p) \sim \pi$ tels que

$$(X_1, X_2, +b(X_1^2 - \sigma_1^2), X_3, \dots, X_p) \sim \mathcal{N}_p(0, \Sigma)$$

où $\Sigma = \text{diag}(\sigma_1^2, 1, \dots, 1)$. Dans les applications numériques $p = 10$, $\sigma_1^2 = 100$ et $b = 0.03$. Les lignes de niveau de la loi marginale de (X_1, X_2) sont représentées figure 4.4 (les v.a. X_3, \dots, X_p sont indépendantes de (X_1, X_2) et de loi gaussienne centrée réduite).

On cherche la loi instrumentale q dans la famille des mélanges de 9 densités de Student, à 9 degrés de liberté; la loi est mise à jour tous les 10 k tirages. La figure 4.5 montre l’évolution de M-PMC au

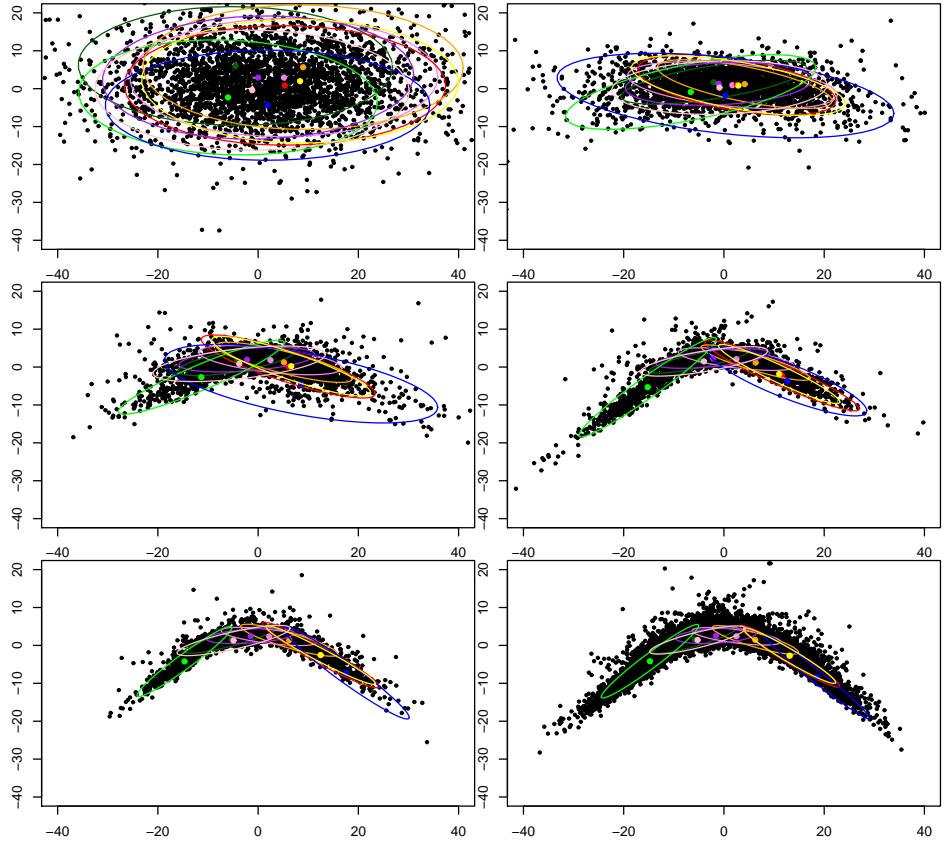


Fig. 4.5: [de gauche à droite et haut en bas] Evolution de la loi instrumentale q (itérations 1, 3, 5, 7, 9 et 11) et du nuage de points (1 point sur 3 est tracé). Les couleurs indiquent les différentes composantes du mélange q : les centres sont indiqués par des points et les ellipses délimitent les régions de plus haute densité a posteriori à 95%.

fur et à mesure des mises à jour : on représente les $10 k$ points, et les centres et ellipses de plus haute densité a posteriori à 95% des 9 composantes du mélange estimées à chaque itération.

La distance entre la loi cible π et la loi instrumentale courante $q^{(t)}$, mesurée au sens de la divergence de Kullback, peut être approchée par la “perplexité normalisée”¹ - nombre compris entre 0 et 1 et d'autant plus proche de 1 que la distance est faible - On trace sur la figure 4.6[haut] l'évolution de la perplexité en fonction du nombre de mise à jour de la loi instrumentale.

Les algorithmes d'échantillonnage construisent un nuage de points pondérés comme approximation de π , les poids étant proportionnels au ratio d'importance π/q . La taille effective de l'échantillon (*effective sample size (ESS)*) est un indicateur de non-dégénérescence des poids : c'est un nombre compris entre 0 et N , N étant le nombre de points du nuage, et d'autant plus proche de N que les poids sont uniformes. On trace sur la figure 4.6[bas] l'évolution de l'ESS normalisée (ESS/N) en

¹ estimation de $\exp(-DK(\pi, q^{(t)}))$ - DK désigne la distance de Kullback - à partir des poids $\{\omega_k^t, k \in \{1, \dots, N\}\}$ du nuage de points $X_k^{(t)}$, v.a. i.i.d. de loi $q^{(t)}$

$$\frac{1}{N} \exp \left[- \left(\sum_{k=1}^N \bar{\omega}_k^t \ln(\bar{\omega}_k^t) \right) \right] \quad \bar{\omega}_k^t := \frac{\omega_k^t}{\sum_{j=1}^N \omega_j^t} \quad \omega_k^t := \frac{\pi(X_k^{(t)})}{q^{(t)}(X_k^{(t)})}$$

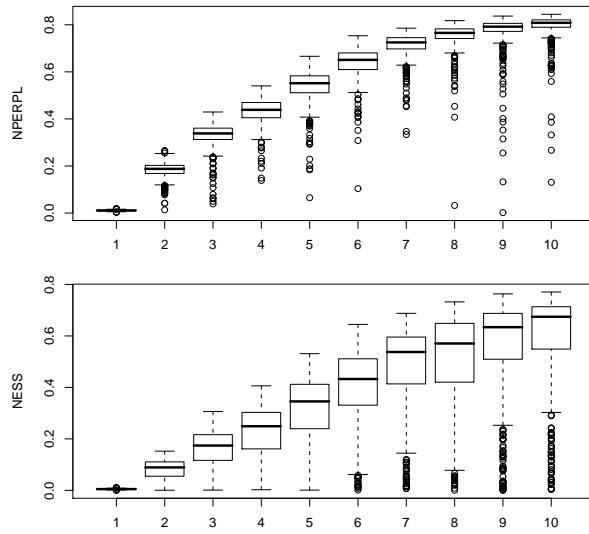


Fig. 4.6: Distribution de la perplexité normalisée [haut] et de l'ESS [bas] en fonction du nombre de mise à jour de la loi instrumentale. Ces distributions sont calculées à partir de 500 répétitions de l'algorithme M-PMC. Ces boîtes représentent : la médiane de la distribution (trait noir épais horizontal); les quantiles 0.25 et 0.75 (rectangles en traits fins); et les points aberrants (cercles).

fonction du nombre de mise à jour de la loi instrumentale.

Ces figures montrent qu'après quelques itérations de M-PMC, la loi de proposition ressemble à la loi cible (cf. figure 4.5) et qu'au bout d'une dizaine d'itérations, la perplexité et l'ESS n'augmentent plus significativement. On peut donc cesser d'adapter la loi de proposition.

Nous avons observé que M-PMC était sensible à l'initialisation: une mauvaise initialisation peut ralentir la convergence du processus d'adaptation et dégrader l'estimation de q (qui, à la différence de ce que l'on observe sur la figure 4.5, ne va plus couvrir tout l'espace où la densité π est grande).

4.2.2 Comparaison de M-PMC et de l'échantillonneur *Adaptive Metropolis* sur données simulées

Nous avons comparé les algorithmes M-PMC et *Adaptive Metropolis* (cf. section 1.2.1 pour l'estimation de l'espérance des fonctions

$$\begin{aligned} f_a(x) &= x_1 & f_b(x) &= x_2 \\ f_c(x) &= \mathbb{I}_{68.3}(x) & f_d(x) &= \mathbb{I}_{95}(x) \\ f_e(x) &= \mathbb{I}_{68.3}(x_1, x_2) & f_g(x) &= \mathbb{I}_{95}(x_1, x_2) \\ f_h(x) &= \mathbb{I}_{68.3}(x_1) & f_i(x) &= \mathbb{I}_{95}(x_1); \end{aligned}$$

en notant \mathbb{I}_q l'indicatrice de la région de plus haute densité a posteriori de niveau $q\%$, sous la loi π décrite en section 4.2.1. Dans M-PMC comme dans *Adaptive Metropolis*, la mise à jour des paramètres d'implémentation est faite tous les 10 k points.

Le tableau 4.3 et les figures 4.7 et 4.8 présentent les résultats obtenus: estimation des moyennes et de la précision de l'estimation calculée à partir de 500 répétitions des algorithmes.

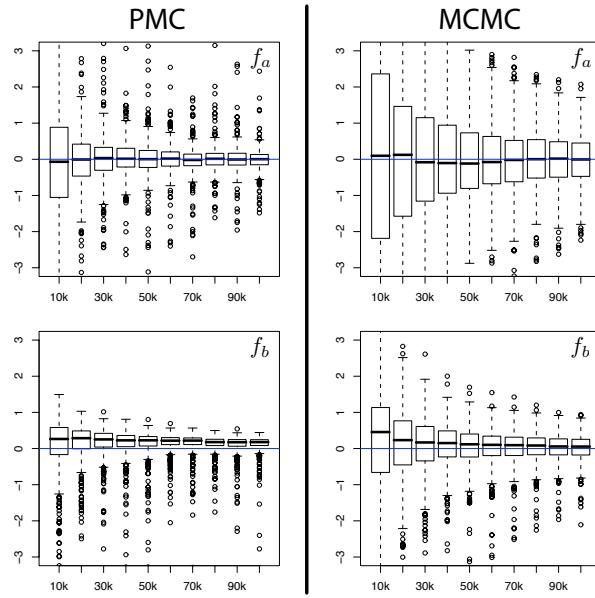


Fig. 4.7: Evolution de la distribution empirique des estimateurs Monte Carlo des quantités $\pi(f_a)$ [haut] et $\pi(f_b)$ [bas] en fonction du nombre de mises à jour, pour les échantillonneurs M-PMC [gauche] et *Adaptive Metropolis* [droite]. En bleu, la vraie valeur de $\pi(f_a)$ et $\pi(f_b)$. Ces boîtes sont calculées à partir de 500 réalisations des estimateurs.

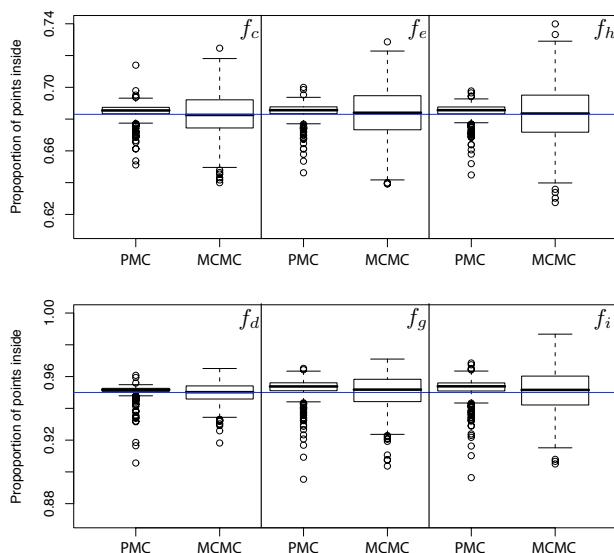


Fig. 4.8: Distribution empirique des estimateurs Monte Carlo des quantités $\pi(f_v)$ pour les échantillonneurs M-PMC [gauche] et *Adaptive Metropolis* [droite]. En bleu, la vraie valeur de $\pi(f_v)$. Ces boîtes sont calculées à partir de 500 réalisations des estimateurs.

Ces résultats montrent que la variabilité des estimations par l'échantillonneur M-PMC est plus faible que celle des estimations par l'algorithme *Adaptive Metropolis*. En particulier, sur la figure 4.7, la variabilité obtenue après $20 k$ points avec M-PMC est du même ordre que celle obtenue avec $100 k$ points avec l'algorithme *Adaptive Metropolis*.

En revanche, nous avons observé que l'algorithme *Adaptive Metropolis* est bien plus robuste à une mauvaise initialisation de la loi de proposition que ne l'est M-PMC. Cette sensibilité de M-PMC à l'initialisation est d'autant plus forte que la dimension de l'espace de simulation est grande.

4.2.3 M-PMC pour l'estimation de paramètres cosmologiques

Nous avons comparé M-PMC et un algorithme de Hastings-Metropolis à marche aléatoire symétrique (SRWM) pour l'exploration de la loi a posteriori des paramètres cosmologiques, et pour différents jeux de données.

Pour la plupart des paramètres, le prior est plat; ou uniforme sur un hypercube (pour éviter d'explorer des zones de l'espace des paramètres qui n'ont pas de sens physique, et qui en conséquence sont en dehors des valeurs des paramètres pour lesquelles on sait évaluer la vraisemblance).

On cherche à adapter la loi instrumentale q parmi la famille des mélanges gaussiens; nous avons considéré des mélanges de 5 à 10 densités, et la mise à jour de cette loi se fait tous les $10 k$ tirages. L'algorithme M-PMC a été initialisé en estimant (par une méthode de gradient conjugué) le maximum de la vraisemblance et la matrice de Fisher en ce point; les paramètres des composantes gaussiennes utilisées dans la loi initiale $q^{(0)}$ sont obtenus en perturbant ces estimations du mode et de la dispersion.

La figure 4.9 montre le comportement de M-PMC pour la vraisemblance WMAP5: on trace l'évolution de la perplexité normalisée et de l'ESS normalisé en fonction du nombre de mises à jour, ainsi que l'évolution de la distribution des poids des points du nuage. On atteint une perplexité assez élevée (0.95) et un ESS normalisé de 0.7; et l'on observe que si au début de l'algorithme, beaucoup de points ont des poids très faibles, les particules migrent ensuite vers des zones de l'espace de simulation de plus forte probabilité.

Sur la figure 4.10, on montre l'allure de deux densités marginales (en dimension 2) de la loi a posteriori; ainsi que l'évolution des moyennes des composantes du mélange (projection de ces moyennes en dimension 2). On observe en particulier la forte structure de corrélation entre les paramètres, ce qui est certainement à l'origine de la lente convergence des algorithmes MCMC.

		PMC	MCMC
$\pi(f_a)$	mean	0.097	-0.028
"	std	0.218	0.536
$\pi(f_b)$	mean	0.013	0.002
"	std	0.163	0.315
acceptance		-	0.11
perplexity		0.80	-

Tableau 4.3: Estimation des moyennes et de la précision. Résultats obtenus à partir de 500 répétitions de chaque algorithme.

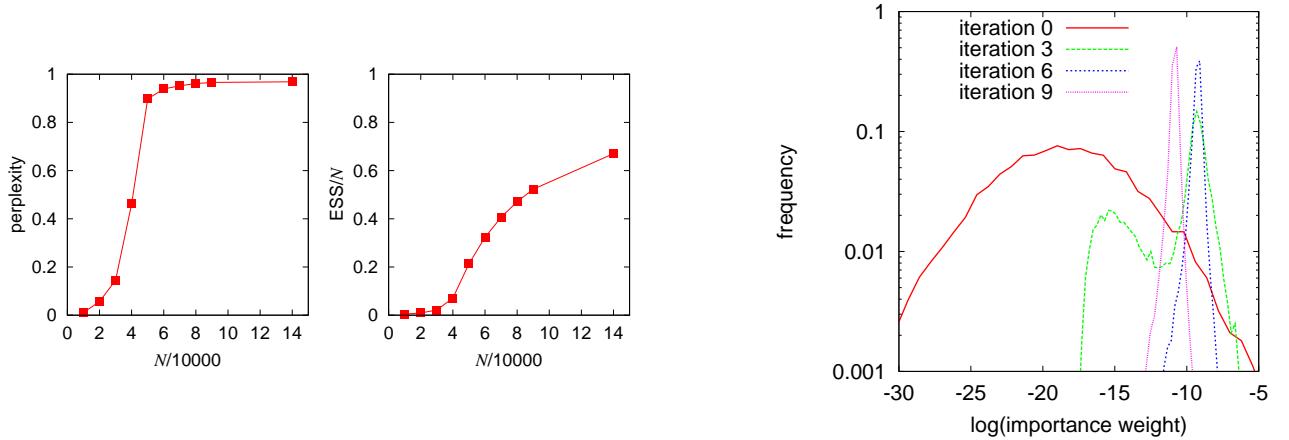


Fig. 4.9: Perplexité [gauche] et ESS [centre] en fonction du nombre de mise à jour de la loi instrumentale (tous les 10 k tirages). [droite] Evolution de la distribution des poids des 10 k points, au fur et à mesure des mises à jour de la loi instrumentale.

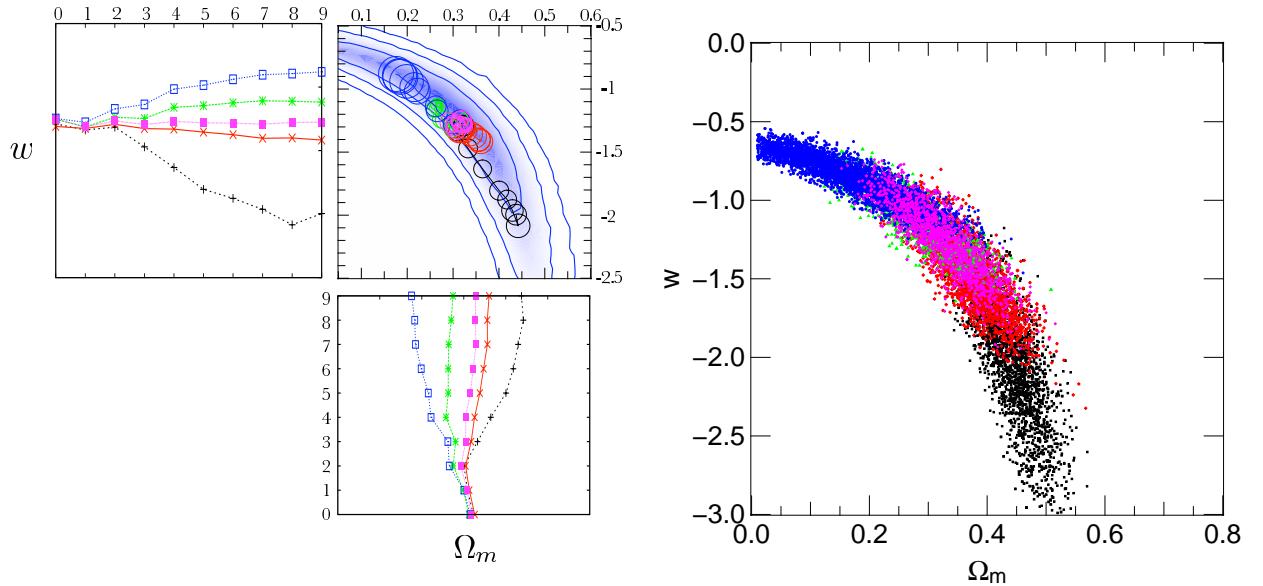


Fig. 4.10: Jeu de données SNIa. [droite] Nuage de points approchant la loi a posteriori de (ω, Ω_m) (1 point sur 5 est représenté). [haut gauche et bas] Evolution de la moyenne de chacune des 5 composantes gaussiennes de la loi instrumentale. [haut centre] Loi a posteriori des paramètres (ω, Ω_m) : on représente les régions de plus haute densité a posteriori à 68%, 95% et 99.7%. On montre aussi l'évolution de la loi instrumentale: les cercles sur la figure centrale sont centrés en les moyennes des composantes du mélange, et leur rayon est proportionnel au poids de la composante dans le mélange.

Les tableaux 4.4 et 4.5, et la figure 4.11 présentent les résultats de la comparaison de M-PMC et de SRWM. Ces deux algorithmes donnent des résultats équivalents pour l'estimation des paramètres par l'estimateur de moyenne a posteriori - cf. tableaux (4.4 et 4.5). Sur la figure 4.11, on trace les régions de plus haute densité a posteriori à 68%, 95% et 99.7% de lois marginales a posteriori (de

Paramètre	PMC	MCMC
Ω_b	$0.04424^{+0.00321}_{-0.00290}$	$0.04418^{+0.00321}_{-0.00294}$
Ω_m	$0.2633^{+0.0340}_{-0.0282}$	$0.2626^{+0.0359}_{-0.0280}$
τ	$0.0878^{+0.0181}_{-0.0160}$	$0.0885^{+0.0181}_{-0.0160}$
n_s	$0.9622^{+0.0145}_{-0.0143}$	$0.9628^{+0.0139}_{-0.0145}$
$10^9 \Delta_R^2$	$2.431^{+0.118}_{-0.113}$	$2.429^{+0.123}_{-0.108}$
h	$0.7116^{+0.0271}_{-0.0261}$	$0.7125^{+0.0274}_{-0.0268}$

Tableau 4.4: Jeu de données WMAP5. Estimation des paramètres et intervalles de confiance à 68%

dimension 2) pour le jeu de données SNIa: pour les marginales des paramètres $(\alpha, -M)$, ces courbes sont quasiment elliptiques sauf celle à 99.7% obtenue par l'échantillonneur SRWM; cela s'explique par une longue excursion de la chaîne dans les queues de distribution.

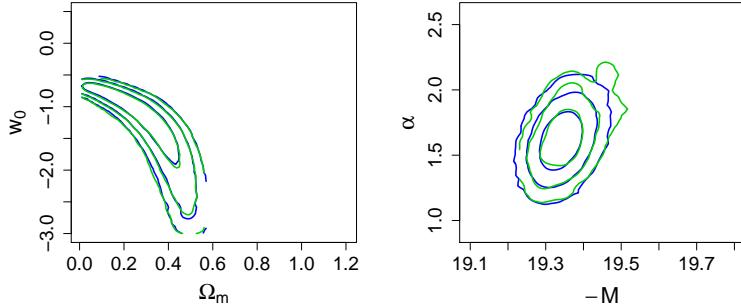


Fig. 4.11: Jeu de données SNIa. Exemples de lois (marginales) a posteriori obtenues par M-PMC (bleu) et SRWM (vert). On représente les contours estimés des zones de plus haute densité a posteriori à 68%, 95% et 99.7%.

Si les estimations sont peu sensibles à la méthode d'échantillonnage pour l'exploration de la loi a posteriori, le temps de calcul est lui très nettement supérieur pour SRWM; à nombre d'évaluations égal de la vraisemblance, M-PMC permet une réduction du temps de calcul d'un facteur 100.

Conclusion

Nous avons montré que, dans ce problème d'estimation bayésienne de paramètres cosmologiques, l'algorithme M-PMC a un comportement équivalent à celui d'autres échantillonneurs MCMC plus classiques; comportement équivalent en termes d'exploration de la loi a posteriori et par suite, d'estimation des paramètres.

Paramètre	PMC	MCMC
Ω_b	$0.0432^{+0.0027}_{-0.0024}$	$0.0432^{+0.0026}_{-0.0023}$
Ω_m	$0.254^{+0.018}_{-0.017}$	$0.253^{+0.018}_{-0.016}$
τ	$0.088^{+0.018}_{-0.016}$	$0.088^{+0.019}_{-0.015}$
w	-1.011 ± 0.060	$-1.010^{+0.059}_{-0.060}$
n_s	$0.963^{+0.015}_{-0.014}$	$0.963^{+0.015}_{-0.014}$
$10^9 \Delta_R^2$	$2.413^{+0.098}_{-0.093}$	$2.414^{+0.098}_{-0.092}$
h	$0.720^{+0.022}_{-0.021}$	$0.720^{+0.023}_{-0.021}$
a	$0.648^{+0.040}_{-0.041}$	$0.649^{+0.043}_{-0.042}$
b	$9.3^{+1.4}_{-0.9}$	$9.3^{+1.7}_{-0.9}$
c	$0.639^{+0.084}_{-0.070}$	$0.639^{+0.082}_{-0.070}$
$-M$	19.331 ± 0.030	$19.332^{+0.029}_{-0.031}$
α	$1.61^{+0.15}_{-0.14}$	$1.62^{+0.16}_{-0.14}$
$-\beta$	$-1.82^{+0.17}_{-0.16}$	-1.82 ± 0.16
σ_8	$0.795^{+0.028}_{-0.030}$	$0.795^{+0.030}_{-0.027}$

Tableau 4.5: Combinaison des trois jeux de données. Estimation des paramètres et intervalles de confiance à 68%

Néanmoins, M-PMC est préférable dans le contexte de cette application, pour les raisons suivantes:

(i) possibilité de paralléliser les simulations. Par suite, le temps de calcul pour l'obtention d'un nuage de points approchant la loi a posteriori est réduit de façon drastique, et ce, d'autant plus que l'on dispose de CPU.

(ii) possibilité de diagnostiquer simplement la convergence de l'algorithme à partir du suivi de l'évolution d'indicateurs très facilement calculables, comme la perplexité et l'ESS.

Bibliographie

- [1] C. Andrieu and G. Fort. Explicit control of subgeometric ergodicity. Technical Report 05:17, University of Bristol, UK, 2005.
- [2] Y. Atchadé and G. Fort. Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. *Bernoulli*, 2009. Accepted for publication.
- [3] Y. Atchadé, G. Fort, E. Moulines, and P. Priouret. Adaptive Markov chain Monte Carlo : Theory and Methods. Technical report, 2009.
- [4] S. Connor and G. Fort. State-dependent Foster-Lyapunov criteria for subgeometric convergence of Markov chains. *Stoch. Process. Appl.*, 119:4176–4193, 2009.
- [5] R. Douc, G. Fort, and A. Guillin. Subgeometric rates of convergence of f-ergodic strong Markov processes. *Stoch. Process Appl.*, 119(3):897–923, 2009.
- [6] R. Douc, G. Fort, E. Moulines, and P. Priouret. Forgetting of the initial distribution for Hidden Markov Models. *Stoch. Process Appl.*, 119(4):1235–1256, 2009.
- [7] R. Douc, G. Fort, E. Moulines, and P. Soulier. Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.*, 14(3):1353–1377, 2004.
- [8] P. Etoré, G. Fort, B. Jourdain, and E. Moulines. On adaptive stratification. *Annals of Operations Research*, 2009. Accepted for publication.
- [9] F. Forbes and G. Fort. Combining Monte Carlo and Mean Field-like methods for inference in Hidden Markov Random Fields. *IEEE Trans. Image Process.*, 16(3):824–837, 2007.
- [10] G. Fort. *Contrôle explicite d'ergodicité de chaînes de Markov : Applications à l'analyse de convergence de l'algorithme Monte-Carlo EM*. PhD thesis, Univ. Paris VI, France, 2001.
- [11] G. Fort. Computable bounds for V-geometric ergodicity of Markov transition kernels. Technical Report RR 1047-M, Univ. J. Fourier, France, 2002.
- [12] G. Fort. Partial Least Squares for classification and feature selection in Microarray gene expression data. Technical report, UNIV. J. Fourier, France, 2004.
- [13] G. Fort. Fluid limit-based tuning of some hybrid MCMC samplers. Technical report, 2007.
- [14] G. Fort and S. Lambert-Lacroix. Classification using Partial Least Squares with Penalized Logistic Regression. *Bioinformatics*, 21(7):1104–1111, 2005.

- [15] G. Fort, S. Lambert-Lacroix, and J. Peyre. Réduction de dimension dans les modèles généralisés : Application à la classification de données issues des biopuces. *Journal de la SFDS*, 146(1-2):117–152, 2005.
- [16] G. Fort, S. Meyn, E. Moulines, and P. Priouret. The ODE method for the stability of skip-free Markov Chains with applications to MCMC. *Ann. Appl. Probab.*, 18(2):664–707, 2008.
- [17] G. Fort and E. Moulines. V-subgeometric ergodicity for a Hastings-Metropolis algorithm. *Stat. Probab. Lett.*, 49(4):401–410, 2000.
- [18] G. Fort and E. Moulines. Convergence of the Monte-Carlo EM for curved exponential families. *Ann. Stat.*, 31(4):1220–1259, 2003.
- [19] G. Fort and E. Moulines. Polynomial ergodicity of Markov transition kernels. *Stochastic Processes Appl.*, 103(1):57–99, 2003.
- [20] G. Fort, E. Moulines, G.O. Roberts, and J.S. Rosenthal. On the geometric ergodicity of hybrid samplers. *J. Appl. Probab.*, 40(1):123–146, 2003.
- [21] G. Fort, E. Moulines, and P. Soulier. *Inference in Hidden Markov Models*, chapter "Elements of Markov Chain Theory" Chapter 14), pages 511–562. Springer, 2005.
- [22] G. Fort and G.O. Roberts. Subgeometric ergodicity of strong Markov processes. *Ann. Appl. Probab.*, 15(2):1565–1589, 2005.
- [23] D. Wraith, M. Kilbinger, K. Benabed, O. Cappé, J.F. Cardoso, G. Fort, S. Prunet, and C.P. Robert. Estimation of cosmological parameters using adaptive importance sampling. *Phys. Rev. D*, 80(2):18 pages, 2009.
- [24] S.J. Ahamed, V.S. Borkar, and S. Juneja. Adaptive Importance Sampling Technique for Markov Chains using Stochastic Approximation. *Oper. Res.*, 54(3):489–504, 2004.
- [25] C. Andrieu, A. Jasra, A. Doucet, and P. Del Moral. On non-linear Markov chain Monte Carlo via self-interacting approximations. Technical report, Available at http://stats.ma.ic.ac.uk/a/aj2/public_html/, 2007.
- [26] C. Andrieu and E. Moulines. On the ergodicity properties of some adaptive MCMC algorithms. *Ann. Appl. Probab.*, 16(3):1462–1505, 2006.
- [27] C. Andrieu and J. Thoms. A tutorial on adaptive MCMC. *Statistics and Computing*, 18(4):343–373, 2008.
- [28] B. Arouna. Adaptative Monte Carlo method, a variance reduction technique. *Monte Carlo Methods Appl.*, 10(1):1–24, 2004.

- [29] S. Asmussen and P.W. Glynn. *Stochastic simulation*. Stochastic Modelling and Applied Probability, Springer, 2007.
- [30] Y. Atchade, G.O. Roberts, and J.S Rosenthal. Optimal scaling of Metropolis-Coupled Markov chain Monte Carlo. Technical report, 2009.
- [31] Y. F. Atchade and J. S. Rosenthal. On adaptive Markov chain Monte Carlo algorithm. *Bernoulli*, 11(5):815–828, 2005.
- [32] P. Baxendale. Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.*, 15:700–738, 2005.
- [33] P.H. Baxendale. Uniform estimates for geometric ergodicity of recurrent Markov processes. University of Southern California, 1993.
- [34] L. Breyer and G. O. Roberts. From Metropolis to diffusions : Gibbs states and optimal scaling. *Stochastic Proc. Appl.*, 90:181–206, 2000.
- [35] W. Byrne and A. Gunawardana. Convergence theorems of Generalized Alternating Minimization Procedures. *Journal of Machine Learning Research*, 1:1–48, 2004.
- [36] M. Bédard. Weak convergence of Metropolis algorithms for non-IID target distributions. *Ann. Appl. Probab.*, 17:1222–1244, 2007.
- [37] O. Cappé, R. Douc, A. Guillin, J.M. Marin, and C.P. Robert. Adaptive Importance Sampling in General Mixture Classes. *Stat. Comput.*, 18:447–459, 2008.
- [38] O. Cappé, A. Guillin, J.M. Marin, and C.P. Robert. Population Monte Carlo. *J. Comput. Graph. Statist.*, 13:907–929, 2005.
- [39] G. Celeux and J. Diebolt. The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Comput. Statist. Quart.*, 2(1):73–82, 1985.
- [40] G. Celeux, F. Forbes, and N. Peyrard. EM procedures using mean-field approximations for Markov model-based image segmentation. *Pattern Recognition*, 36:131–144, 2003.
- [41] B. Chalmond. An iterative gibbsian technique for reconstruction of m-array images. *Pattern Recognition*, 22(6):747–761, 1989.
- [42] D. Chandler. *Introduction to Modern Statistical Mechanics*. Oxford University Press, 1987.
- [43] D. Chauveau and P. Vandekerkhove. Un algorithme de Hastings-Metropolis avec apprentissage séquentiel. *C. R. Acad. Sci. Paris Sér. I Math.*, 329(2):173–176, 1999.
- [44] D. Chauveau and P. Vandekerkhove. Algorithmes de Hastings-Metropolis en interaction. *C. R. Acad. Sci. Paris Sér. I Math.*, 333(9):881–884, 2001.
- [45] H. Chen and D. D. Yao. *Fundamentals of queueing networks*, volume 46 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 2001. Performance, asymptotics, and optimization, Stochastic Modelling and Applied Probability.

- [46] H. Chen and Y.M. Zhu. Stochastic approximation procedures with randomly varying truncations. *Sci. Sinica. Ser. A*, 29:914–926, 1986.
- [47] H.F. Chen, L. Guo, and A.J. Gao. Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds. *Stochastic Process. Appl.*, 27:217–231, 1988.
- [48] O.F. Christensen, G.O. Roberts, and J.S. Rosenthal. Scaling limits for the Transient Phase of Local Metropolis-Hastings Algorithms. *J. Roy. Statist. Soc. Ser. B*, 67:253–268, 2005.
- [49] I. Csiszar and G. Tusnady. Information geometry and alternating minimization procedures. *Stat. & Dec.*, 1:205–237, 1984. Supp. Iss.
- [50] J.G. Dai and S.P. Meyn. Stability and convergence of moments for multiclass queueing networks via fluid limit models. *IEEE Trans. Automat. Control*, 40(11):1889–1904, 1995.
- [51] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. B*, 39(1):1–38, 44.
- [52] R. Douc, E. Moulines, and J. Rosenthal. Quantitative bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.*, 14(4):1643–1665, 2004.
- [53] R. Douc, E. Moulines, and P. Soulier. Computable bounds for subgeometric ergodic Markov chains. *Bernoulli*, 13(3):831–848, 2007.
- [54] N. Down, S.P. Meyn, and R.L. Tweedie. Exponential and uniform ergodicity of Markov processes. *Ann. Prob.*, 23(4):1671–1691, 1995.
- [55] S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Phys. Lett. B*, 195:216–22, 1987.
- [56] P. Etore and B. Jourdain. Adaptive optimal allocation in stratified sampling methods. *Methodol. Comput. Appl. Probab.*, 9(2):117–152, to appear (2009).
- [57] J. Fessler. Comments on "the convergence of mean field procedures for MRF's". *IEEE Trans. Image Process.*, 7(6):917, 1998.
- [58] L. Fu, E. Sambolini, H. Hoekstra, M. Kilbinger, L. van Waerbeke, I. Tereno, Y. Mellier, and et al. Very weak lensing in the CFHTLS Wide: Cosmology from cosmic shear in the linear regime. *Astronomy and Astrophysics*, 479(1):9–25, 2008.
- [59] H. Ganidis, B. Roynette, and F. Simonot. Convergence rate of semi-groups to their invariant probability. *Stochastic Process. Appl.*, 79:243–263, 1999.
- [60] C. J. Geyer. Markov chain Monte Carlo maximum likelihood. *Computing Science and Statistics: Proc. 23rd Symposium on the Interface, Interface Foundation, Fairfax Station, VA*, pages 156–163, 1991.
- [61] P. Glasserman. *Monte Carlo methods in Financial Engineering*. Springer, 2004.

- [62] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Math. Finance*, 9(2):117–152, 1999.
- [63] U. Grenander, editor. *Unitary dilations of Markov transition operators and the corresponding integral representation for transition-probability matrices*, pages 139–161. Almqvist and Wiksell, Stockholm, 1959.
- [64] H. Haario, E. Saksman, and J. Tamminen. Adaptive proposal distribution for random walk Metropolis algorithm. *Comput. Stat.*, 14(3):375–395, 1999.
- [65] H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001.
- [66] R.Z. Has'minskii. *Stochastic stability of differential equations*. Sijthoff and Noordhoff, 1980.
- [67] G. Hinshaw, J.L. Weiland, R.S. Hill, N. Odegard, D. Larson, C.L. Bennett, J. Dunkley, B. Gold, M.R. Greason, N. Jarosik, and et al. Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Data Processing, Sky Maps, and Basic Results. *Astrophys. J. Suppl.*, 180:246–264, 2009.
- [68] S. Jarner and E. Hansen. Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.*, 85:341–361, 2000.
- [69] S. Jarner and G.O. Roberts. Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.*, 12:224–247, 2002.
- [70] S. F. Jarner and G.O. Roberts. Convergence of Heavy-tailed Monte Carlo Markov Chain Algorithms. *Scand. J. Stat.*, 34:781–815, 2007.
- [71] M.I. Jordan, Z. Ghahramani, T.S. Jaakkola, and L.K. Saul. An introduction to variational methods for graphical models. In M.I. Jordan, editor, *Learning in Graphical Models*, pages 105–162. 1998.
- [72] M. Kilbinger, K. Benabed, J. Guy, P. Astier, I. Tereno, L. Fu, D. Wraith, J. Coupon, Y. Mellier, C. Balland, and et al. Dark-energy constraints and correlations with systematics from CFHTLS weak lensing, SNLS supernovae Ia and WMAP5. *Astronomy and Astrophysics*, 497(3):677–688, 2009.
- [73] J.E. Kolassa. Explicit bounds for geometric convergence of Markov chains. *J. Appl. Probab.*, 37:642–651, 2000.
- [74] S. Kou, Q. Zhou, and W. Wong. Equi-energy sampler with applications in statistical inference and statistical mechanics. *Ann. Stat.*, 34:1581–1619, 2006.
- [75] M. Krasnosel'skii and Y. Rutickii. *Convex functions and Orlicz spaces*. Noordhoff, Groningen, 1961.
- [76] T. Kurtz. Solutions of ordinary differential equations as limits of pure jump Markov processes. *J. Appl. Probab.*, 7(1):49–58, 1970.

- [77] R.B. Lund and R.L. Tweedie. Geometric convergence rates for stochastically ordered Markov Chains. *Math. Oper. Res.*, 20(1):182–194, 1996.
- [78] E. Marinari and G. Parisi. Simulated tempering: A new Monte Carlo schemes. *Europhysics letters*, 19:451–458, 1992.
- [79] M. Maxwell and M. Woodroffe. Central limit theorems for additive functional of Markov chains. *Ann. Probab.*, 28(2):713–724, 2000.
- [80] K.L. Mengersen and R.L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.*, 24(1):101–121, 1996.
- [81] F. Merlevede, M. Peligrad, and S. Utev. Recent advances in invariance principles for stationary sequences. *Probability surveys*, 3:1–36, 2006.
- [82] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1091, 1953.
- [83] S Meyn. *Control techniques for Complex Networks*. Cambridge University Press, 2008.
- [84] S. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer, 1993.
- [85] S. Meyn and R.L. Tweedie. State dependent criteria for convergence of Markov chains. *Ann. Appl. Probab.*, pages 149–168, 1994.
- [86] S.P. Meyn and R.L. Tweedie. *Generalized resolvents and Harris Recurrence of Markov processes*, pages 227–250. American Mathematical Society, Providence, RI, 1993.
- [87] S.P. Meyn and R.L. Tweedie. Stability of Markovian processes II: continuous-time processes and sampled chains. *Adv. Appl. Prob.*, 25:487–517, 1993.
- [88] S.P. Meyn and R.L. Tweedie. Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes. *Adv. Appl. Prob.*, 25:518–548, 1993.
- [89] S.P. Meyn and R.L. Tweedie. Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.*, 4(4):981–1011, 1994.
- [90] P. Neal and G.O. Roberts. Optimal scaling for partially updating MCMC algorithms. *Ann. Appl. Probab.*, 16(2):475–515, 2006.
- [91] R.M. Neal and G.E. Hinton. A view of the EM algorithm that justifies incremental, sparse and other variante. In M.I. Jordan, editor, *Learning in Graphical Models*, pages 355–368. 1998.
- [92] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009.
- [93] G. Newell. *Applications of queueing theory. 2nd ed.* Monographs on Statistics and Applied Probability. London New York: Chapman and Hall. XVI, 303 p., 1982.

- [94] S.F. Nielsen. The stochastic EM: Estimation and asymptotic results. *Bernoulli*, 6(3):457–489, 2000.
- [95] M. Oh and J. Berger. Integration of multimodal functions by Monte Carlo importance sampling. *J. American Statist. Assoc.*, 88:450–456, 1993.
- [96] G. Parisi. *Statistical Field Theory*. Addison-Wesley, 1988.
- [97] C.P. Robert. *Méthodes de Monte Carlo par Chaînes de Markov*. Economica, 1996.
- [98] G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7(1):110–120, 1997.
- [99] G. O. Roberts and J. S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statist. Sci.*, 16(4):351–367, 2001.
- [100] G. O. Roberts and J. S. Rosenthal. Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Probab.*, 44(2):458–475, 2007.
- [101] G.O. Roberts and O. Stramer. Langevin diffusions and Metropolis-Hastings algorithms. *Method. Comput. Appl. Probab.*, 4:337–357, 2002.
- [102] G.O. Roberts and R.L. Tweedie. Exponential convergence of Langevin diffusions and their discrete approximations. *Bernoulli*, pages 341–364, 1996.
- [103] G.O. Roberts and R.L. Tweedie. Geometric convergence and central limit theorem for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83:95–110, 1996.
- [104] G.O. Roberts and R.L. Tweedie. Bounds on regeneration times and convergence rates for Markov Chains. *Stochastic Process. Appl.*, 80(2):211–229, 1999.
- [105] J. S. Rosenthal. *MCMC Handbook*, chapter Optimal Proposal Distributions and Adaptive MCMC. Chapman & Hall/CRC Press, 2009.
- [106] J.S. Rosenthal. Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.*, 90(430):558–566, 1995.
- [107] J.S. Rosenthal. Quantitative convergence rates of Markov chains: A simple account. *Elec. Comm. Prob.*, 7(13):123–128, 2002.
- [108] R. Rubinstein. The cross-entropy methos for combinatorial and continuous optimization. *Methodol. Comput. Appl. Probab.*, 1:127–190, 1999.
- [109] O. Stramer and R.L. Tweedie. Langevin-type models I: diffusions with given stationary distributions and their discretizations. *Methodol. Comput. Appl. Probab.*, 1:307–328, 1999.
- [110] O. Stramer and R.L. Tweedie. Langevin-type models II: self-targeting candidates for MCMC algorithms. *Methodol. Comput. Appl. Probab.*, 1:283–306, 1999.

- [111] R. Swendsen and J.S. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, 58:86–88, 1987.
- [112] T. Tanaka. Information geometry of mean-field approximation. In M. Opper and D. Saad, editors, *Advanced Mean Field Methods*, chapter 17. MIT Press, 2001.
- [113] P. Tuominen and R.L. Tweedie. Subgeometric rates of convergence of f -ergodic Markov chains. *Adv. in Appl. Probab.*, 26:775–798, 1994.
- [114] R. Tweedie and G.O. Roberts. Understanding MCMC. to be published, 2010.
- [115] F. Wang and D.P. Landau. Efficient, Multiple Range Random Walk Algorithm to Calculate the Density of States. *Phys. Rev. Lett.*, 86:2050–5053, 2001.
- [116] M. West. *Bayesian Statistics*, volume 4, chapter Modelling with mixtures, pages 503–525. Oxford University Press, Oxford, 1992.
- [117] J. Zhang. The Mean Fields Theory in EM Procedures for Markov Random Fields. *IEEE Trans. Signal Process.*, 40(10):2570–2583, 1992.
- [118] J. Zhang. The convergence of mean field procedures for MRF's. *IEEE Trans. Image Process.*, 5(12):1662–1665, 1996.

Liste des figures

1.1	Figures extraites de [114]. Réalisation d'une chaîne de Hastings-Metropolis à marche aléatoire symétrique, de loi de proposition gaussienne $\mathcal{N}(0, \sigma^2)$ pour σ^2 grand [gauche], σ^2 petit [centre] et σ^2 approprié [droite]	22
1.2	[de gauche à droite] Densité cible, échantillonneur <i>Equi-Energy</i> , échantillonneur <i>Parallel Tempering</i> [60] et échantillonneur SRWM	25
2.1	Réalisations du processus $t \mapsto \eta_r^\alpha(t; x)$ pour différentes valeurs du point initial x sur la sphère unité de $[-\pi/2, \pi/2]$ et pour différents facteurs d'échelle $r : r \in \{100, 1000, 5000, +\infty\}$ (gauche à droite, et haut en bas)	35
3.1	Figures extraites de [48]. [Gauche] Trajectoire d'un SRWM lorsque π est une loi a posteriori pour l'estimation de paramètres d'une diffusion partiellement observée. [Droite] Evolution de $k \mapsto \ X_k\ ^2$ pour une chaîne SRWM de loi cible $\pi \sim \mathcal{N}_{1000}(0, \text{Id})$ issue de l'origine ($X_0 = 0$)	54
3.2	Ligne de niveau $x + C_x$ délimitant la zone d'acceptation $x + A_x$ et la zone de rejet $x + R_x$. Comparaison de la zone de rejet et du demi-espace défini par l'hyperplan de vecteur normal $\ell(x) := \nabla \log \pi(x)$. $n(x)$ désigne le vecteur normé $x/ x $.	56
3.3	[Gauche] Courbes de niveau de la densité (3.4) et [Droite] de la densité (3.5) lorsque $\delta = 0.4$	57
3.4	Cas de la densité (3.4) [Gauche]: Fonctions Δ en vert et h en bleu. [Droite]: En bleu, trajectoires du processus (2.6) issues d'un ensemble de points sur la sphère unité de $[0, \pi/2]$; en rouge, les trajectoires de l'EDO.	57
3.5	Cas de la densité (3.5) [Gauche]: Fonctions Δ en vert et h en bleu. [Droite]: En vert, trajectoires du processus (2.6) issues d'un ensemble de points sur la sphère unité de $[0, \pi/2]$; en rouge, les trajectoires de l'EDO.	59
3.6	[Gauche] Courbes de niveau de la densité (3.6) avec $\delta = 0.4$ et [Droite] de la densité (3.7) avec $\delta = 0.4$	60
3.7	Cas de la densité (3.6) avec $\delta = 0.4$ [Gauche]: Fonctions Δ en vert et h en bleu. [Droite]: En pointillés, trajectoires du processus (2.6) issues d'un ensemble de point sur la sphère unité de $[0, \pi/2]$; en trait rouge plein, les trajectoires de l'EDO.	60
3.8	Cas de la densité (3.7) avec $\delta = 0.4$. [Gauche]: Fonctions Δ en vert et h en bleu. [Droite]: En pointillés, trajectoires du processus (2.6) issues du point de coordonnées $(\sqrt{2}/2, \sqrt{2}/2)$; en trait plein, les trajectoires de l'EDO.	62

3.9	En rouge, limites fluides issues de la sphère unité, de l'algorithme Metropolis-within-Gibbs dans le cas [gauche] $\pi^{(1)}$ et $\omega^{(1)}$; et [droite] $\pi^{(1)}$ et $\omega^{(2)}$	62
3.10	En rouge, limites fluides issues de la sphère unité, de l'algorithme Metropolis-within-Gibbs dans le cas [gauche] $\pi^{(2)}$ et $\omega^{(1)}$; et [droite] $\pi^{(2)}$ et $\omega^{(2)}$	63
3.11	En rouge, limites fluides issues de la sphère unité de l'algorithme Metropolis-within-Gibbs dans le cas $\pi^{(3)}$ et $\omega^{(1)}$	63
3.12	Option Asiatique lorsque $(v, K, \nu) = (0.1, 45, \nu_*)$. [gauche] directions de stratification successives $t \mapsto \mu^{(t)}$. $\mu^{(0)}$ est proportionnelle au vecteur $(1, \dots, 1)$ donc les d courbes sont issues du même point $1/\sqrt{d}$. Par convention, la première composante de $\mu^{(t)}$ est positive. [droite, haut] valeurs successives de l'estimation $t \mapsto \mathcal{E}^{(t)}$ de la quantité d'intérêt $\mathbb{E}[\phi(X)]$. [droite, bas] valeurs successives de l'estimation de la variance de $\mathcal{E}^{(t)}$; la valeur limite est 0.0016.	78
3.13	Option asiatique [gauche] dans le cas $(v, K, \nu) = (0.1, 45, 0)$ et [droite] dans le cas $(v, K, \nu) = (0.1, 45, \nu_*)$. Directions de stratification μ_g , $\mu^{(N)}$, ν_* (ν_* est normalisé pour être de norme 1), μ_g et μ_l	79
3.14	Option Barrière dans le cas $(K, B) = (50, 60)$ et [gauche] $\nu = 0$ ou [droite] $\nu = \nu_*$: directions $\mu^{(N)}$, μ_{reg} , μ_l , ν_* et μ_g . ν_* a été normalisée pour être de norme 1.	79
4.1	Image Logo : [haut, de gauche à droite] image d'origine, image à segmenter, segmentation initiale obtenue par un algorithme $k\text{-means}$, ind-EM, MC2-EM; [bas, gauche à droite] Gibbsian-EM, Simulated Field, Mean Field, MCVEM, MCVEM + Filtre médian	89
4.2	Image Triangle : [haut, de gauche à droite] image d'origine, image à segmenter, segmentation initiale obtenue par un algorithme $k\text{-means}$, ind-EM, MC2-EM; [bas, gauche à droite] Gibbsian-EM, Simulated Field, Mean Field, MCVEM, MCVEM + Filtre médian	89
4.3	Image SPOT : [haut, de gauche à droite] image à segmenter, segmentation initiale obtenue par un algorithme $k\text{-means}$, ind-EM, MC2-EM; [bas, gauche à droite] Gibbsian-EM, Simulated Field, Mean Field, MCVEM	90
4.4	Projection de la densité cible dans le plan (x_1, x_2) plane. Les contours représentent les régions de plus haute densité a posteriori à 68.3% (bleu), 95% (vert) et 99% (noir)	92
4.5	[de gauche à droite et haut en bas] Evolution de la loi instrumentale q (itérations 1, 3, 5, 7, 9 et 11) et du nuage de points (1 point sur 3 est tracé). Les couleurs indiquent les différentes composantes du mélange q : les centres sont indiqués par des points et les ellipses délimitent les régions de plus haute densité a posteriori à 95%.	93
4.6	Distribution de la perplexité normalisée [haut] et de l'ESS [bas] en fonction du nombre de mise à jour de la loi instrumentale. Ces distributions sont calculées à partir de 500 répétitions de l'algorithme M-PMC. Ces boîtes représentent : la médiane de la distribution (trait noir épais horizontal); les quantiles 0.25 et 0.75 (rectangles en traits fins); et les points aberrants (cercles).	94
4.7	Evolution de la distribution empirique des estimateurs Monte Carlo des quantités $\pi(f_a)$ [haut] et $\pi(f_b)$ [bas] en fonction du nombre de mises à jour, pour les échantilleurs M-PMC [gauche] et <i>Adaptive Metropolis</i> [droite]. En bleu, la vraie valeur de $\pi(f_a)$ et $\pi(f_b)$. Ces boîtes sont calculées à partir de 500 réalisations des estimateurs.	95

4.8	Distribution empirique des estimateurs Monte Carlo des quantités $\pi(f_v)$ pour les échantillonneurs M-PMC [gauche] et <i>Adaptive Metropolis</i> [droite]. En bleu, la vraie valeur de $\pi(f_v)$. Ces boîtes sont calculées à partir de 500 réalisations des estimateurs.	95
4.9	Perplexité [gauche] et ESS [centre] en fonction du nombre de mise à jour de la loi instrumentale (tous les 10 k tirages). [droite] Evolution de la distribution des poids des 10 k points, au fur et à mesure des mises à jour de la loi instrumentale.	97
4.10	Jeu de données SNIa. [droite] Nuage de points approchant la loi a posteriori de (ω, Ω_m) (1 point sur 5 est représenté). [haut gauche et bas] Evolution de la moyenne de chacune des 5 composantes gaussiennes de la loi instrumentale. [haut centre] Loi a posteriori des paramètres (ω, Ω_m) : on représente les régions de plus haute densité a posteriori à 68%, 95% et 99.7%. On montre aussi l'évolution de la loi instrumentale: les cercles sur la figure centrale sont centrés en les moyennes des composantes du mélange, et leur rayon est proportionnel au poids de la composante dans le mélange.	97
4.11	Jeu de données SNIa. Exemples de lois (marginales) a posteriori obtenues par M-PMC (bleu) et SRWM (vert). On représente les contours estimés des zones de plus haute densité a posteriori à 68%, 95% et 99.7%.	98

Liste des tableaux

3.1	Option asiatique: estimateurs stratifiés et estimateur de Monte Carlo usuel	80
3.2	Option Barrière: estimateurs stratifiés et estimateur de Monte Carlo usuel	81
4.1	Image Logo: Estimation des paramètres et taux d'erreur de segmentation.	88
4.2	Image Triangle: Estimation des paramètres et taux d'erreur de segmentation.	90
4.3	Estimation des moyennes et de la précision. Résultats obtenus à partir de 500 répétitions de chaque algorithme.	96
4.4	Jeu de données WMAP5. Estimation des paramètres et intervalles de confiance à 68%	98
4.5	Combinaison des trois jeux de données. Estimation des paramètres et intervalles de confiance à 68%	99

Appendix A

Récapitulatif des activités administratives, d'enseignement et de recherche

A.1 Curriculum Vitae

Gersende Fort,
née le 6 Avril 1974.

url: <http://www.tsi.enst.fr/~gfort/>

LTCI, CNRS - TELECOM ParisTech
46, rue Barrault
75 634 Paris Cedex 13
tél: 01 45 81 80 83
fax : 01 45 88 79 35
email: gersende.fort@telecom-paristech.fr

Chargée de Recherche CNRS (CR1), affectée au LTCI (UMR 5141) - TELECOM ParisTech, Département Traitement du Signal et des Images (TSI).

Thèmes de recherche **(a)** Méthodes de simulation : méthodes de réduction de variance, méthodes Monte Carlo par Chaînes de Markov, méthodes adaptatives. **(b)** Théorie des chaînes de Markov : critères d'ergodicité, contrôle de convergence, modèles fluides.

Autres fonctions

- Enseignant vacataire à TELECOM ParisTech (2006 - · · ·). Vacations à l'Univ. J. Fourier, l'Univ. P. Mendès-France, l'Univ. Paris-Dauphine, l'Univ. Paris VI, l'Univ. Evry-Val-d'Essonne entre 1997 et 2006.
- Coordonnateur du projet ANR-BLAN08 Méthodes de Monte Carlo en grande dimension (2009 - · · ·).
- Coordonnateur du projet ANR-BLAN05 Méthodes de Monte Carlo adaptatives (2006 - 2009).
- Membre du projet ANR-BLAN05 Exploration du modèle cosmologique par fusion statistique de grands relevés hétérogènes (2006 - 2010).

- Organisateur d'un séminaire mensuel BigMC "Méthodes de Monte Carlo en grande dimension" (2006 - · · ·).
- Co-organisateur du Séminaire Parisien de Statistiques (2008 - · · ·).
- Responsable de l'organisaton d'un Workshop international *New directions in Monte Carlo methods*, France, 2007.
- Membre de la Commission de Spécialistes, section 26, Univ. P. Mendès-France (2003) et Univ. J. Monnet (2004 - 2006).

Emplois précédents

2001-2005 Chargée de Recherche CNRS, affectée au LMC (UMR 5523) - Université J. Fourier, Grenoble.

2000-2001 ATER à l'Univ. Paris VI, Paris.

Formation

2001 Thèse de doctorat de l'Univ. Paris VI (spécialité *Mathématiques Appliquées*) ; "Contrôle explicite d'ergodicité de chaînes de Markov : Applications à l'analyse de convergence de l'algorithme Monte Carlo EM", sous la co-direction d'E. Moulines et P. Priouret.

1997 Diplôme d'ingénieur de l'Ecole Nationale Supérieure des Télécommunications (ENST, Paris).

A.2 Encadrement, animation et administration de la recherche

A.2.1 Responsabilités dans l'animation de programmes et projets

2009-2012 Coordonnateur du projet BigMC *Méthodes de Monte Carlo en grande dimension* (programme ANR blanc 2008) qui regroupe trois partenaires académiques: le CEREMADE (Univ. Paris-Dauphine); le CERMICS (ENPC) ; le LTCI (CNRS-Institut TELECOM).

2006-2009 Coordonnateur du projet ADAP'MC *Méthodes de Monte Carlo adaptatives* (programme ANR blanc 2005) qui regroupe trois partenaires académiques: le CEREMADE (Univ. Paris-Dauphine); le CERMICS (ENPC) ; le LTCI (CNRS-Institut TELECOM).

A.2.2 Autres responsabilités

2004-2006 Membre de la Commission de Spécialistes (section 26) de l'Univ. J. Monnet

2003 Membre de la Commission de Spécialistes (section 26) de l'Univ. P. Mendès-France .

A.3 Diffusion de la culture scientifique

A.3.1 Participation à l'organisation de conférences, workshops, numéros spéciaux de revues

2007 Responsable de l'organisation d'un Workshop international *New directions in Monte Carlo methods*, France. (Env. 60 participants).

A.3.2 Organisation de séminaires

2008-… Co-responsable de l'organisation d'un séminaire mensuel, à l'IHP (Paris), “Séminaire Parisien de Statistiques”

2006-… Co-responsable de l'organisation d'un séminaire mensuel, à l'IHP (Paris), sur le thème des “Méthodes de Monte Carlo”.

A.4 Enseignement, formation

A.4.1 Enseignement (à partir de 2001)

2009-… ENS Cachan. Cours sur les Méthodes de Simulation (6h, M2)

2008-… Institut TELECOM. Cours d'optimisation (15h, L3); Cours d'Introduction aux Chaînes de Markov (3h, M1).

2006-… TELECOM ParisTech. Cours de Probabilités (45h, L3); Cours sur les méthodes de Monte Carlo et applications à l'ingénierie financière (18h, M2); Encadrement de projets et stages.

2005-2006 Univ. Paris-Dauphine. Cours et TD de Séries Chronologiques (50 h, M1).

2003-2004 Univ. P. Mendès-France et Univ. J. Fourier. Cours de Probabilités (52h, L1). Cours d'Ingénierie Statistique (12h, M2).

2002-2003 Univ. P. Mendès-France et Univ. J. Fourier. Cours de Probabilités (52h, L1). TD de Théorie de la Mesure (40h, M2).

2001-2002 Univ. P. Mendès-France. Cours de Probabilités (52h, L1). TD de Régression Linéaire (31h, L2).

A.4.2 Thèses co-dirigées

- SYLVAIN LE CORFF Co-encadrement avec E. Moulines (TELECOM ParisTech), bourse EDITE, depuis Septembre 2009. *Méthodes particulières pour la cartographie et localisation simultanées.*
- MOUHAMAD ALLAYA Co-encadrement avec J.M. Bardet (Univ. Paris-Tolbiac), depuis Avril 2009. *Méthodes de Monte Carlo séquentielles pour l'inférence dans des modèles à données cachées.*

A.4.3 Participation à des jurys de thèse

- JULIETTE BLANCHET *Modèles markoviens et extensions pour la classification de données complexes*, Univ. J. Fourier, Octobre 2007.

A.5 Transfert technologique, Relations industrielles et valorisation

A.5.1 Relations Industrielles

2004 Contrat avec Hewlett-Packard (site Grenoble) : *Etude de faisabilité et étude préliminaire à l'élaboration d'une méthodologie pour le traitement de données marketing.* Etude menée conjointement avec A. Antoniadis (Univ. J. Fourier) et S. Lambert-Lacroix (Univ. J. Fourier)

A.6 Production scientifique

A.6.1 Revues avec comité de lecture

1. P. Etoré, G. Fort, B. Jourdain, and E. Moulines. On adaptive stratification. *Annals of Operations Research*, 2009. Accepted for publication.
2. Y. Atchadé and G. Fort. Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. *Bernoulli*, 2009. Accepted for publication.
3. D. Wraith, M. Kilbinger, K. Benabed, O. Cappé, J.F. Cardoso, G. Fort, S. Prunet, and C.P. Robert. Estimation of cosmological parameters using adaptive importance sampling. *Phys. Rev. D*, 80(2):18 pages, 2009.
4. S. Connor and G. Fort. State-dependent Foster-Lyapunov criteria for subgeometric convergence of Markov chains. *Stoch. Process. Appl.*, 119:4176–4193, 2009.

5. R. Douc, G. Fort, E. Moulines, and P. Priouret. Forgetting of the initial distribution for Hidden Markov Models. *Stoch. Process Appl.*, 119(4):1235–1256, 2009.
6. R. Douc, G. Fort, and A. Guillin. Subgeometric rates of convergence of f-ergodic strong Markov processes. *Stoch. Process Appl.*, 119(3):897–923, 2009.
7. G. Fort, S. Meyn, E. Moulines, and P. Priouret. The ODE method for the stability of skip-free Markov Chains with applications to MCMC. *Ann. Appl. Probab.*, 18(2):664–707, 2008.
8. F. Forbes and G. Fort. Combining Monte Carlo and Mean-Field-like methods for inference in Hidden Markov Random Fields. *IEEE Transactions on Image Processing*, 16(3):824–837, 2007.
9. G. Fort, S. Lambert-Lacroix, and J. Peyre. Réduction de dimension dans les modèles généralisés : Application à la classification de données issues des biopuces. *Journal de la SFDS*, 146(1-2):117–152, 2005.
10. G. Fort and S. Lambert-Lacroix. Classification using Partial Least Squares with Penalized Logistic Regression. *Bioinformatics*, 21(7):1104–1111, 2005.
11. G. Fort and G.O. Roberts. Subgeometric ergodicity of strong Markov processes. *Ann. Appl. Probab.*, 15(2):1565–1589, 2005.
12. R. Douc, G. Fort, E. Moulines, and P. Soulier. Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab.*, 14(3):1353–1377, 2004.
13. G. Fort, E. Moulines, G.O. Roberts, and J.S. Rosenthal. On the geometric ergodicity of hybrid samplers. *J. Appl. Probab.*, 40(1):123–146, 2003.
14. G. Fort and E. Moulines. Polynomial ergodicity of Markov transition kernels. *Stochastic Processes Appl.*, 103(1):57–99, 2003.
15. G. Fort and E. Moulines. Convergence of the Monte-Carlo EM for curved exponential families. *Ann. Stat.*, 31(4):1220–1259, 2003.
16. G. Fort and E. Moulines. V-subgeometric ergodicity for a Hastings-Metropolis algorithm. *Stat. Probab. Lett.*, 49(4):401–410, 2000.

A.6.2 Chapitres de livre

G. Fort, E. Moulines, and P. Soulier. *Inference in Hidden Markov Models*, chapitre 14 : "Elements of Markov Chain Theory", pages 511–562. Springer, 2005.

A.6.3 Actes de conférences internationales avec comité de lecture

1. G. Fort, S. Meyn, E. Moulines and P. Priouret. ODE methods for Markov chain stability with applications to MCMC. *Proceedings of the 1st International Conference on Performance Evaluation Methodologies and Tools*, Valuetools, Art. 42, 2006.

2. G. Fort and S. Lambert-Lacroix. Ridge-Partial Least Squares for Generalized Linear Models with binary response. COMPSTAT'04, *Proceedings in Computational Statistics*, pages 1019-1026, 2004.
3. G. Fort, E. Moulines, and P. Soulier. On the convergence of iterated random maps with applications to the MCEM algorithm. COMPSTAT'98, *CProceedings in Computational Statistics*, 1998.
4. G. Fort, O. Cappé, E. Moulines, and P. Soulier. Optimization via simulation for maximum likelihood estimation in incomplete data models. In Proc. IEEE Workshop on Stat. Signal and Array Proc., pages 80-83, 1998.

A.6.4 Rapports de Recherche

1. Y. Atchadé, G. Fort, E. Moulines, and P. Priouret. Adaptive Markov chain Monte Carlo : Theory and Methods. Technical report, 2009.
2. G. Fort. Fluid limit-based tuning of some hybrid MCMC samplers. Technical report, 2007.
3. C. Andrieu and G. Fort. Explicit control of subgeometric ergodicity. Technical Report 05:17, University of Bristol, UK, 2005.
4. G. Fort. Partial Least Squares for classification and feature selection in Microarray gene expression data. Technical report, UNiv. J. Fourier, France, 2004.
5. G. Fort. Computable bounds for V-geometric ergodicity of Markov transition kernels. Technical Report RR 1047-M, Univ. J. Fourier, France, 2002.

A.6.5 Thèse de Doctorat

G. Fort. *Contrôle explicite d'ergodicité de chaînes de Markov : Applications à l'analyse de convergence de l'algorithme Monte Carlo EM*. Thèse de Doctorat de l'Université Paris VI, soutenue en Juin 2001.

A.6.6 Conférences invitées

1. G. Fort. Adaptive MCMC : theory and methods Conference "Optimization in MCMC", Warwick, GB, Juin 2009
2. G. Fort. Stability of Markov Chains based on fluid limit techniques. Applications to MCMC Congrès SSC-SFDS, Ottawa, Canada, Mai 2008.
3. G. Fort. Fluid limit-based tuning of some hybrid MCMC samplers ADAP'ski, Bormio, Italie, Janvier 6-12 2008.
4. G. Fort and G.O. Roberts. Criteria for subgeometric ergodicity of strong Markov processes. Workshop New Developments in MCMC (Diffusions, Images and Other Challenges), Warwick (GB), Août 2006.

5. G. Fort, S. Meyn, E. Moulines and P. Priouret. Limites fluides de quelques échantillonneurs MCMC. *Journées MAS, Lille*, Septembre 2006.
6. G. Fort. Some recent results on Hybrid Samplers. *Workshop: MCMC methodology, Lancaster (GB)*, Décembre 2001.
7. G. Fort. Convergence of the MCEM algorithm. *Workshop: MCMC methodology, Lancaster (GB)*, Décembre 2001.
8. G. Fort and E. Moulines. *European Conference on Spatial and Computational Statistics, Ambleside (GB)*, Septembre 2000.

A.6.7 Exposés dans des conférences ou séminaires (à partir de 2001)

1. G. Fort. On adaptive stratification *2009 INFORMS Applied Probability Society conference, Ithaca, USA*, Juillet 2009.
2. G. Fort. Théorèmes limites pour les échantillonneurs MCMC adaptatifs *Séminaire Méthodes de Monte Carlo en grande dimension, Paris*, Juin 2009.
3. G. Fort. Limit theorems for Adaptive MCMC algorithms *41èmes Journées de Statistiques de la SFDS, Bordeaux*, Mai 2009.
4. G. Fort. MCMC et approximations en champs moyen pour les modèles de Markov *Journée GDR ISIS "Problèmes inverses"*, Mars 2009.
5. G. Fort. Stability of Markov Chains based on fluid limit techniques. Applications to MCMC *Séminaire du département de Statistique, Warwick, UK*, Juin 2008.
6. G. Fort. Stabilité des chaînes de Markov par la méthode des limites fluides. Applications à l'étude de méthodes MCMC adaptatives *Séminaire de l'Institut Mathématique de Bordeaux*, Mars 2008.
7. G. Fort. Stabilité des chaînes de Markov par la méthode des limites fluides. Applications à l'étude de méthodes MCMC adaptatives *Séminaire de Statistique, LSP Toulouse*, Janvier 2008.
8. G. Fort. Fluid limit for Hybrid MCMC samplers *INFORMS Applied Probability, Eindhoven*, 9-11 Juillet 2007.
9. G. Fort. Limites fluides de quelques échantillonneurs MCMC *Séminaire de Statistiques, Grenoble*, Mai 2007.
10. G. Fort. Limites fluides et Stabilité des Chaînes de Markov. Application aux échantillonneurs MCMC *Séminaire TREK, ENS Paris*, Février 2007.
11. G. Fort. Limites fluides et Stabilité des Chaînes de Markov. Application aux échantillonneurs MCMC *Séminaire Parisien de Statistiques, Paris*, Novembre 2006

12. G. Fort. L'algorithme Ridge-Partial Least Squares et application à la classification de puces à ADN. *Séminaire de Statistique de Grenoble*, Juin 2004.
13. G. Fort. L'algorithme Ridge-Partial Least Squares et application à la classification de puces à ADN. *Séminaire SAMOS, Paris 1*, Mai 2004.
14. G. Fort. A drift criterion for subgeometric ergodicity of Markov transition kernel. *Séminaire de Statistique, Lancaster (GB)*, Mars 2003.
15. G. Fort. A drift criterion for subgeometric ergodicity of Markov transition kernel. *Séminaire de Statistique, Toronto (Canada)*, Novembre 2002.
16. G. Fort. Critère d'ergodicité polynomiale: applications à l'analyse de convergence d'algorithmes MCMC. *Séminaire de Statistique de Lille*, Janvier 2002, et *Séminaire de Statistique de Grenoble*, Décembre 2001.