
Numéro d'ordre : 4787

Thèse présentée à l'université Bordeaux 1

Ecole Doctorale de Mathématiques et Informatique de Bordeaux

par GIRAUD François

Pour obtenir le grade de Docteur en Mathématiques

**ANALYSE DES MODÈLES PARTICULAIRES
DE FEYNMAN-KAC
ET APPLICATION À LA RÉOLUTION
DE PROBLÈMES INVERSES
EN ÉLECTROMAGNÉTISME**

Thèse dirigée par Pierre DEL MORAL, et co-encadrée par Pierre MINVIELLE

Soutenue le 29 mai 2013

Jury :

MALRIEU Florent, Maître de Conférences HDR, Université Rennes 1

LAMBERT Marc, Chargé de Recherche HDR, CNRS

JOURDAIN Benjamin, Professeur à l'École Nationale des Ponts et Chaussées

GIOVANNELLI Jean-François, Professeur d'Université, Université Bordeaux 1

GOBET Emmanuel, Professeur à l'École Polytechnique

DEL MORAL Pierre, Directeur de Recherche, INRIA Bordeaux Sud-Ouest

MINVIELLE Pierre, Ingénieur Chercheur, CEA-CESTA

La présente thèse a été préparée au centre CEA-CESTA, 15 avenue des sablières, CS 60001, 33116 Le Barp Cedex.

Résumé

L'algorithme SMC (pour Sequential Monte Carlo) est un algorithme stochastique particulier de type génétique permettant la simulation de lois de probabilité complexes sur des ensembles multidimensionnels. Dans une première partie théorique, nous nous penchons sur une analyse rigoureuse des performances de l'algorithme, basée sur l'étude de la stabilité des flots de Feynman-Kac, et conduisant à des résultats de type bornes L^p et inégalités de concentration. Nous abordons notamment le cas particulier des SMC associés à des schémas de température, et analysons sur ce sujet un processus à schéma adaptatif. Dans une seconde partie appliquée, nous illustrons son utilisation par la résolution de problèmes inverses concrets en électromagnétisme. Le plus important d'entre eux consiste à estimer les propriétés radioélectriques de matériaux recouvrant un objet de géométrie connue, et cela à partir de mesures de champs rétrodiffusés. Nous montrons comment l'algorithme SMC, couplé à des calculs analytiques, permet une inversion bayésienne en tirant profit de la structure spécifique du problème, et fournit des estimées robustes enrichies d'estimations des incertitudes.

Mots-clés : algorithmes génétiques, formules de Feynman-Kac, problèmes inverses.

Abstract

Sequential and Quantum Monte Carlo methods, as well as genetic type search algorithms, can be interpreted as a mean field and interacting particle approximation of Feynman-Kac models in distribution spaces. The performance of these population Monte Carlo algorithms is strongly related to the stability properties of nonlinear Feynman-Kac semigroups. In a first theoretical part, we analyze these models in terms of Dobrushin ergodic coefficients of the reference Markov transitions and the oscillations of the potential functions. Sufficient conditions for uniform concentration inequalities w.r.t. time are expressed explicitly in terms of these two quantities. We provide an original perturbation analysis that applies to annealed and adaptive FK models, yielding what seems to be the first results of this kind for these type of models. Special attention is devoted to the particular case of Boltzmann-Gibbs measures' sampling. In this context, we design an explicit way of tuning the number of Markov Chain Monte Carlo iterations with temperature schedule. We also propose and analyze an alternative interacting particle method based on an adaptive strategy to define the temperature increments. In a second, applied part, we illustrate the use of these SMC algorithms in the field of inverse problems. Mainly, the following electromagnetism (EM) inverse problem is addressed. It consists in estimating local radioelectric properties of materials recovering an object from global EM scattering measurements, at various incidences and wave frequencies. This large scale ill-posed inverse problem is explored by an intensive exploitation of an efficient 2D Maxwell solver, distributed on high performance computing machines. Applied to a large training data set, a statistical analysis reduces the problem to a simpler probabilistic metamodel, on which Bayesian inference can be performed. Considering the radioelectric properties as a hidden dynamic stochastic process, that evolves in function of the frequency, it is shown how the Sequential Monte Carlo methods can take benefit of the structure and provide local EM property estimates.

Keywords : genetic algorithms, Feynman-Kac formulae, inverse problems.

Remerciements

En premier lieu, je remercie mes directeurs, Pierre Del Moral et Pierre Minvielle, pour leur investissement, leurs suggestions et leur soutien du début à la fin de cette thèse.

Merci au CEA-CESTA d'avoir initié, encadré et financé ce projet, tout en me laissant jouir d'une grande liberté. Merci en particulier au service SSPP pour son aide, sa convivialité, et son intérêt pour mes travaux, fussent-ils en marge de ses activités propres.

Merci à toute l'équipe ALEA pour le cadre intellectuel très riche qu'elle m'a fait partager. Merci en particulier à François Caron et Pierrick Legrand pour l'aide généreuse qu'ils m'ont apportée à chaque fois que j'ai fait appel à leurs conseils.

De façon plus personnelle, je tiens à remercier tous les thésards du bureau *B-417*, notamment mes plus anciens camarades, Frédéric, Romain et Philippe. L'ambiance qui règne dans ce lieu, mélange de dérision et de solidarité, a été pour moi un socle affectif important pendant toute la thèse. J'ai pu en outre y régler tous mes problèmes de \LaTeX , ce qui est sans doute encore plus considérable.

Un grand merci au collectif Visan, né en même temps que cette thèse. Mois après mois, j'en ai apprécié les grands rassemblements comme les petites complicités, les Olympiades épiques comme les belotes improvisées. Parmi ses membres, je remercie tout particulièrement M. Maël Mevel. Pour son soutien moral, son énergie et sa sollicitude d'abord, mais aussi pour son entreprise de rhumerie domestique, sa frappe croisée pied droit, son goût pour les bonnes choses et surtout son mauvais-perdantisme congénital (charme ultime s'il en est).

Un grand merci à tous les membres du Century Funk, pour la musique, l'aventure, et l'amitié dans l'aventure. Il est paradoxal de voir apparaître ici ce nom tant les impératifs de groupe et les impératifs de thèse se sont fait la guerre pendant cette singulière année 2010/2011. Mais il me semble que des remerciements dignes ne sauraient ignorer ces moments-là. Merci à toute la bande, et j'espère à bientôt.

Je remercie chaleureusement M. Nicolas Lelong pour les mille joies quotidiennes que m'a offertes sa colocation, au nombre desquelles sa version piano de *Moving On*, de George Duke, ses parodies

de l'émission *En quête d'action à qui profite ce business*, son indifférence totale à l'avancée de ma thèse, et son travail remarquable sur le jeu *Super Smash Bros. Brawl*. A ce propos, je tiens néanmoins à contre-remercier fermement M. Pit pour sa piètre prestation dans ce jeu, digne à mon sens du plus insupportable personnage de jeux vidéos jamais créé.

Je profite également de cette section pour remercier chaleureusement MM. Paul McCartney, Stevie Wonder et Antonio Carlos Jobim, qui sont pour moi les trois plus grands compositeurs du XX^{ème} siècle, ainsi que Breakbot et Amy Winehouse, détenteurs provisoires du titre pour le XXI^{ème}. Bon gré mal gré, les choses sont ainsi faites que leurs mélodies seront pour moi à jamais associées aux processus particuliers de Feynman-Kac... Je les cite donc ici au titre légitime de fantômes de ce document.

Je remercie de même les développeurs des jeux *Super Mario Kart*, *Mario Kart 64*, *Mario Kart Double Dash* et *Mario Kart Wii*, très rarement remerciés par la communauté scientifique, ce qui est une hypocrisie, vu la quantité de plaisir et de défoulement qu'elle leur doit. A titre personnel, je remercie tout particulièrement le personnage de Bowser pour sa spectaculaire inertie, son rire gras et sa constance dans la cruauté, qui font de lui mon partenaire de course préféré.

Enfin et plus sérieusement, merci à ma Judith, elle sait pourquoi.

Préambule

Le présent manuscrit résulte de la combinaison de deux travaux. Premièrement, une étude théorique des processus de Feynman-Kac, en collaboration avec Pierre Del Moral, étude visant à l'analyse rigoureuse de la convergence des méthodes SMC. Deuxièmement, une étude appliquée, en collaboration avec Pierre Minvielle, consistant à élaborer et tester des méthodes de résolution de problèmes inverses mettant en oeuvre ces méthodes SMC. En bien des sens, ces travaux se sont entremêlés. D'un point de vue simplement chronologique, d'abord, les trois années et demi de thèse se caractérisant par le passage fréquent de l'un à l'autre. D'un point de vue scientifique, ensuite, les résultats théoriques m'invitant à les illustrer en pratique, et les méthodes développées en pratique m'invitant à les analyser de façon théorique. Comme tout travail de recherche scientifique, ce travail de thèse est le fruit d'un certain cheminement de pensée, quelque peu masqué par la structure finale du manuscrit. Ce préambule est l'occasion d'en donner quelques grandes lignes.

Le point de départ de la partie théorique est l'algorithme ISA (pour *Interacting Simulated Annealing*). Cette méthode, sorte de version particulière du recuit simulé, est une application du SMC à la simulation d'une suite de mesures de Gibbs, utilisée en optimisation globale. Malgré plusieurs travaux non asymptotiques sur le SMC général, il lui manquait une étude théorique globale, explicitant des conditions sur ses paramètres pour assurer sa convergence directement en termes d'optimisation. Cet objectif a naturellement conduit à l'entreprise de dépliage des équations de Feynman-Kac développée dans la section 2.2, puis à sa traduction dans les modèles à schémas de température, développée section 2.3. L'approche considérée est entièrement non asymptotique, et s'appuie sur les notions de stabilité des flots de Feynman-Kac à travers les coefficients de Dobrushin des noyaux de Markov des mutations, ainsi que les oscillations des potentiels de sélection. Elle fournit en définitive des résultats assez parallèles à ceux obtenus pour le recuit simple, mais sous des formes intéressantes, notamment des inégalités de concentration très riches en interprétation.

Les premiers travaux appliqués sont focalisés sur l'imagerie radar SAR/ISAR (pour *Inverse Synthetic Aperture Radar*). Le problème consiste à obtenir une image des contributeurs électromagnétiques d'un objet à partir de mesures radar monochromatiques réalisées à des fréquences et des angles d'incidence différents. Il s'agit en soi d'un problème inverse. Les méthodes classiques d'inversion reposent entre autres sur une modélisation détermi-

niste, une approximation champ lointain et un modèle d'approximation linéaire (modèle des points brillants, voir par exemple [58, 54]), ramenant le problème à une inversion des équations de Fourier. Plusieurs améliorations sont envisageables. Au niveau de la modélisation, d'abord : adaptation en ondes sphériques en champ proche, prise en compte du diagramme de rayonnement des antennes, introduction d'un bruit de mesure et gestion d'éventuelles données manquantes. Au niveau des résultats, ensuite : disparition des artefacts propres aux modèles classiques, plus grande précision dans la localisation des contributeurs électromagnétiques principaux et estimation des incertitudes. Dans ce contexte, l'approche bayésienne est intéressante à plusieurs titres. D'abord, elle propose un formalisme général non linéaire, et très souple, permettant de complexifier le modèle direct. Ensuite, elle se caractérise par la notion de vraisemblance, qui est adaptée aux modèles aléatoires de bruit de mesure, et la notion de distribution a priori, qui permet de formaliser des informations sur l'inconnue. Enfin, elle présente la solution du problème inverse sous la forme d'une densité de probabilité, résultat plus riche en interprétation et en fiabilité qu'un estimateur unique. En tant que méthodes efficaces d'échantillonnage de loi a posteriori, les méthodes SMC paraissent un outil pertinent pour réaliser numériquement l'inversion.

Cependant, après une période de documentation, tests, concertations et réflexions sur ces modèles, cette piste fut abandonnée, en tout cas sous cette forme. Un des arguments clés, révélé par une étude bibliographique sur les méthodes stochastiques en imagerie, était le problème de la dimension, souvent central en simulation stochastique (voir par exemple [42, 75], [47] ou [72] et ses références). En outre, dans ce contexte linéaire de grande dimension, une piste beaucoup plus naturelle se dessinait : les méthodes de régularisation. Ces méthodes, très courantes en imagerie (voir par exemple [78, 70, 71]), consistent essentiellement à interpréter la solution du problème inverse comme le minimiseur d'une fonctionnelle impliquant deux termes : un terme d'adéquation aux données, lié au modèle de vraisemblance, et un critère de pénalisation, lié à l'a priori sur la solution. Sauf pour certaines situations spécifiques, comme par exemple l'estimation d'hyperparamètres du modèle, qui peuvent faire intervenir des méthodes stochastiques, ces approches sont basées sur des résolutions par optimisation convexe. La poursuite dans cette voie nous a semblé éloignée de l'orientation probabiliste de la thèse, et incompatible avec les travaux théoriques.

La partie appliquée de la thèse s'est donc réorientée vers un autre type de problèmes inverses en électromagnétisme, proche de l'imagerie : le contrôle. Deux applications ont été abordées, et sont développées respectivement dans la section 3.2 et le chapitre 4. La première consiste à déterminer des propriétés électromagnétiques d'un matériau multicouche 1D à partir de mesures de coefficients de réflexion. Il s'agit d'un cas d'étude simple et relativement artificiel d'un point de vue physique, mais dont l'étude révèle un grand intérêt mathématique. Elle met en lumière une méthodologie de *modélisation-inversion bayésienne-échantillonnage stochastique* capable de résoudre un problème, certes simple dans sa formulation, mais très complexe dans sa résolution : problème mal posé, très forte irrégularité du modèle direct, densités multimodales, ensemble solution de taille infime. En outre, sa mise en oeuvre informatique révèle l'importance cruciale des méthodes adapta-

tives de sélection/mutation dans l'implémentation pratique du SMC.

La seconde application consiste à déterminer des propriétés électromagnétiques de matériaux recouvrant un objet 3D de géométrie connue, à partir de coefficients de rétrodiffusion. L'inconnue admet une dépendance à la fois spatiale et fréquentielle, ce qui fait de ce problème un problème de grande dimension. Dans sa résolution, l'algorithme SMC s'inscrit au sein d'une large démarche probabiliste mêlant apprentissage statistique, hiérarchie bayésienne, construction d'une loi a priori séquentielle, calcul analytique et simulation stochastique.

Dans chacun de ces problèmes concrets, l'implémentation pratique de l'algorithme SMC a mis en évidence deux principes importants pour garantir de bonnes performances. Premièrement, le fait de "tempérer" toute distribution difficile à simuler par un schéma de "température", pour aborder sa complexité progressivement. Deuxièmement, la gestion de ce schéma de température de façon adaptative, c'est-à-dire en tenant compte du comportement du nuage de particules courant. Formellement, le processus ainsi défini ne s'inscrit pas dans la classe d'algorithme étudié théoriquement, car il implique des quantités aléatoires (incrémentations du paramètre de température) supposées déterministes dans l'analyse classique du SMC. Comme à ma connaissance ces processus, d'utilisation pourtant courante, n'avaient pas été analysés en tant que tels, j'ai cherché à les étudier selon l'approche non asymptotique qui m'était familière. L'idée est simple : mettre formellement cet incrément aléatoire sous la forme d'une fonction déterministe de la mesure d'occupation des particules, ce qui permet, en remontant aux propriétés de régularité de cette fonction, de voir le processus comme une perturbation d'une suite de Feynman-Kac théorique, mais parfaitement déterministe. Ce schéma d'analyse est développé dans la section 2.4, et aboutit à des estimées de type bornes L^p ainsi qu'à des inégalités de concentration.

Finalement, le manuscrit est organisé en quatre chapitres. Le chapitre 1 est une simple introduction aux méthodes particulières, SMC en tête. Le lecteur profane y trouvera de quoi se familiariser avec les principes de base des méthodes de simulation stochastique, et le lecteur averti quelques définitions et notations (plus probabilistes que statisticiennes) récurrentes dans le document. Le chapitre 2 regroupe tous les résultats théoriques : étude des semigroupes de Feynman-Kac, estimées uniformes en temps dans le cas général, performances de l'ISA et analyse des modèles particuliers à schéma de température adaptatif. Le chapitre 3 est consacré au lien entre l'analyse théorique des méthodes SMC et leur utilisation pratique. D'abord via des exemples jouets, construits dans des contextes parfaitement contrôlés, pour illustrer la pertinence de nos estimées. Puis à l'inverse via le problème inverse multicouche 1D, qui permet d'aborder les problèmes pratiques. Enfin, le chapitre 4 est dédié au problème inverse en grande dimension mentionné ci-avant.

Un soin particulier a été apporté au référencement des notations et définitions, afin de rendre ces quatre chapitres lisibles indépendamment les uns des autres.

Table des matières

1	Introduction aux méthodes particulières	1
1.1	Noyaux de Markov classiques	3
1.1.1	Noyaux de Metropolis-Hastings	3
1.1.2	Noyau de Gibbs	4
1.1.3	Noyau de Metropolis-Hastings hybride	5
1.2	Méthodes de simulation stochastiques	6
1.2.1	Méthodes MCMC	6
1.2.2	Optimisation par recuit simulé	7
1.2.3	Parallel Tempering Algorithm	8
1.2.4	Metropolis-Hastings hybride en interaction	9
1.3	Introduction à l’algorithme SMC	10
1.3.1	Définition générale et implémentation	10
1.3.2	Mesures cibles, flots de Feynman-Kac	12
1.3.3	L’exemple du filtrage particulière	13
2	Analyse théorique	15
2.1	Préliminaires	21
2.1.1	Coefficients de Dobrushin	21
2.1.2	Modèles de Feynman-Kac	22
2.1.3	L’algorithme SMC classique	24
2.1.4	Quelques résultats non asymptotiques	25
2.2	Analyse non asymptotique générale	27
2.2.1	Stabilité du semigroupe	28
2.2.2	Théorèmes de concentration uniformes	31
2.3	Méthodes SMC à schéma de température, recuits simulés en interaction	36
2.3.1	Motivations	36
2.3.2	Un modèle d’optimisation ISA	39
2.4	Une méthode SMC à schéma de température adaptatif	42
2.4.1	Représentation de Feynman-Kac	43
2.4.2	Un processus particulière adaptatif	44
2.4.3	Analyse de perturbation	45
2.4.4	Résultats de convergence non asymptotiques	50

3	Théorie et pratique des méthodes SMC : illustrations	57
3.1	Exemples "jouets"	58
3.1.1	Illustration sur un espace fini, limites de l'approche Dobrushin . . .	59
3.1.2	Illustration sur un exemple en 2D	60
3.2	Problème inverse multicouche 1D	66
3.2.1	Paramétrisation du problème, définition de l'espace d'état E	67
3.2.2	Modélisation probabiliste	68
3.2.3	Echantillonnage par algorithme SMC	70
3.2.4	Résultats	73
4	Diffraction inverse	83
4.1	Problème de diffraction inverse	85
4.1.1	Mesures de rétrodiffusion électromagnétique	85
4.1.2	Contrôle non destructif	87
4.1.3	Un problème inverse aux équations de Maxwell	87
4.2	Formulation statistique du problème	89
4.2.1	Modélisation à une fréquence fixée f_k	89
4.2.2	Formulation du problème global	93
4.2.3	Processus autorégressif généralisé	93
4.2.4	Structure conditionnelle et chaîne de Markov cachée	94
4.3	Inversion probabiliste	95
4.3.1	L'approche Rao-Blackwellisée	95
4.3.2	Algorithme SMC et chemin d'interpolation dans l'espace des mesures de probabilité	96
4.3.3	Estimation globale	98
4.4	Applications	100
4.4.1	Scénario de test non destructif	100
4.4.2	Processus d'inversion	102
4.4.3	Analyse de performance	104
4.4.4	Analyses de variantes	108
Bibliographie		117

Chapitre 1

Introduction aux méthodes particulières

Notations fondamentales

Soit (E, r) un espace métrique complet séparable, et soit \mathcal{E} la tribu borélienne associée. On note $\mathcal{P}(E)$ l'ensemble des mesures de probabilité sur E . Soit $\mathcal{B}(E)$ l'espace des fonctions de E dans \mathbb{R} mesurables bornées. On note $\mathcal{B}_1(E) \subset \mathcal{B}(E)$ le sous-ensemble des fonctions bornées par 1.

Fixons une mesure de probabilité $\mu \in \mathcal{P}(E)$, une fonction bornée $f \in \mathcal{B}(E)$ et trois noyaux de Markov K, K_1, K_2 sur E pour définir les quantités suivantes :

- le produit mesure-fonction, noté $\mu.f$ ou $\mu(f)$ désigne l'intégrale de f contre μ soit

$$\mu(f) = \int_E f(x)\mu(dx) \quad (\in \mathbb{R});$$

- le produit noyau-noyau $K_1.K_2$ désigne le noyau de Markov défini par

$$(K_1.K_2)(x, A) = \int_E K_1(x, dy)K_2(y, A).$$

Pour tout élément x de E , obtenir une simulation z du noyau $(K_1.K_2)(x, \cdot)$ peut se faire en simulant d'abord $y \sim K_1(x, \cdot)$ puis en simulant $z \sim K_2(y, \cdot)$;

- le produit noyau-fonction $K.f$ désigne la fonction de $\mathcal{B}(E)$ définie par

$$K.f(x) = \int_E K(x, dy)f(y).$$

Si X_n désigne une chaîne de Markov homogène de noyau de transition K , alors $K.f(x)$ peut s'interpréter comme l'espérance de $f(X_{n+1})$ sachant $X_n = x$;

- le produit mesure-noyau $\mu.K$ désigne la mesure de probabilité définie par

$$\mu.K(A) = \int_E K(x, A)\mu(dx). \tag{1.1}$$

Simuler $z \sim \mu.K(\cdot)$ peut se faire en simulant d'abord $x \sim \mu(\cdot)$, puis $z \sim K(x, \cdot)$.

Ces opérations entre mesures, noyaux de Markov et fonctions bornées sont clairement linéaires, et se comportent entre elles de façon associative.

Si $G \in \mathcal{B}(E)$ est une fonction strictement positive, alors $\psi_G : \mathcal{P}(E) \rightarrow \mathcal{P}(E)$ désigne la transformation de Boltzmann-Gibbs associée à G , définie par

$$\forall \mu \in \mathcal{P}(E), \forall f \in \mathcal{B}(E), \quad \psi_G(\mu)(f) = \frac{\mu(G \times f)}{\mu(G)}. \quad (1.2)$$

Enfin, pour tout $x \in E$, δ_x désigne la mesure de Dirac centrée en x .

Le principe d'approximation de Monte Carlo

Soit η une mesure de probabilité sur l'espace mesuré (E, \mathcal{E}) , que l'on cherche à étudier (recherche de l'espérance, variance, modes principaux, probabilité de dépassement d'un seuil, estimation de la densité complète, etc.). Le principe très général des méthodes de Monte Carlo est d'approcher cette mesure η par une mesure discrète

$$\eta^N := \sum_{k=1}^N \omega_k \delta_{X_k}$$

où les X_k sont des variables aléatoires et les ω_k des nombres positifs (possiblement aléatoires) tels que $\sum_k \omega_k = 1$. Il est à noter d'emblée que η^N est par définition une mesure discrète, donc non-représentable par une fonction de densité continue, ce qui exclut toute forme d'approximation de η au sens fort (comme la convergence uniforme des densités par exemple). D'autre part, η^N est une mesure aléatoire, ce qui signifie que les résultats d'approximation à espérer sont des résultats dans des sens probabilistes (convergence en loi, en probabilité, convergence presque sûre, convergence de moments, inégalités de concentration, etc.).

L'exemple le plus célèbre, et le plus simple, est la méthode de Monte Carlo pour le calcul de l'intégrale d'une fonction f par l'utilisation d'un échantillon i.i.d. (X_1, \dots, X_N) . Plus précisément, si l'on cherche à calculer, pour une fonction f et une mesure η la quantité :

$$\eta(f) = \int_E f(x) \eta(dx)$$

et si (X_1, \dots, X_N) sont N variables aléatoires indépendantes de loi η , alors la mesure discrète $\eta^N := \frac{1}{N} \sum_{k=1}^N \delta_{X_k}$ vérifie, sous la condition $\sigma_f := \text{Var}(f(X_k)) = \eta(f^2) - \eta(f)^2 < +\infty$, les propriétés bien connues suivantes :

1. $\eta^N(f)$ est un estimateur non biaisé de $\eta(f)$ vérifiant $\eta^N(f) \xrightarrow[N \rightarrow \infty]{} \eta(f)$ p.s. (loi des grands nombres)
2. la variable aléatoire $\sqrt{N} (\eta^N(f) - \eta(f))$ converge en loi vers la gaussienne $\mathcal{N}(0, \sigma_f^2)$ quand $N \rightarrow +\infty$. (théorème limite central)

1.1. NOYAUX DE MARKOV CLASSIQUES

Vis-à-vis de l'intérêt de ces résultats en intégration numérique, il est important de remarquer qu'ils ne dépendent pas de la dimension de l'espace E , et surtout qu'ils ne dépendent pas de la régularité de f , qui n'est supposée que mesurable sur (E, \mathcal{E}) . D'un point de vue probabiliste, ces résultats (et d'autres) affirment qu'en de nombreux sens, η^N est une approximation de η .

Il est néanmoins difficile dans la plupart des cas de simuler directement une loi η d'intérêt. Nous présentons dans la suite de ce chapitre quelques méthodes classiques d'approximation stochastique.

1.1 Noyaux de Markov classiques

De nombreuses méthodes de simulation reposent sur la capacité de savoir simuler non pas la mesure cible η , mais un noyau de Markov K qui laisse stable η , c'est-à-dire qui vérifie

$$\eta.K = \eta.$$

Définissons ici quelques noyaux fondamentaux, très utiles dans la suite.

1.1.1 Noyaux de Metropolis-Hastings

Le noyau de Metropolis-Hastings (voir [63] pour une description plus précise) est utilisé dans la simulation de lois de probabilité de la forme

$$\eta(dx) = \frac{1}{Z} h(x) m(dx)$$

où h est une fonction positive bornée $E \rightarrow \mathbb{R}_+$, m une mesure de référence quelconque sur E , et Z une constante de normalisation. Il s'agit là d'une forme très générale, pouvant décrire à peu près toutes les distributions rencontrées en pratique, mais le point important de ce noyau est que sa simulation ne requiert pas de connaître la constante Z . Autrement dit, il suffit de savoir calculer les ratios $h(x)/h(y)$.

En plus de cette fonction h , le noyau de Metropolis-Hastings fait intervenir un noyau de proposition de la forme $K(x, dy) = T(x, y).m(dy)$ choisi a priori par l'utilisateur, et se définit par les relations :

$$\begin{aligned} K_{h,T}(x, dy) &= K(x, dy). \min(1, r(x, y)) && \forall y \neq x \\ K_{h,T}(x, \{x\}) &= 1 - \int_{y \neq x} K(x, dy). \min(1, r(x, y)) \end{aligned} \tag{1.3}$$

$$\text{où } r(x, y) := \frac{h(y) T(y, x)}{h(x) T(x, y)}.$$

1.1. NOYAUX DE MARKOV CLASSIQUES

Une façon simple de simuler ce noyau $K_{h,T}(x, dy)$ en pratique est de commencer par simuler un $y_{prop} \sim K(x, dy)$ (étape de proposition), puis de choisir $y := y_{prop}$ avec probabilité $\min(1, r(x, y_{prop}))$, et $y := x$ avec proba $1 - \min(1, r(x, y_{prop}))$ (étape d'acceptation/rejet).

Il est aisé de vérifier que quel que soit le noyau de proposition K utilisé, $K_{h,T}$ laisse η invariant, soit $\eta.K_{h,T} = \eta$. Dans la perspective pratique de simulation de η par le biais d'une ou plusieurs particules, indépendantes ou en interaction, ce choix de K est décisif car c'est ce noyau qui guide l'exploration de E par la ou les particules.

Un cas particulier simple mais important est celui du noyau du recuit simulé. Si l'on se fixe un potentiel borné $V : E \rightarrow \mathbb{R}$ et un noyau de proposition de la forme $K(x, dy) = T(x, y).m(dy)$, alors pour tout paramètre $\beta \in \mathbb{R}$, on note K_β le noyau de Metropolis-Hastings $K_{h_\beta, T}$ où

$$\forall x \in E, \quad h_\beta(x) = e^{-\beta V(x)}.$$

Autrement dit, K_β est défini par les relations

$$\begin{aligned} K_\beta(x, dy) &= K(x, dy) \cdot \min(1, r(x, y)) & \forall y \neq x \\ K_\beta(x, \{x\}) &= 1 - \int_{y \neq x} K(x, dy) \cdot \min(1, r(x, y)) \end{aligned} \tag{1.4}$$

$$\text{où } r(x, y) := e^{-\beta(V(y)-V(x))} \frac{T(y, x)}{T(x, y)}.$$

Ce noyau paramétré par β laisse stable la mesure de Gibbs $\mu_\beta(dx) := \frac{1}{Z_\beta} e^{-\beta.V(x)} m(dx)$ (où Z_β est une constante de normalisation) ce qui lui donne une importance capitale dans l'algorithme du recuit simulé (voir la section 1.2.2 page 7) et de façon générale dans les applications liées aux modèles avec schéma de température (nombreuses dans ce manuscrit).

1.1.2 Noyau de Gibbs

Soit E un ensemble produit $E = E_1 \times \dots \times E_d$, et soit η une mesure sur E que l'on souhaite simuler. Soit $X = (X_1, \dots, X_d)$ une variable aléatoire sur E de loi η . Le principe de l'échantillonneur de Gibbs est que pour obtenir des réalisations de X , il suffit de pouvoir simuler les marginales conditionnelles de cette loi, soit, selon la notation bayésienne, les $p(x_p|x_{-p})$, où $\forall p \in \{1, \dots, d\}$,

$$x_{-p} := (x_1, \dots, x_{p-1}, x_{p+1}, \dots, x_d) \in E_1 \times \dots \times E_{p-1} \times E_{p+1} \times \dots \times E_d =: E_{-p}.$$

Plus précisément, si $p \in \{1, \dots, d\}$, la loi de (X_1, \dots, X_d) peut se décomposer conditionnellement à X_{-p} selon la formule :

$$p(x_1, \dots, x_d) = p(x_{-p}).p(x_p|x_{-p})$$

1.1. NOYAUX DE MARKOV CLASSIQUES

En notations probabilistes classiques, $p(x_{-p})$ désigne la loi de X_{-p} c'est-à-dire une mesure de probabilité η_{-p} sur E_{-p} définie pour toute partie A de E_{-p} par

$$\eta_{-p}(A) = \eta(A \times E_p).$$

L'objet $p(x_p|x_{-p})$ est lui formalisé par un noyau de Markov \mathcal{K}_p de E_{-p} dans E_p (noyau conditionnel de X_p par rapport à X_{-p}) et la relation s'écrit de façon complètement équivalente :

$$\eta = \eta_{-p} \otimes \mathcal{K}_p$$

Ce sont ces noyaux \mathcal{K}_p que l'on appelle marginales conditionnelles. Chacun de ces noyaux est canoniquement prolongeable à un noyau K_p de E dans E par la relation :

$$\forall(x, y) \in E^2, K_p(x, dy) := \delta_{x_{-p}}(dy_{-p}) \otimes \mathcal{K}_p(x_{-p}, dy_p)$$

autrement dit pour tout $i \neq p$, on prend $y_i = x_i$, puis pour y_p , on tire selon \mathcal{K}_p .

Le résultat essentiel est que chacun des K_p laisse invariant η (presque par définition, détails dans [6]), et par conséquent tout noyau de Markov défini comme composition déterministe ou aléatoire de ces noyaux laisse invariant η . Citons deux exemples notoires (voir [63]) :

- Balayage régulier (systematic-scan Gibbs sampler) : $K_{reg} := K_1 \dots K_d$, qui consiste à simuler les marginales conditionnelles dimension après dimension, de la première à la dernière dans l'ordre ;
- Balayage aléatoire (random-scan Gibbs sampler) : $K_{al} := K_I$ où I est une variable aléatoire discrète sur $\{1, \dots, d\}$.

1.1.3 Noyau de Metropolis-Hastings hybride

Toujours dans le cadre de la simulation d'une mesure η sur un espace produit $E = E_1 \times \dots \times E_d$, remarquons que l'échantillonneur de Gibbs présuppose de savoir simuler chacune des lois marginales conditionnelles, ce qui n'est pas toujours possible en pratique. L'algorithme de Metropolis-Hastings hybride propose de remplacer l'étape de simulation marginale conditionnelle par un noyau de Metropolis-Hastings unidimensionnel, conçu pour laisser invariant cette marginale conditionnelle. Plus précisément, supposons que les noyaux $\mathcal{K}_p(x_{-p}, dx_p)$ ne soient connus qu'à une constante près, typiquement :

$$\mathcal{K}_p(x_{-p}, dx_p) = \frac{1}{Z_p^{x_{-p}}} e^{-V_p^{x_{-p}}(x_p)} m(dx_p)$$

alors d'après la section 1.1.1 sur Metropolis-Hastings, tout noyau $\mathcal{M}_{MH,p}^{x_{-p}}(x_p, dx'_p)$ de E_p dans E_p faisant intervenir un noyau de proposition $\mathcal{M}_p^{x_{-p}}(x_p, dx'_p)$ et la procédure d'acceptation/rejet décrite en 1.1.1 vérifie, $\forall x_{-p} \in E_{-p}$:

$$\mathcal{K}_p(x_{-p}, dx_p) \cdot \mathcal{M}_{MH,p}^{x_{-p}} = \mathcal{K}_p(x_{-p}, dx_p)$$

1.2. MÉTHODES DE SIMULATION STOCHASTIQUES

soit $\mathcal{K}_p \cdot \mathcal{M}_{MH,p}^{x-p} = \mathcal{K}_p$, sans abus de notation. Tout comme \mathcal{K}_p , les $\mathcal{M}_{MH,p}^{x-p}$ sont représentables par un noyau M_p de E dans E défini par :

$$M_{MH,p}(x, dy) := \delta_{x-p}(dy_{-p}) \otimes \mathcal{M}_{MH,p}^{x-p}(x_p, dy_p)$$

(ce qui correspond à prendre $y_i = x_i$ pour tout $i \neq p$ et tirer y_p selon $\mathcal{M}_{MH,p}^{x-p}(x_p, dy_p)$), et la relation demeure : $K_p \cdot M_{MH,p} = K_p$. En prenant alors le résultat de la section 1.1.2 sur l'échantillonneur de Gibbs, à savoir $\eta \cdot K_p = \eta$, et en la multipliant à droite par $M_{MH,p}$, on obtient :

$$\eta \cdot \underbrace{K_p \cdot M_{MH,p}}_{K_p} = \eta \cdot M_{MH,p}$$

soit $\eta \cdot M_{MH,p} = \eta$. Ainsi, comme pour l'échantillonneur de Gibbs, tout noyau M_{MH} sous la forme d'une composition déterministe ou aléatoire de ces noyaux $M_{MH,p}$, c'est-à-dire qui réalise un Metropolis-Hastings dimension par dimension comme par exemple $M_{MH} := M_{MH,1} \dots M_{MH,d}$, laisse stable η .

1.2 Méthodes de simulation stochastiques

Nous introduisons dans cette section quelques-unes des méthodes de simulation stochastiques les plus classiques. Bien qu'elle ne puisse pas vraiment être qualifiée de "particulaire", la méthode MCMC (Monte Carlo Markov Chain) mérite d'être introduite pour son caractère fondamental, tant d'un point de vue historique qu'intellectuel.

1.2.1 Méthodes MCMC

Les algorithmes MCMC sont des techniques de simulation d'utilisation très courante dans de nombreux domaines d'application, notamment en statistique bayésienne (pour plus de détails, le lecteur pourra par exemple se référer à [48, 9] ainsi qu'à [3] pour les méthodes adaptatives). Le principe de base est simple : dans le but d'échantillonner une loi de probabilité η sur E , on simule une chaîne de Markov X_n de loi initiale η_0 choisie arbitrairement, et selon une transition K_η qui laisse la mesure η invariante. Les exemples de la section 1.1 illustrent bien que de tels noyaux peuvent être construits relativement facilement, et surtout sans avoir besoin de savoir simuler η directement.

Définie de cette façon, la chaîne X_n suit la loi $\eta_n = \eta_0 \cdot K_\eta^n$. Sous de bonnes hypothèses de contraction sur le noyau K_η , et compte tenu de la remarque

$$\eta_n - \eta = \eta_0 \cdot K_\eta^n - \underbrace{\eta}_{=\eta \cdot K_\eta^n} = (\eta_0 - \eta) \cdot K_\eta^n,$$

on est en droit d'espérer la convergence $\eta_n - \eta \xrightarrow[n \rightarrow \infty]{} 0$ en un sens à préciser. Autrement dit, la chaîne X_n suit asymptotiquement la loi cible η . De plus, sous de bonnes hypothèses de

1.2. MÉTHODES DE SIMULATION STOCHASTIQUES

mélange sur le noyau K_η , on peut démontrer que les *trajectoires* $(X_{k_0+1}, X_{k_0+2}, \dots, X_{k_0+n-1}, X_{k_0+n})$, considérées à partir d'un temps k_0 quelconque, vérifient le théorème ergodique

$$\eta_{k_0}^n := \left(\frac{1}{n} \sum_{k=k_0+1}^{k_0+n} \delta_{X_k} \right) \xrightarrow{n \rightarrow \infty} \eta.$$

Pour toute fonction bornée f , cette approximation peut notamment être utilisée pour approcher l'intégrale

$$\eta(f) = \int_E f(x) \eta(dx) \quad \text{par} \quad \eta_{k_0}^n(f) = \frac{1}{n} \sum_{k=k_0+1}^{k_0+n} f(X_k).$$

Le paramètre k_0 , appelé "temps de chauffe", est à choisir de façon à ce que le début de la trajectoire de X_n , lié à son initialisation, ne biaise pas la qualité de l'approximation.

Les méthodes MCMC sont en général analysées séparément selon le type de transition markovienne utilisée. La méthode la plus simple (et la plus ancienne) est probablement l'algorithme de Metropolis-Hastings (voir l'article historique [65]), qui réalise l'exploration au moyen des noyaux de Markov du même nom. L'échantillonneur de Gibbs, c'est-à-dire l'algorithme MCMC avec des transitions de Gibbs, est intuitivement adapté pour échantillonner une loi η sur un espace produit. Le lecteur pourra trouver dans [62] des résultats de convergence à vitesse géométrique de ces échantillonneurs sous des hypothèses de régularité peu restrictives.

Derrière ce formalisme simple se cachent plusieurs difficultés de mise en oeuvre pratique : la lenteur de la convergence, qui implique souvent des temps de calcul importants, l'estimation du temps de chauffe, l'élaboration d'une transition adéquate, c'est-à-dire adaptée à la géométrie de l'ensemble E et suffisamment mélangeante pour assurer la convergence, et de façon générale la garantie d'une bonne exploration de l'ensemble E par la chaîne X_n , notamment dans le cas d'une loi η multimodale. Ce dernier inconvénient est peut-être le plus important, et justifie en partie l'utilisation de méthodes à plusieurs particules : l'exploration de E se faisant par le mouvement d'une unique particule X_n , si η est multimodale, cette particule risque de se retrouver piégée dans un mode sans jamais visiter les autres, fournissant ainsi une estimation erronée.

1.2.2 Optimisation par recuit simulé

Fixons un potentiel borné $V : E \rightarrow \mathbb{R}$. Pour tout paramètre réel β , la mesure de Gibbs μ_β associée à V et β est définie par :

$$\mu_\beta(dx) := \frac{1}{Z_\beta} e^{-\beta \cdot V(x)} m(dx)$$

où m est une mesure de référence sur E et Z_β une constante de normalisation pour assurer une masse totale de 1. Lorsque le paramètre β tend vers $+\infty$, il est bien connu que la mesure μ_β se concentre sur les minimiseurs globaux du potentiel V sur E (voir par exemple

1.2. MÉTHODES DE SIMULATION STOCHASTIQUES

le lemme 6 page 40). Par conséquent, simuler les mesures μ_β pour des β grands permet de minimiser V . Or, la définition (1.4) page 4 fournit une famille de noyaux de Markov qui laissent ces mesures invariantes. L'idée du recuit simulé est donc de combiner la logique du MCMC avec les propriétés des mesures de Gibbs pour estimer le minimum global du potentiel V .

Plus précisément, pour tout paramètre β on se fixe un noyau K_β défini par (1.4) page 4. Fixons une suite croissante $(\beta_n)_{n \geq 1}$ qui tend vers $+\infty$. L'algorithme consiste à simuler une chaîne de Markov non homogène X_n dont les transitions sont définies par

$$\mathbb{P}(X_n \in dx | X_{n-1}) = K_{\beta_n}(X_{n-1}, dx).$$

Cette méthode peut être interprétée comme une méthode MCMC dans laquelle la loi cible varie petit à petit au cours de la simulation. Intuitivement, on parie sur le fait que si la suite β_n croît suffisamment lentement, la loi de la chaîne X_n est toujours proche de μ_{β_n} . Ainsi, β_n augmentant, cette loi charge logiquement les minima globaux de V ; autrement dit, X_n minimise V . Sous certaines hypothèses, notamment sur la suite β_n , les propriétés de mélange du noyau de proposition K et la géométrie du potentiel V , on peut démontrer (voir par exemple [64]) des résultats du type

$$\forall \varepsilon > 0, \quad \mathbb{P}(V(X_n) \geq V_{\min} + \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0.$$

Dans le cas général, de tels résultats sont obtenus sous l'hypothèse d'une croissance logarithmique de la suite β_n (voir [6, 35]), ce qui limite leur portée pratique. Dans son application, le recuit simulé demande très peu de connaissances (grosso modo savoir calculer $V(x)$ pour tout x) pour être appliqué, mais souffre des mêmes difficultés que le MCMC : temps de calcul importants, explorations de l'espace E problématiques, difficultés à pouvoir valider le résultat, et surtout risques de voir la chaîne X_n piégée dans un minimum local du potentiel V et ne jamais atteindre le minimum global.

1.2.3 Parallel Tempering Algorithm

Le parallel Tempering algorithm (PT) est une méthode particulière visant à améliorer le MCMC classique, en particulier le problème des modes secondaires. Soit η une mesure de probabilité sur E définie par une densité h (par rapport à une mesure de référence quelconque) calculable à une constante multiplicative près. Le principe est d'assister la chaîne de Markov de loi cible η par des sous-chaînes ayant pour lois invariantes des mesures η_k définies par des densités h_k plus "plates", ce au moyen d'échanges aléatoires. Formellement, soit $(\eta_k)_{1 \leq k \leq k_0}$ une suite finie de mesures dont la dernière, η_{k_0} , est la loi d'intérêt. Typiquement, si η_{n_f} est une loi a posteriori type $p(x|y)$, on peut prendre η_k proportionnelle à $p(x|y)^{\alpha_k}$, où α_k est une suite croissante allant de 0 à 1 (voir [8]). Supposons que l'on dispose pour chacune de ces mesures de probabilité η_k d'un noyau de Markov M_k laissant η_k invariante, soit tel que $\eta_k.M_k = \eta_k$ et vérifiant le théorème ergodique

$$\forall \pi \in \mathcal{P}(E), \quad \pi.M_k^n \xrightarrow[n \rightarrow \infty]{} \eta_k.$$

1.2. MÉTHODES DE SIMULATION STOCHASTIQUES

Le PT consiste à construire k_0 chaînes de Markov $(X_n^{(k)})_{n \geq 0}$, où $1 \leq k \leq k_0$, chacune réalisant ses transitions selon le noyau M_k .

Définies de cette façon, les chaînes $(X_n^{(k)})_n$ tendent à converger en loi vers η_k quand $n \rightarrow \infty$, en particulier $(X_n^{(k_0)})_n$, la chaîne d'intérêt principal, converge en loi vers la loi cible η , et jusqu'ici, nos k particules $X_n^{(k)}$ évoluent indépendamment les unes des autres. Puis à un certain instant n , déterministe ou aléatoire, on offre la possibilité d'un échange entre 2 chaînes, selon la procédure suivante : on tire deux chaînes voisines au hasard k et $k + 1$. Puis avec la probabilité

$$r = \min \left(1, \frac{h_k(X_n^{(k+1)})}{h_k(X_n^{(k)})} \frac{h_{k+1}(X_n^{(k)})}{h_{k+1}(X_n^{(k+1)})} \right)$$

on réalise l'échange $X_n^{(k)} \leftrightarrow X_n^{(k+1)}$.

L'idée sous-jacente est que les chaînes dites "chaudes", c'est-à-dire associées aux k petits, réalisent une exploration plus rapide mais moins sélective que les chaînes "froides", des grands k , qui se concentrent dans les différents modes des η_k . Les échanges entre chaînes sont donc ainsi pensés pour tempérer cette haute sélectivité des chaînes froides, en proposant des sauts pour les faire sortir des modes dans lesquels elles pourraient rester coincées en pratique.

1.2.4 Metropolis-Hastings hybride en interaction

Le Metropolis-Hastings hybride en interaction est un processus particulière dont le principe est de mêler l'évolution d'acceptation/rejet dimension par dimension du Metropolis-Hastings à une logique génétique de sélection/mutation/cross-over. Considérons donc, comme pour le Metropolis-Hastings hybride, une mesure cible η sur un ensemble produit $E = E_1 \times \dots \times E_d$, et pour chaque dimension $p \in \{1, \dots, d\}$ un noyau de Metropolis-Hastings $M_{MH,p}$ associé à un noyau de proposition \mathcal{M}_p^{x-p} . Remarquons que pour simplifier les notations, on peut représenter ces $\mathcal{M}_p^{x-p}(x_p, dy_p)$ de E_p dans E_p par les noyaux $\mathcal{M}_p(x, dy_p)$ prolongés de E tout entier dans E_p , avec la définition : $\forall x \in E$, décomposable en $x = (x_p, x_{-p})$ (sa composante selon E_p et les autres),

$$\mathcal{M}_p(x, dy_p) := \mathcal{M}_p^{x-p}(x_p, dy_p).$$

\mathcal{M}_p est donc le noyau qui à chaque $x \in E$ associe une proposition x_p^{prop} dans la dimension E_p .

On cherche à approcher η à l'aide d'un nuage dynamique de N particules ζ_n^k , $k \in \{1, \dots, N\}$, n servant ici d'indice temporel. Pour chacun des ζ_n^k vivant dans $E = E_1 \times \dots \times E_d$, on désigne par $\zeta_{n,p}^k$ sa composante selon E_p . L'itération $(\zeta_n^k)_{1 \leq k \leq N} \rightarrow (\zeta_{n+1}^k)_{1 \leq k \leq N}$ consiste à choisir une particule k et une composante p (par exemple choisies au hasard). Chaque particule ζ_n^i propose alors un candidat $\tilde{\zeta}_{n,p}^i \in E_p$ selon le noyau de proposition $\mathcal{M}_p(\zeta_n^i, dy_p)$, candidat associé à une probabilité d'acceptation α_i définie comme dans la section 1.1.1.

1.3. INTRODUCTION À L'ALGORITHME SMC

$\zeta_{n,p}^k$ est alors mis à jour en choisissant parmi l'ensemble de ces candidats, chaque $\tilde{\zeta}_{n,p}^i$ étant choisi avec probabilité $\frac{1}{N}\alpha_i$, ou en restant le même, avec probabilité $1 - \sum_{i=1}^N \frac{1}{N}\alpha_i$.

Ainsi défini, le processus $(\zeta_n^1, \dots, \zeta_n^N)_{n \geq 0}$ est markovien, et sa densité invariante est $\eta^{\otimes N}$ (démonstration dans [12]), ce qui signifie que l'algorithme fournit asymptotiquement en n un N -échantillon de la densité cible η . Autrement dit, la mesure d'occupation $\eta_n^N = \frac{1}{N} \sum_{k=0}^N \delta_{\zeta_n^k}$ admet pour n grand les mêmes propriétés d'approximation de η que celles des tirages i.i.d.

Le Metropolis-Hastings hybride en interaction a été développé pour accélérer la convergence vers la mesure cible, notamment dans le cadre de modèles de Markov cachés (voir [11]), l'algorithme donnant empiriquement de meilleurs résultats sur certains problèmes simulés.

1.3 Introduction à l'algorithme SMC

Les méthodes SMC (pour Sequential Monte Carlo) sont des méthodes d'échantillonnage particulières d'utilisation courante dans de nombreux domaines scientifiques (voir page 15 pour une introduction complète). Avant d'étudier ces algorithmes en détail de façon théorique dans le chapitre 2, nous en donnons ici les bases, notamment leur implémentation dans le cas général, leur interprétation en termes d'approximation d'un flot de mesures théoriques, et une de leurs applications les plus directes : le filtrage particulière.

1.3.1 Définition générale et implémentation

L'algorithme SMC (voir la section 2.1.3 page 24 pour une description formelle) consiste à simuler l'évolution aléatoire de N particules $\zeta_n^i \in E$ où $n \in \mathbb{N}$ désigne l'indice temporel et $i \in \{1, \dots, N\}$ l'indice de la particule. Cette évolution se fait par l'alternance de deux étapes de type génétique. Tout d'abord une étape de sélection, au moyen de potentiels $G_n : E \rightarrow \mathbb{R}$ strictement positifs bornés, où certaines particules disparaissent et d'autres se dupliquent. Puis une étape de mutation, au moyen de noyaux de Markov $M_n : E \rightarrow E$, où les particules sélectionnées évoluent indépendamment les unes des autres.

Fixons une mesure de probabilité initiale $\eta_0 \in \mathcal{P}(E)$, une suite (finie ou infinie) de potentiels $(G_n)_{n \geq 1}$ et une suite de noyaux de Markov $(M_n)_{n \geq 1}$. Il existe de nombreuses variantes d'implémentation de l'algorithme SMC. La plus simple et la plus classique est sans doute la suivante :

1.3. INTRODUCTION À L'ALGORITHME SMC

Initialisation : $\forall i \in \{1, \dots, N\}$, simuler $\zeta_0^i \sim \eta_0$.

Pour tout $n \geq 1$:

– SÉLECTION

1. Calculer les poids $w_i \leftarrow G_n(\zeta_{n-1}^i)$.
2. Normaliser $w_i \leftarrow \frac{w_i}{\sum_{1 \leq j \leq N} w_j}$.
3. $\forall i \in \{1, \dots, N\}$, simuler $\hat{\zeta}_{n-1}^i$ selon la loi multinomiale associée aux valeurs $(\zeta_{n-1}^j)_{j=1, \dots, N}$ et aux poids $(w_j)_{j=1, \dots, N}$, i.e. :

$$\forall i \in \{1, \dots, N\}, \quad \mathbb{P}(\hat{\zeta}_{n-1}^i = \zeta_{n-1}^j) = w_j.$$

– MUTATION

$\forall i \in \{1, \dots, N\}$, simuler $\zeta_n^i \sim M_n(\hat{\zeta}_{n-1}^i, \cdot)$.

Au sein de la sélection, l'étape 3, en particulier, se prête à de nombreuses variantes. Dans le cas où l'on sait borner les fonctions G_n , ou autrement dit que l'on connaît des $\varepsilon_n \geq 0$ tels que $\varepsilon_n G_n \leq 1$, on lui préférera la variante suivante :

3. Avec probabilité $\varepsilon_n G_n(\zeta_{n-1}^i)$, $\hat{\zeta}_{n-1}^i \leftarrow \zeta_{n-1}^i$ et avec probabilité $1 - \varepsilon_n G_n(\zeta_{n-1}^i)$, simuler $\hat{\zeta}_{n-1}^i$ selon la loi multinomiale associée aux valeurs $(\zeta_{n-1}^j)_{j=1, \dots, N}$ et aux poids $(w_j)_{j=1, \dots, N}$.

qui est plus stable algorithmiquement et plus générale (la première correspond au cas particulier $\varepsilon_n = 0$). Mis à part ces éventuelles quantités ε_n , l'algorithme nécessite trois prérequis :

$$\left\{ \begin{array}{l} \text{savoir simuler } \eta_0 ; \\ \text{savoir calculer } G_n(x), \quad \forall x \in E ; \\ \text{savoir simuler } M_n(x, \cdot), \quad \forall x \in E. \end{array} \right. \quad (1.5)$$

En sortie de cet algorithme, nous nous intéressons essentiellement à deux types de quantités.

– Les mesures d'occupation des particules $(\zeta_n^1, \dots, \zeta_n^N)$, ou mesures empiriques :

$$\eta_n^N(\cdot) := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\zeta_n^i}(\cdot)$$

– Les mesures empiriques dites "non normalisées" :

$$\gamma_n^N(\cdot) := \left(\prod_{p=1}^n \eta_{p-1}^N(G_p) \right) \times \eta_n^N(\cdot)$$

et en particulier la quantité $\gamma_n^N(1) = \prod_{p=1}^n \eta_{p-1}^N(G_p)$.

Ces mesures empiriques (la première est une mesure de probabilité, pas la seconde) sont les estimateurs naturels de mesures théoriques η_n et γ_n définies ci-après.

1.3.2 Mesures cibles, flots de Feynman-Kac

Si l'on se fixe une mesure de probabilité η_0 , des potentiels G_n et des transitions de Markov M_n , alors la suite de Feynman-Kac η_n (voir section 2.1.2 page 22 pour une description détaillée de ces modèles) associée à ces objets est la suite de mesures de probabilité sur E définie par la relation de récurrence

$$\eta_n = \psi_{G_n}(\eta_{n-1}) \cdot M_n \tag{1.6}$$

où ψ_{G_n} désigne la transformation de Boltzmann-Gibbs définie en (1.2) page 2 et le symbole \cdot désigne le produit mesure-noyau défini en (1.1) page 1. Intuitivement, on passe de η_{n-1} à η_n via premièrement une pondération par une fonction strictement positive bornée G_n , et deuxièmement un transport markovien M_n .

Ces mesures sont les mesures cibles de l'algorithme SMC, qui peut se voir comme une interprétation particulière de ces équations. Ainsi peut-on démontrer sous des hypothèses très faibles la convergence faible suivante :

$$\forall n \in \mathbb{N}, \quad \forall f \in \mathcal{B}(E), \quad \eta_n^N(f) \xrightarrow{N \rightarrow \infty} \eta_n(f).$$

De même les mesures empiriques "non normalisées" γ_n^N sont-elles des approximations des mesures non normalisées théoriques γ_n définies par l'initialisation

$$\gamma_0 = \eta_0 \quad \text{et la formule de récurrence} \quad \gamma_n(f) = \gamma_{n-1}(G_n \times M_n \cdot f).$$

On vérifie facilement que ces mesures vérifient

$$\eta_n(f) = \frac{\gamma_n(f)}{\gamma_n(1)} \quad \text{et} \quad \gamma_n(1) = \prod_{p=1}^n \eta_{p-1}(G_p),$$

ce qui rend tout à fait lisible la définition des γ_n^N .

Ces flots de mesures constituent une classe de modèles très large, apparaissant naturellement dans des domaines variés. En génétique des populations, naturellement, ainsi qu'en dynamique des populations, où les potentiels G_n quantifient la pression de sélection de l'environnement, et M_n les mouvements des individus. Mais aussi dans des applications a priori tout à fait étrangères à la logique génétique. Au sein du présent manuscrit, nous nous penchons longuement sur deux d'entre elles. Tout d'abord l'optimisation globale, via les flots de mesures de Gibbs, où le SMC s'interprète comme l'évolution de N algorithmes

1.3. INTRODUCTION À L'ALGORITHME SMC

de recuit simulé interagissant les uns avec les autres (voir la section 2.3 page 36). Ensuite l'inférence bayésienne : pour échantillonner une loi a posteriori η impossible à simuler directement, on construit artificiellement une suite de Feynman-Kac η_n passant progressivement d'une mesure η_0 facile à simuler à la mesure cible η . Dans ces modèles, la vraisemblance des données est alors quantifiée par la quantité $\gamma_n(1)$, d'où l'importance de son estimation. La partie 3.2 (voir notamment la section 3.2.3.1 page 71) et le chapitre 4 (voir notamment la section 4.3.2 page 96) illustrent ces techniques sur des problèmes inverses en électromagnétisme. La section suivante présente une quatrième application fondamentale des méthodes SMC.

1.3.3 L'exemple du filtrage particulaire

En statistiques bayésiennes, le filtrage consiste grossièrement à déterminer la loi d'une suite de variable d'intérêt X_n conditionnellement à des variables observées Y_n présentes (détermination de $p(x_n|y_n)$ ou de la prédiction $p(x_{n+1}|y_n)$) ou antérieures (détermination de $p(x_n|y_0 \dots y_n)$ ou de la prédiction $p(x_{n+1}|y_0 \dots y_n)$). L'un des modèles les plus simples et classiques pour formaliser cette problématique est le modèle de chaîne de Markov cachée (voir par exemple [14]). Il s'agit d'un modèle bayésien selon lequel la loi a priori sur les X_n est la loi définie par une loi initiale sur X_0 et des relations markoviennes données

$$\mathbb{P}(X_n \in dx | X_0, \dots, X_{n-1}) = \mathbb{P}(X_n \in dx | X_{n-1}) = M_n(X_{n-1}, dx).$$

Pour la vraisemblance des observations, on suppose que chaque variable Y_n ne dépend que de X_n selon une fonction de vraisemblance $g_n(X_n, Y_n)$ donnée, vérifiant

$$g_n(x_n, y_n) \propto p(y_n | x_n),$$

d'où finalement la représentation suivante des dépendances :

$$\begin{array}{ccccccccccc} X_0 & \xrightarrow{M_1} & X_1 & \xrightarrow{M_2} & X_2 & \xrightarrow{M_3} & \cdots & \xrightarrow{M_n} & X_n & \xrightarrow{M_{n+1}} & \cdots \\ \downarrow & & \downarrow & & \downarrow & & & & \downarrow & & \cdots \\ Y_0 & & Y_1 & & Y_2 & & \cdots & & Y_n & & \cdots \end{array}$$

Cette situation étant posée, le constat suivant est central : si l'on fixe une suite d'observations $(y_0, y_1, \dots, y_n, \dots)$ et que l'on pose pour tout n les potentiels

$$G_{n+1}(x) := g_n(x, y_n),$$

alors

$$\text{Loi}(X_n | Y_0 = y_0, \dots, Y_{n-1} = y_{n-1}) = \eta_n$$

et

$$\text{Loi}(X_n | Y_0 = y_0, \dots, Y_n = y_n) = \psi_{G_{n+1}}(\eta_n),$$

où η_n est la suite de Feynman-Kac associée aux potentiels G_n et aux noyaux M_n (voir par exemple [27]). Par conséquent, l'algorithme SMC permet d'échantillonner ces lois conditionnelles à condition de savoir simuler les noyaux M_n et de savoir calculer les $p(y_n | x_n)$ à

1.3. INTRODUCTION À L'ALGORITHME SMC

une constante multiplicative près. Les mutations jouent le rôle de prédiction, en utilisant le modèle a priori, et les sélections le rôle de correction, c'est-à-dire de prise en compte des observations.

Dans ce contexte, la quantité $\gamma_n(1)$ coïncide avec la vraisemblance des observations $p(y_0, \dots, y_n)$. On comprend l'importance de cette quantité en sélection de modèle : si les quantités du modèle (fonctions de vraisemblance, dynamique markovienne, etc.) dépendent d'un paramètre θ , $\gamma_n^{(\theta)}(1)$ fournit $p(y_0, \dots, y_n \mid \theta)$, soit intuitivement la vraisemblance de ce modèle.

Chapitre 2

Analyse théorique

Introduction

Les méthodes SMC (pour Sequential Monte Carlo), également appelées Quantum Monte Carlo, Diffusion Monte Carlo ou méthodes particulières de Feynman-Kac (on abrégera FK), sont des algorithmes stochastiques permettant l'échantillonnage de lois de probabilité complexes, éventuellement sur des espaces de dimension élevée. Elles sont d'utilisation courante en physique numérique [10, 4, 50] pour calculer des niveaux d'énergie dans les systèmes moléculaires. Elles sont en outre utilisées en statistiques, en traitement du signal et en sciences de l'information [14, 27, 29, 32] pour estimer les lois a posteriori de signaux ou paramètres qui sont observés partiellement. Au sein de la littérature des algorithmes évolutionnaires, ces méthodes de Monte Carlo sont utilisées de façon naturelle comme algorithmes d'exploration pour résoudre des problèmes d'optimisation. D'un point de vue purement mathématique, il s'agit d'un système de particules en interaction que l'on peut voir comme une interprétation des équations de FK. En plus des concepts présentés dans cette thèse, le lecteur pourra trouver des détails sur les formules de FK dans la monographie [28], et dans les références qu'elle suggère. Le principe (voir aussi [29] et ses références) est d'approcher une suite de mesures de probabilité cible $(\eta_m)_n$ au moyen d'un large nuage de particules au comportement aléatoire. Au départ de l'algorithme, le nuage est initialisé par N tirages indépendants selon la première loi η_0 . S'enclenche ensuite une alternance entre deux types d'opération : un schéma d'acceptation-rejet enrichi d'un mécanisme de recyclage de type sélectif, et une exploration libre de l'espace d'état.

Lors de l'étape de recyclage, le nuage de particules courant est modifié par des duplications et éliminations aléatoires des particules, selon un protocole similaire aux étapes de sélection des modèles en génétique des populations. Lors de l'étape d'exploration, les particules bougent indépendamment les unes des autres selon un noyau de Markov commun (étape de mutation).

Etant donné leur structure séquentielle intrinsèque, les algorithmes SMC trouvent naturellement application dans la résolution de problèmes séquentiels, comme le filtrage (voir par exemple [14, 43, 27] ou la section 1.3.3), mais pas uniquement. Dans d'autres types

de problèmes, ces algorithmes s'avèrent efficaces pour échantillonner une mesure unique η trop complexe pour être simulée directement de façon classique. Dans ce contexte, l'idée centrale est de trouver une suite de mesures d'interpolation $(\eta_k)_{0 \leq k \leq n}$, dont la complexité augmente progressivement, initialisée sur une mesure de proba η_0 facile à simuler et finissant sur la loi cible $\eta_n = \eta$. Les mesures consécutives η_k et η_{k+1} doivent être proches, assez, par exemple, pour qu'il soit envisageable de faire de l'échantillonnage d'importance, ou de l'échantillonnage par acceptation-rejet l'une par rapport à l'autre. L'aspect séquentiel de l'approche n'est alors qu'un moyen artificiel d'introduire la difficulté de simulation progressivement. A ce titre, l'exemple des modèles de recuit est fondamental, car il propose une interpolation dans l'espace des mesures à la fois simple, intuitive et universelle, basée sur la variation progressive d'un unique paramètre dit de température. Du point de vue de l'efficacité algorithmique, l'utilisation d'une grande population de particules se justifie par son avantage de pouvoir couvrir plusieurs modes simultanément. A l'inverse, les méthodes MCMC standards sont plus susceptibles de se voir piégées dans des modes locaux. Les algorithmes SMC ont donc été utilisés avec succès dans plusieurs domaines d'application, comme la simulation d'événements rares (voir [17]), l'optimisation stochastique et plus généralement l'échantillonnage des mesures de Gibbs (voir [29]).

Jusqu'à présent, les algorithmes SMC ont été essentiellement étudiés au moyen de techniques asymptotiques (c'est-à-dire lorsque le nombre de particules N tend vers $+\infty$), notamment via des décompositions des biais et variance, TLC et principes de grandes déviations (voir par exemple [26, 31], [32, 34, 36],[56], [20], [27], [14] et [28] pour un état de l'art).

Des théorèmes non asymptotiques ont été obtenus récemment ([18, 38]), mais aucun d'entre eux ne s'applique à l'analyse des modèles de FK adaptatifs et des modèles de recuit. Or ces modèles, et les algorithmes SMC non homogènes qui leur sont associés sont d'utilisation courante en physique numérique et en sciences de l'ingénierie (voir par exemple [7, 69], [22, 41, 67, 75], [52, 76] et la partie appliquée du présent manuscrit). Par manque d'estimées non asymptotiques, ces algorithmes particuliers sont utilisés comme heuristiques efficaces mais sans garantie théorique.

La principale contribution de notre travail est d'analyser ces deux classes de modèles. Notre approche est basée sur des techniques de semigroupe ainsi que sur une analyse de perturbation relativement inédite, impliquant entre autres des estimées uniformes en temps.

Plus précisément, dans le cas des modèles de type recuit, nous estimons explicitement les propriétés de stabilité du semigroupe de FK en fonction du coefficient de Dobrushin des noyaux de Markov de référence et de l'oscillation des potentiels de sélection. Ces techniques sont combinées avec des théorèmes non asymptotiques de bornes L^p ([36]) et des inégalités de concentration ([39, 37]). Puis nous explicitons des stratégies de réglage des paramètres de l'algorithme qui permettent d'assurer certaines inégalités de concentration uniformes en temps. Ces résultats s'appliquent en particulier aux modèles non homogènes de FK de type recuit, dans lesquels la suite de mesures η_n est associée à un paramètre décroissant de température. Mentionnons ici que d'autres approches indépendantes, comme celles de Whiteley ([81]) et Schweizer ([77]), sont basées, entre autres, sur des conditions de drift,

de trou spectral et des décompositions non asymptotiques du biais et de la variance. Ces approches permettent d'obtenir des résultats applicables au cas d'espaces d'état non compacts. A notre connaissance, ces techniques sont néanmoins restreintes aux théorèmes sur la variance non asymptotique, et ne peuvent être utilisées, en tout cas directement, pour déduire des inégalités de concentration exponentielles uniformes. Il paraît également difficile, en tout cas clairement non immédiat, d'étendre ces approches à l'analyse des modèles SMC adaptatifs étudiés dans ce manuscrit. Pour aborder cette question, nous développons une technique de perturbation de semigroupes de FK stochastiques : le schéma particulier adaptatif proposé est basé sur des potentiels de sélection aléatoires, dépendant d'un "schéma de refroidissement" adapté à la variabilité de la population aléatoire.

Du reste, cette partie théorique du manuscrit est organisée de la façon suivante. Dans une section préliminaire, nous introduisons les notions essentielles liées aux coefficients de Dobrushin et semigroupes de FK. Nous y soulignons en outre les résultats non asymptotiques importants que nous utilisons dans la suite. La section 2.2 se concentre sur l'analyse de la stabilité des semigroupes de FK dans le cas général. Nous y établissons des estimées pour le SMC uniformes en temps, de type bornes L^p et inégalités de concentration. Dans la section 2.3 nous appliquons ces résultats dans le cas particulier des modèles de Boltzmann-Gibbs, associés aux schémas de température. Dans ce contexte, les algorithmes SMC peuvent être interprétés comme une population de recuits simulés en interaction (on abrégera ISA pour Interacting Simulated Annealing). Nous élaborons une façon de régler le nombre d'itérations MCMC à chaque mutation en fonction du schéma de température, et cela de façon à garantir des bornes uniformes en temps. Enfin, dans la section 2.4, nous proposons une méthode SMC alternative pour l'échantillonnage de mesures de Gibbs, basée sur une stratégie adaptative pour définir au fur et à mesure la valeur des variations de température. Nous y apportons une étude théorique non asymptotique, basée sur une analyse de perturbation, et débouchant sur des estimées de type bornes L^p et inégalités de concentration.

Présentation des résultats principaux

Les algorithmes particuliers de FK consistent à faire évoluer un système de particules en interaction $(\zeta_n)_n = (\zeta_n^1, \dots, \zeta_n^N)_n$ de taille N , sur un espace d'état donné E . Cette évolution se met sous la forme d'une alternance entre deux types de transitions génétiques : une étape de sélection, faisant intervenir une fonction de sélection positive G_n , et une étape de mutation, où les particules sélectionnées évoluent aléatoirement selon une transition de Markov M_n donnée (une description plus formelle de ces algorithmes SMC est fournie section 2.1.3 page 24 puis section 2.4.2 page 44 pour une variante adaptative). Dans ce contexte, les mesures d'occupation $\eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\zeta_n^i}$ sont des approximations d'une suite de mesures de probabilité η_n définies par les formules récursives de FK :

$$\eta_n(f) = \frac{\eta_{n-1}(G_n \times M_n \cdot f)}{\eta_{n-1}(G_n)},$$

pour toute fonction f mesurable et bornée sur E (une discussion détaillée sur ces équations d'évolution est donnée section 2.1.2.1).

Pour pouvoir décrire avec une certaine pertinence nos principaux résultats, nous introduisons le couple de paramètres (g_n, b_n) défini ci-dessous.

$$g_n := \sup_{x, y \in E} \frac{G_n(x)}{G_n(y)} \quad \text{et} \quad b_n = \beta(M_n) := \sup_{\substack{x, y \in E \\ A \subset E}} |M_n(x, A) - M_n(y, A)|$$

La quantité $\beta(M_n)$ est appelée coefficient de Dobrushin de la transition de Markov M_n . Un de nos premiers résultats peut être grossièrement formulé comme suit.

Théorème 1. *On suppose*

$$\sup_{p \geq 1} g_p \leq M \quad \text{et} \quad \sup_{p \geq 1} b_p \leq \frac{a}{M(1+a)}$$

pour une certaine constante finie $M < \infty$ et un certain $a \in]0, 1[$. Sous cette hypothèse, pour tout $n \geq 0$, $N \geq 1$, $y \geq 0$ et $f \in \mathcal{B}_1(E)$, la probabilité de l'événement

$$|\eta_n^N(f) - \eta_n(f)| \leq \frac{r_1^* N + r_2^* y}{N^2}$$

est supérieure à $1 - e^{-y}$, où r_1^ et r_2^* sont des constantes définies explicitement en fonction de (a, M) .*

Dans la section 2.2.2, sous les mêmes hypothèses que le théorème 1, on prouve également des bornes L^p uniformes, ainsi que des inégalités de concentration pour les modèles particuliers non normalisés. Nous étendons aussi l'analyse à la situation où $g_n \xrightarrow{n \rightarrow +\infty} 1$.

Ces conditions de régularité sur b_n peuvent paraître difficiles à vérifier dans la mesure où les noyaux de Markov sont souvent imposés par l'application concernée. Remarquons cependant que c'est un problème que l'on peut résoudre à partir du moment où l'on sait simuler un noyau de Markov K_n vérifiant $\eta_n \cdot K_n = \eta_n$. En effet, pour stabiliser le système, le concepteur de l'algorithme SMC peut rajouter artificiellement plusieurs évolutions MCMC selon K_n à la suite de l'étape de mutation selon M_n . D'un point de vue formel, la suite de mesures cible $(\eta_n)_n$ est aussi solution des équations de FK associées aux noyaux de Markov $M'_n = M_n \cdot K_n^{m_n}$, où les nombre d'itérations m_n peuvent être choisis librement. Ce système est clairement plus stable dans la mesure où les quantités b'_n correspondantes vérifient

$$b'_n = \beta(M'_n) \leq b_n \cdot \beta(K_n^{m_n}) \leq b_n \cdot \beta(K_n)^{m_n}. \quad (2.1)$$

Dans ces cas de figure, le théorème 1 et ses extensions fournissent des conditions suffisantes sur les nombres d'itérations m_n pour assurer la convergence et les propriétés de stabilité de l'algorithme.

Ces résultats s'appliquent dans le cas des problèmes d'optimisation stochastique. Soit $V : E \rightarrow \mathbb{R}$ un potentiel borné, β_n une suite positive qui tend vers l'infini et m une mesure de référence quelconque sur E . Il est bien connu que la suite des mesures de Gibbs

$$\eta_n(dx) \propto e^{-\beta_n \cdot V(x)} m(dx)$$

se concentre sur les minima globaux de V (au sens du $\text{essinf}(V)$ selon m). Une remarque centrale est que ces mesures peuvent être vues comme un flot de FK, entre autres comme le flot associé aux potentiels $G_n = e^{-(\beta_n - \beta_{n-1}) \cdot V}$ et aux noyaux de Markov $M_n = K_{\beta_n}^{m_n k_0}$ où K_{β_n} est un noyau de type recuit simulé (voir la définition (1.4) page 4 et la section 2.3.2 page 39) et (m_n, k_0) sont des paramètres d'itération donnés. Dans la suite de cette section, K désigne le noyau de proposition du noyau du recuit simulé K_β . Dans ce contexte, les méthodes SMC peuvent être employées pour minimiser V . Les conditions sur b_n et g_n sont, elles, réinterprétées comme des conditions sur le schéma de température β_n et sur le nombre d'itérations de MCMC m_n . De plus, en combinant nos résultats avec des propriétés standards des mesures de Gibbs, nous obtenons des résultats de convergence exprimés directement en termes de performance d'optimisation. Sous ces notations, notre second résultat principal peut être formulé comme suit.

Théorème 2. *Soit $a \in]0, 1[$ fixé. On suppose que pour tout $x \in E$, la mesure $K^{k_0}(x, \cdot)$ vérifie la minoration suivante :*

$$K^{k_0}(x, \cdot) \geq \delta \nu(\cdot),$$

et cela pour une certaine mesure ν sur E , un certain $\delta > 0$ et un certain $k_0 \geq 1$ indépendants de x . On suppose également que le schéma de température β_p et les nombres d'itérations m_p vérifient les conditions suivantes :

$$\sup_{p \geq 1} \Delta_p \leq \Delta \quad \text{et} \quad m_p \geq \frac{\log\left(\frac{e^{\Delta \cdot \text{osc}(V)}(1+a)}{a}\right) e^{\text{osc}(V) \cdot \beta_p}}{\delta}$$

pour une certaine constante Δ . Pour tout $\varepsilon > 0$, on note $p_n^N(\varepsilon)$ la proportion des particules (ζ_n^i) telles que $V(\zeta_n^i) \geq V_{\min} + \varepsilon$. Ainsi, pour tout $n \geq 0$, $N \geq 1$, $y \geq 0$ et pour tout $\varepsilon' < \varepsilon$, la probabilité de l'événement

$$p_n^N(\varepsilon) \leq \frac{e^{-\beta_n(\varepsilon-\varepsilon')}}{m_{\varepsilon'}} + \frac{r_1^*N + r_2^*y}{N^2}$$

est supérieure à $1 - e^{-y}$, avec $m_{\varepsilon'} = m(V \leq V_{\min} + \varepsilon')$, (r_1^*, r_2^*) désignant les mêmes constantes que dans le théorème 1 (avec $M = e^{\Delta \cdot \text{osc}(V)}$).

Il est intéressant de comparer les estimées fournies par le précédent théorème avec l'analyse des performances du recuit simulé traditionnel (on abrégera SA, pour Simulated Annealing). Tout d'abord, l'essentiel de la littérature sur les modèles SA se focalise sur une convergence faible de la loi de la variable d'état. Lorsque la température initiale du schéma est supérieure à une certaine valeur critique, et sous réserve d'un schéma de refroidissement logarithmique, il est bien connu que la probabilité que la variable d'état se trouve dans un des extrema globaux tend vers 1 quand le paramètre temporel tend vers $+\infty$. Le schéma de refroidissement présenté dans le théorème 2 est lui aussi logarithmique. Mais à la différence du modèle SA, le théorème 2 permet de quantifier la performance de l'algorithme SMC par des inégalités de concentration uniformes en temps, qui ne dépendent par ailleurs d'aucun paramètre critique.

En pratique, le choix de la suite d'incrément $\Delta_n = (\beta_n - \beta_{n-1})$ à l'avance pose des problèmes de mise en oeuvre informatique. Une façon de les résoudre consiste à élaborer des stratégies adaptatives, c'est-à-dire de faire dépendre l'incrément Δ_n du passé de l'algorithme, et en particulier du nuage de particules courant ζ_{n-1} . Ces stratégies sont d'utilisation courante dans l'ingénierie et la recherche appliquée (voir par exemple [52, 76], [22, 41, 67]). Dans cette logique, nous proposons d'étudier le cas où l'incrément Δ_n^N (noté ainsi entre autres pour le démarquer du cas classique) est choisi de telle sorte que

$$\eta_{n-1}^N(e^{-\Delta_n^N \cdot V}) = \varepsilon,$$

où $\varepsilon > 0$ est une constante donnée (voir la section 2.4.2 pour une description détaillée de l'algorithme). D'un point de vue numérique, ε peut se voir comme l'espérance de la proportion des particules qui ne sont pas concernées par le mécanisme de recyclage dans l'étape de sélection. Nous interprétons ce processus particulière comme une perturbation d'une certaine suite FK théorique η_n associée à un schéma de température théorique β_n . Notre principal résultat consiste en cette estimation de l'erreur L^p :

Théorème 3. *Pour tout $p \geq 1$, $n \geq 0$, $N \geq 1$ et toute fonction f bornée par 1, on a*

$$\mathbb{E} \left(\left| \eta_n^N(f) - \eta_n(f) \right|^p \right)^{1/p} \leq \frac{B_p}{\sqrt{N}} \sum_{k=0}^n \prod_{i=k+1}^n (b_i g_i (1 + c_i)),$$

2.1. PRÉLIMINAIRES

avec $c_n = \frac{V_{\max} e^{\Delta_n V_{\max}}}{\varepsilon \cdot \eta_{n-1}(V)}$, $\Delta_n = \beta_n - \beta_{n-1}$ et B_p définis ci-dessous.

$$B_{2p}^{2p} = \frac{(2p)!}{2^p \cdot p!} \quad ; \quad B_{2p+1}^{2p+1} = \frac{(2p+1)!}{2^p \cdot p! \sqrt{2p+1}}. \quad (2.2)$$

Sous des hypothèses de régularité appropriées sur les paramètres b_n, g_n, c_n , nous verrons par la suite que ces estimées L^p impliquent également des inégalités de concentration uniformes. Les preuves des théorèmes 1, 2, et 3 sont détaillées respectivement dans les sections 2.2.2, 2.3.2 et 2.4.4.

2.1 Préliminaires

Comme pour l'essentiel du chapitre introductif 1, on se fixe un espace métrique complet séparable (E, r) dont la tribu borélienne associée est notée \mathcal{E} . On note $\mathcal{P}(E)$ l'ensemble des mesures de probabilité sur E , $\mathcal{B}(E)$ l'espace des fonctions de E dans \mathbb{R} mesurables bornées et $\mathcal{B}_1(E) \subset \mathcal{B}(E)$ le sous-ensemble des fonctions bornées par 1. En plus des objets et opérations définis page 1, on pose les notations suivantes :

Pour tout $f \in \mathcal{B}(E)$, on note $\|f\|_\infty = \sup_{x \in E} |f(x)|$ et $osc(f) = (\max(f) - \min(f))$.

On note $\mathcal{O}_1(E) \subset \mathcal{B}(E)$ l'ensemble des fonctions $f \in \mathcal{B}(E)$ telles que $osc(f) \leq 1$.

Pour toute variable aléatoire $X : \Omega \rightarrow \mathbb{R}$ définie sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, et tout $p \geq 1$, $\|X\|_p$ désigne la norme L^p de X , soit $\mathbb{E}(|X|^p)^{1/p}$.

On note $\mathcal{P}_\Omega(E)$ l'ensemble des mesures de probabilité aléatoires sur E .

Enfin, pour tout $p \geq 1$, on note d_p la distance sur l'espace $\mathcal{P}_\Omega(E)$ définie pour toutes mesures aléatoires $\hat{\mu}, \hat{\nu}$ par

$$d_p(\hat{\mu}, \hat{\nu}) = \sup_{f \in \mathcal{O}_1(E)} \|\hat{\mu}(f) - \hat{\nu}(f)\|_p. \quad (2.3)$$

2.1.1 Coefficients de Dobrushin

Rappelons ici quelques définitions et propriétés simples indispensables pour la suite.

Définition 1. Soit $\mu, \nu \in \mathcal{P}(E)$. La distance en variation totale entre μ et ν est définie par

$$\|\mu - \nu\|_{tv} = \sup\{|\mu(A) - \nu(A)|; A \in \mathcal{E}\}.$$

Définition 2. A tout noyau de Markov K sur E est associé son coefficient de Dobrushin $\beta(K) \in [0, 1]$ défini par

$$\beta(K) = \sup\{K(x, A) - K(y, A); x, y \in E, A \in \mathcal{E}\}$$

ou de manière équivalente :

$$\beta(K) = \sup\left\{\frac{\|\mu.K - \nu.K\|_{tv}}{\|\mu - \nu\|_{tv}}; \mu, \nu \in \mathcal{P}(E), \mu \neq \nu\right\}.$$

Le paramètre $\beta(K)$ caractérise les propriétés de mélange du noyau de Markov K . On remarque facilement que la fonction β est une norme d'opérateur, au sens où $\beta(K_1.K_2) \leq \beta(K_1).\beta(K_2)$, cela pour tout couple de noyaux de Markov K_1, K_2 . Par définition, pour toutes mesures $\mu, \nu \in \mathcal{P}(E)$ et tout noyau de Markov K , on a $\|\mu.K - \nu.K\|_{tv} \leq \beta(K) \cdot \|\mu - \nu\|_{tv}$. Par ailleurs, pour toute fonction $f \in \mathcal{B}(E)$,

$$\text{osc}(K.f) \leq \beta(K) \cdot \text{osc}(f). \quad (2.4)$$

On pourra trouver plus de détails sur ces coefficients ergodiques dans la monographie [28], comme par exemple le lemme suivant que l'on rappelle pour son utilité ultérieure.

Lemme 1. *Soit $\mu, \nu \in \mathcal{P}(E)$ et G une fonction positive bornée sur E vérifiant $\sup_{x,y \in E} \frac{G(x)}{G(y)} \leq g$, pour une certaine constante $g \geq 0$. Dans cette situation, on a*

$$\|\Psi_G(\mu) - \Psi_G(\nu)\|_{tv} \leq g \cdot \|\mu - \nu\|_{tv}.$$

2.1.2 Modèles de Feynman-Kac

Nous rappelons ici les outils standards liés aux modèles de FK. Ces modèles fournissent les outils théoriques nécessaires à la fois pour la formalisation et pour l'analyse des algorithmes SMC (on pourra trouver de nombreux détails supplémentaires dans [32, 39, 36]).

2.1.2.1 Equations d'évolution

On considère une suite de mesures de probabilité $(\eta_n)_n$, définie par une mesure initiale η_0 et par la relation de récurrence

$$\forall f \in \mathcal{B}(E), \quad \eta_n(f) = \frac{\eta_{n-1}(G_n \times M_n.f)}{\eta_{n-1}(G_n)} \quad (2.5)$$

où les $G_n \in \mathcal{B}(E)$ sont des fonctions strictement positives données et les M_n des noyaux de Markov avec $M_n(x, \cdot) \in \mathcal{P}(E)$ et $M_n(\cdot, A) \in \mathcal{B}_1(E)$. Il s'agit de la suite de mesures que l'on veut essentiellement approcher par le biais des méthodes SMC. De façon équivalente, $(\eta_n)_n$ peut être définie par la relation

$$\eta_n = \phi_n(\eta_{n-1})$$

où $\phi_n : \mathcal{P}(E) \rightarrow \mathcal{P}(E)$ est la transformation de FK associée au potentiel G_n et au noyau de Markov M_n , définie par

$$\phi_n(\eta_{n-1}) = \psi_{G_n}(\eta_{n-1}).M_n$$

2.1. PRÉLIMINAIRES

avec

$$\psi_{G_n}(\eta_{n-1})(dx) := \frac{1}{\eta_{n-1}(G_n)} G_n(x) \eta_{n-1}(dx).$$

La formule qui suit fournit une interprétation de la transformation de Boltzmann-Gibbs en termes de transport markovien non linéaire

$$\Psi_{G_n}(\eta_{n-1})(dy) = \left(\eta_{n-1} S_{n,\eta_{n-1}} \right) (dy) := \int \eta_{n-1}(dx) S_{n,\eta_{n-1}}(x, dy)$$

où le noyau de Markov S_{n,η_n} est défini par

$$S_{n,\eta_{n-1}}(x, dy) = \varepsilon_n \cdot G_n(x) \delta_x(dy) + (1 - \varepsilon_n \cdot G_n(x)) \Psi_{G_n}(\eta_{n-1})(dy),$$

(pour toute constante $\varepsilon_n > 0$ vérifiant $\varepsilon_n \cdot G_n \leq 1$). Cela implique

$$\eta_n = \eta_{n-1} K_{n,\eta_{n-1}} \quad \text{avec} \quad K_{n,\eta_{n-1}} = S_{n,\eta_{n-1}} M_n \quad (2.6)$$

De cette façon, η_n peut être interprétée comme la loi des états aléatoires \bar{X}_n d'une chaîne de Markov dont les transitions markoviennes

$$\mathbb{P}(\bar{X}_{n+1} \in dy \mid \bar{X}_n = x) := K_{n+1,\eta_n}(x, dy) \quad (2.7)$$

dépendent de la loi courante $\eta_n = \text{Law}(\bar{X}_n)$.

Rappelons enfin que les mesures η_n admettent les représentations fonctionnelles suivantes :

$$\eta_n(f) = \frac{\gamma_n(f)}{\gamma_n(1)}$$

(1 désigne ici la fonction constante) où les mesures de FK non normalisées γ_n sont définies par les formules

$$\gamma_0 = \eta_0 \quad ; \quad \gamma_n(f) = \gamma_{n-1}(G_n \times M_n \cdot f). \quad (2.8)$$

Si l'on compare cette définition avec (2.5), il est clair que la constante de normalisation $\gamma_n(1)$ vérifie

$$\gamma_n(1) = \prod_{p=1}^n \eta_{p-1}(G_p). \quad (2.9)$$

Cette quantité est d'une importance majeure dans de nombreux problèmes, comme en filtrage (voir section 1.3.3 page 13), en statistique bayésienne (un exemple est donné section 2.3.1 page 36) ou en thermodynamique (estimation de niveaux d'énergie, voir section 2.3.1 page 36).

2.1.2.2 Semigroupes de Feynman-Kac

Pour un flot de FK donné, c'est-à-dire des potentiels G_n , des noyaux M_n et les transformations ϕ_n associées, le semigroupe de FK est l'ensemble des transformations $\phi_{p,n}$, $p \leq n$ définies par

$$\phi_{p,n} := \phi_n \circ \phi_{n-1} \circ \cdots \circ \phi_{p+1}.$$

Une importante remarque est que chacune de ces transformations admet une structure comparable à celle de chacune des ϕ_k . Plus précisément, pour tout entier p , définissons l'opérateur intégral non normalisé Q_p :

$$\forall f \in \mathcal{B}(E), \quad Q_p \cdot f = G_p \cdot M_p \cdot f \quad (2.10)$$

ainsi que les opérateurs composés $Q_{p,n}$, définis par les équations à rebours

$$Q_{p,n} = Q_{p+1} \cdot (Q_{p+2} \dots Q_n) = Q_{p+1} \cdot Q_{p+1,n}. \quad (2.11)$$

Précisons la convention $Q_{n,n} = Id$ pour $p = n$. En comparant ces définitions avec (2.8), il est clair que $\gamma_n = \gamma_{n-1} \cdot Q_n$ et plus généralement, pour tout $p \leq n$,

$$\gamma_n = \gamma_p \cdot Q_{p,n}.$$

Le semigroupe $\phi_{p,n}$ peut être formulé en fonction des $Q_{p,n}$ au moyen des formules suivantes

$$\phi_{p,n}(\mu)(f) = \frac{\mu(Q_{p,n} \cdot f)}{\mu(Q_{p,n} \cdot 1)}$$

pour tout $f \in \mathcal{B}(E)$ et $\mu \in \mathcal{P}(E)$. Finalement, si l'on pose

$$P_{p,n} \cdot f = \frac{Q_{p,n} \cdot f}{Q_{p,n} \cdot 1} \quad \text{and} \quad G_{p,n} = Q_{p,n} \cdot 1$$

on obtient alors

$$\phi_{p,n}(\mu)(f) = \frac{\mu(G_{p,n} \cdot P_{p,n} \cdot f)}{\mu(G_{p,n})},$$

ou de façon équivalente $\phi_{p,n}(\mu) = \psi_{G_{p,n}}(\mu) \cdot P_{p,n}$.

2.1.3 L'algorithme SMC classique

L'idée centrale du SMC est d'approcher les mesures η_n en simulant un système de particules en interaction $(\zeta_n)_n = (\zeta_n^1, \dots, \zeta_n^N)_n$ de taille N de sorte que

$$\eta_n^N = \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\zeta_n^i} \xrightarrow{N \rightarrow \infty} \eta_n.$$

La problématique est évidemment de préciser et quantifier cette convergence. Du reste, le processus particulaire est défini de la façon suivante.

L'initialisation du nuage de particules se fait par N tirages indépendants $\zeta_0 = (\zeta_0^1, \dots, \zeta_0^N)$ de même loi η_0 . Puis on alterne deux transitions de type génétique.

Pendant la première étape, chaque particule ζ_n^i évolue en une nouvelle particule $\widehat{\zeta}_n^i$ choisie aléatoirement selon la loi

$$S_{n+1, \eta_n^N}(\zeta_n^i, dx) := \varepsilon_{n+1} \cdot G_{n+1}(\zeta_n^i) \delta_{\zeta_n^i}(dx) + (1 - \varepsilon_{n+1} \cdot G_{n+1}(\zeta_n^i)) \Psi_{G_{n+1}}(\eta_n^N)(dx)$$

2.1. PRÉLIMINAIRES

avec les mesures empiriques "corrigées" (ou pondérées)

$$\Psi_{G_{n+1}}(\eta_n^N) = \sum_{j=1}^N \frac{G_{n+1}(\zeta_n^j)}{\sum_{k=1}^N G_{n+1}(\zeta_n^k)} \delta_{\zeta_n^j}.$$

Cette transition peut être interprétée comme un schéma d'acceptation-rejet muni d'un mécanisme de recyclage. Dans la seconde étape, les particules sélectionnées $\widehat{\zeta}_n^i$ évoluent aléatoirement selon le noyau de Markov M_{n+1} . En d'autres termes, pour tout $1 \leq i \leq N$, on simule la variable aléatoire ζ_{n+1}^i avec la loi $M_{n+1}(\widehat{\zeta}_n^i, dx)$.

Au regard de (2.7), si l'on remplace η_n^N par η_n , alors ζ_n coïncide avec N copies indépendantes de la chaîne de Markov \overline{X}_n définie par (2.6). D'autre part, par la loi des grands nombres, on a $\eta_0^N \simeq \eta_0$ de sorte que

$$\eta_1^N \simeq \eta_0^N \cdot K_{1, \eta_0^N} \simeq \eta_0 \cdot K_{1, \eta_0} = \eta_1.$$

En itérant cette procédure d'approximation, la mesure empirique η_n^N est censée approcher η_n pour tout temps $n \geq 0$. En ce qui concerne les mesures non normalisées γ_n , on définit

$$\gamma_n^N(1) = \prod_{p=1}^n \eta_{p-1}^N(G_p)$$

(définition directement inspirée de (2.9)) et plus généralement $\gamma_n^N(f) = \eta_n^N(f) \times \prod_{p=1}^n \eta_{p-1}^N(G_p)$.

Il est démontré (voir par exemple [27]) que ces modèles particulières fournissent une estimation non biaisée des mesures non normalisées, c'est-à-dire qu'on a

$$\forall f \in \mathcal{B}(E), \quad \mathbb{E}(\gamma_n^N(f)) = \gamma_n(f).$$

En plus de l'analyse de la convergence des η_n^N nous considérerons également par la suite les propriétés de concentration des estimateurs non biaisés $\gamma_n^N(1)$ autour de leurs valeurs limites $\gamma_n(1)$.

2.1.4 Quelques résultats non asymptotiques

Pour quantifier les propriétés de stabilité du semigroupe de FK, il est naturel de définir les paramètres suivants.

Définition 3. *Pour tous entiers $p < n$, on pose*

$$b_n := \beta(M_n) \quad \text{et} \quad b_{p,n} := \beta(P_{p,n}).$$

$$g_n := \sup_{x,y \in E} \frac{G_n(x)}{G_n(y)} \quad \text{et} \quad g_{p,n} := \sup_{x,y \in E} \frac{G_{p,n}(x)}{G_{p,n}(y)}.$$

2.1. PRÉLIMINAIRES

Les quantités $g_{p,n}$, et respectivement $b_{p,n}$, représentent l'oscillation des potentiels $G_{p,n}$, et respectivement les propriétés de mélange des transitions markoviennes $P_{p,n}$ associées au semigrroupe de FK $\phi_{p,n}$ décrit dans (2.1.2.2). De nombreuses inégalités de contraction sur $\phi_{p,n}$ au sens de la norme en variation totale, ou différents types d'entropies relatives peuvent être obtenues en fonction de ces deux quantités (voir par exemple [28]). En décomposant l'erreur de la façon suivante :

$$\eta_n^N - \eta_n = \sum_{p=0}^n \phi_{p,n}(\eta_p^N) - \phi_{p,n}(\phi_p(\eta_{p-1}^N)), \quad (2.12)$$

on comprend alors qu'il est possible d'utiliser ces résultats de contraction pour analyser les performances de l'algorithme SMC. La logique de cette analyse, représentée sur le schéma de la figure 2.1, consiste grossièrement parlant à montrer que si le semigrroupe $\phi_{p,n}$ est contractant (par exemple selon un facteur $a^{-(n-p)}$ où $a < 1$), les erreurs d'échantillonnage successives entre η_p^N et $\phi_p(\eta_{p-1}^N)$ (d'ordre $N^{-1/2}$) sont compensées dans le temps par cette contraction, de façon à rendre (2.12) sommable. L'analyse de performance développée dans

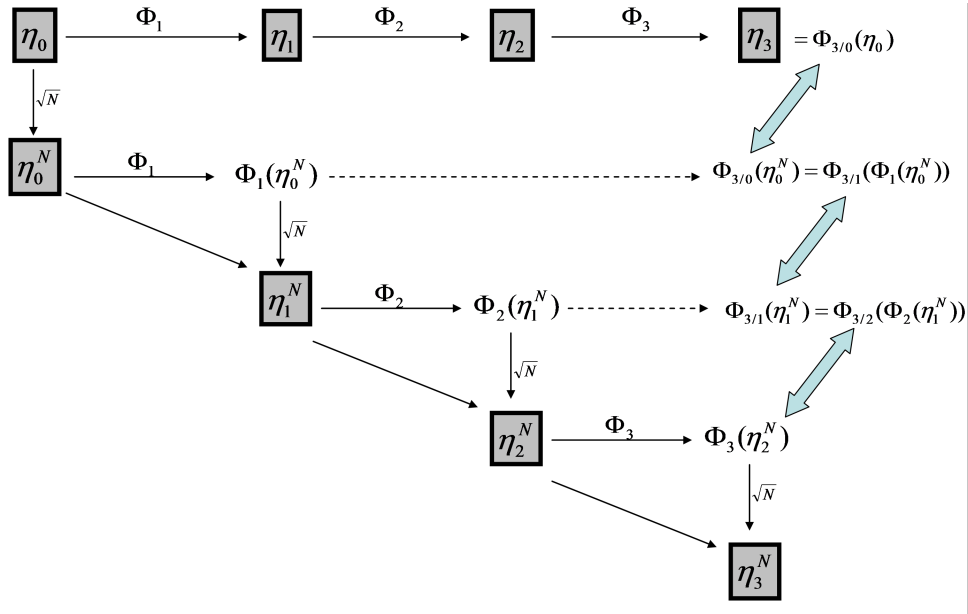


FIGURE 2.1 – Schéma d'analyse du SMC par stabilité du semigrroupe de Feynman-Kac

les sections 2.2.1 et 2.3.2 est en partie basée sur cette analyse, en particulier par le biais des trois inégalités non asymptotiques présentées ci-après.

Premièrement, la majoration de l'erreur L^p suivante, valable pour tout $f \in \mathcal{B}_1(E)$, est démontrée dans [36] :

$$\mathbb{E} \left(\left| \eta_n^N(f) - \eta_n(f) \right|^p \right)^{1/p} \leq \frac{B_p}{\sqrt{N}} \sum_{k=0}^n g_{k,n} b_{k,n} \quad (2.13)$$

2.2. ANALYSE NON ASYMPTOTIQUE GÉNÉRALE

où B_p sont les constantes introduites en (2.2) page 21.

Deuxièmement, les inégalités de concentration exponentielles suivantes sont obtenues dans [37].

Pour tout $f \in \mathcal{B}_1(E)$ et tout $\varepsilon > 0$ on a :

$$\frac{-1}{N} \log \mathbb{P} \left(\left| \eta_n^N(f) - \eta_n(f) \right| \geq \frac{r_n}{N} + \varepsilon \right) \geq \frac{\varepsilon^2}{2} \left[b_n^* \bar{\beta}_n + \frac{\sqrt{2}r_n}{\sqrt{N}} + \varepsilon \left(2r_n + \frac{b_n^*}{3} \right) \right]^{-1} \quad (2.14)$$

où $r_n, \bar{\beta}_n$ et b_n^* sont des constantes vérifiant :

$$\begin{cases} r_n & \leq 4 \sum_{p=0}^n g_{p,n}^3 b_{p,n} \\ \bar{\beta}_n^{-2} & \leq 4 \sum_{p=0}^n g_{p,n}^2 b_{p,n}^2 \\ b_n^* & \leq 2 \sup_{0 \leq p \leq n} g_{p,n} b_{p,n} \end{cases}$$

Troisièmement, l'inégalité de concentration suivante sur les modèles particuliers non normalisés γ_n^N est démontrée dans [39] (voir le théorème 6.5). Si l'on pose les fonctions

$$\begin{aligned} h_0 &:= x \mapsto 2(x + \sqrt{x}) & \text{et} \\ h_1 &:= x \mapsto \frac{x}{3} + \sqrt{2x}, \end{aligned} \quad (2.15)$$

alors, $\forall \varepsilon \in \{+1, -1\}$ et $\forall y \geq 0$ on a :

$$\mathbb{P} \left(\frac{\varepsilon}{n} \log \left(\frac{\gamma_n^N(1)}{\gamma_n(1)} \right) \geq \frac{\bar{r}(n)}{N} h_0(y) + \tau_n^* \bar{\sigma}_n^2 h_1 \left(\frac{y}{N \cdot \bar{\sigma}_n^2} \right) \right) \leq e^{-y} \quad (2.16)$$

où les quantités τ_n^* , $\bar{\sigma}_n^2$ et $\bar{r}(n)$ peuvent être estimées de la façon suivante :

– $\tau_n^* = \sup_{0 \leq q \leq n} \tau_{q,n}$, où $\tau_{q,n}$ vérifie

$$\tau_{q,n} \leq \frac{4}{n} \sum_{p=q}^{n-1} g_{q,p} \cdot g_{p+1} \cdot b_{q,p} ; \quad (2.17)$$

– $\bar{\sigma}_n^2 = \sum_{q=0}^{n-1} \sigma_q^2 \left(\frac{\tau_{q,p}}{\tau_n^*} \right)^2$ où σ_q vérifie $\sigma_q \leq 1$;

– $\bar{r}(n)$ vérifie

$$\bar{r}(n) \leq \frac{8}{n} \sum_{0 \leq q \leq p < n} g_{p+1} \cdot g_{q,p}^3 \cdot b_{q,p}. \quad (2.18)$$

2.2 Analyse non asymptotique générale

Les formules (2.13), (2.14) et (2.16) fournissent des estimées non asymptotiques en fonction des quantités $g_{p,n}$ et $b_{p,n}$, qui sont relatives à la stabilité du semigroupe. Ainsi formulées,

2.2. ANALYSE NON ASYMPTOTIQUE GÉNÉRALE

elles sont difficiles à appliquer directement pour décider des réglages des paramètres d'une méthode SMC. En effet, les seuls objets connus ou calculables sont généralement les noyaux de Markov de référence M_p et les fonctions de sélection élémentaires G_p . Il semble donc nécessaire d'estimer $g_{p,n}$ et $b_{p,n}$ de façon précise en fonction des g_p et b_p . Ce travail est réalisé dans la section 2.2.1. Dans la section 2.2.2, nous combinons ces estimées avec les résultats de concentration présentés dans la section 2.1.4 afin d'obtenir des estimées uniformes en temps.

2.2.1 Stabilité du semigroupe

Introduisons ici quelques lemmes techniques.

Lemme 2. *Soit K un noyau de Markov et G une fonction strictement positive sur E vérifiant $\sup_{x,y \in E} \frac{G(x)}{G(y)} \leq g$, pour une certaine constante positive finie g . Dans cette situation, on a*

$$\sup_{x,y \in E} \frac{K.G(x)}{K.G(y)} \leq 1 + \beta(K)(g - 1).$$

Démonstration. Soient $x, y \in E$ tels que $K.G(x) \geq K.G(y)$. On écrit :

$$\frac{K.G(x)}{K.G(y)} = \frac{K.G(x) - K.G(y)}{K.G(y)} + 1 \leq \frac{\beta(K)(G_{\max} - G_{\min})}{G_{\min}} + 1$$

La dernière inégalité relève du fait que

$$K.G(x) - K.G(y) \leq \text{osc}(K.G) \leq \beta(K) \cdot \text{osc}(G) = \beta(K) \cdot (G_{\max} - G_{\min}).$$

D'autre part, on a $K.G(y) = \int_u G(u)K(y, du) \geq G_{\min}$. On conclut alors par passage à la borne supérieure sur tous les couples $(x, y) \in E^2$. En effet, la quantité $\frac{\beta(K)(G_{\max} - G_{\min})}{G_{\min}} + 1$ se trouve être égale à $1 + \beta(K)(g - 1)$. □

Lemme 3. *Soit M un noyau de Markov et Q un opérateur intégral (pas nécessairement normalisé) qui vérifie $\sup_{x,y \in E} \frac{Q.1(x)}{Q.1(y)} \leq g$, pour une certaine constante finie $g \geq 1$. Dans cette situation, le noyau P défini par*

$$P.f(x) := \frac{M.Q.f(x)}{M.Q.1(x)}$$

vérifie la propriété suivante

$$\beta(P) \leq g \cdot \beta(M) \cdot \beta(P'),$$

où P' est le noyau de Markov défini par $P'.f(x) := \frac{Q.f(x)}{Q.1(x)}$.

2.2. ANALYSE NON ASYMPTOTIQUE GÉNÉRALE

Démonstration. On remarque que $P.f(x)$ peut s'écrire de la façon suivante :

$$P.f(x) = \Psi_{Q,1}(\delta_x.M) (P'.f).$$

Ainsi, pour tous $x, y \in E$, on a

$$\begin{aligned} |P.f(x) - P.f(y)| &= |(\Psi_{Q,1}(\delta_x.M) - \Psi_{Q,1}(\delta_y.M)) (P'.f)| \\ &\leq \|(\Psi_{Q,1}(\delta_x.M) - \Psi_{Q,1}(\delta_y.M))\|_{tv} \cdot osc(P'.f). \end{aligned}$$

Par le lemme 1 (page 22), cela implique

$$\|(\Psi_{Q,1}(\delta_x.M) - \Psi_{Q,1}(\delta_y.M))\|_{tv} \leq g \cdot \|\delta_x.M - \delta_y.M\|_{tv} \leq g \cdot \beta(M).$$

D'après la propriété (2.4) donnée page 22, on a $osc(P'.f) \leq \beta(P').osc(f)$, ce qui conclut la démonstration. \square

Lemme 4. *Pour toute paire d'entiers $p \leq n$, on a :*

$$g_{p,n} - 1 \leq \sum_{k=p+1}^n (g_k - 1) \prod_{i=p+1}^{k-1} (b_i g_i) \quad (2.19)$$

$$b_{p,n} \leq \prod_{k=p+1}^n b_k \cdot g_{k,n} \quad (2.20)$$

Démonstration. Démontrons (2.19). Par définition, on a $G_{p,n} = Q_{p,n} \cdot 1$. En combinant (2.10) et (2.11) (page 24), appliqués à la fonction unité, on a

$$Q_{p-1,n}(1) = Q_{p+1} \cdot [(Q_{p+1} \dots Q_n) \cdot 1] = G_p \times M_p \cdot (Q_{p,n} \cdot 1).$$

Cela implique que les fonctions $G_{p,n}$ vérifient les relations à rebours suivantes :

$$G_{n,n} = 1 \quad ; \quad G_{p-1,n} = G_p \times M_p \cdot G_{p,n}$$

Ainsi, pour tous $x, y \in E$, on obtient

$$\frac{G_{p-1,n}(x)}{G_{p-1,n}(y)} = \underbrace{\frac{G_p(x)}{G_p(y)}}_{E_1} \times \underbrace{\frac{(M_p \cdot G_{p,n})(x)}{(M_p \cdot G_{p,n})(y)}}_{E_2}.$$

On remarque que $E_1 \leq g_p$ (par définition), et par le lemme 2, on a $E_2 \leq 1 + \beta(M_p) \cdot (g_{p,n} - 1)$. Cela implique les inégalités à rebours suivantes :

$$g_{n,n} = 1 \quad ; \quad g_{p-1,n} \leq g_p (1 + b_p (g_{p,n} - 1)) \quad (2.21)$$

De là, (2.19) s'obtient par simple récurrence.

Pour démontrer (2.20), on utilise les formules

$$P_{p-1,n} \cdot f = \frac{Q_{p-1,n} \cdot f}{Q_{p-1,n} \cdot 1} = \frac{G_p \times M_p \cdot Q_{p,n} \cdot f}{G_p \times M_p \cdot Q_{p,n} \cdot 1} = \frac{M_p \cdot Q_{p,n} \cdot f}{M_p \cdot Q_{p,n} \cdot 1}.$$

2.2. ANALYSE NON ASYMPTOTIQUE GÉNÉRALE

En se souvenant que $P_{p,n}.f = \frac{Q_{p,n}.f}{Q_{p,n}.1}$, on applique le lemme 3 pour vérifier que $\beta(P_{p-1,n}) \leq \beta(M_p).g_{p,n}.\beta(P_{p,n})$. De là, on conclut que

$$b_{p-1,n} \leq b_p.g_{p,n}.b_{p,n}.$$

Le résultat (2.20) s'obtient alors par récurrence. □

Terminons cette section avec un lemme très utile sur le contrôle du produit $g_{p,n}b_{p,n}$.

Lemme 5. *Pour toute paire d'entiers $p \leq n$, on a*

$$g_{p,n}b_{p,n} \leq \prod_{k=p+1}^n (b_k.g_{k-1,n}).$$

Démonstration. A l'aide du lemme 4, on a

$$\begin{aligned} g_{p,n}b_{p,n} &\leq g_{p,n} \cdot \prod_{k=p+1}^n b_k g_{k,n} \\ &= g_{p,n} \cdot (b_{p+1}g_{p+1,n}) \cdot (b_{p+2}g_{p+2,n}) \cdots (b_{n-1}g_{n-1,n}) \cdot (b_n \underbrace{g_{n,n}}_{=1}) \\ &= (g_{p,n}b_{p+1}) \cdot (g_{p+1,n}b_{p+2}) \cdots (g_{n-1,n}b_n) = \prod_{k=p+1}^n b_k g_{k-1,n}, \end{aligned}$$

ce qui conclut la démonstration. □

Le terme $g_{p,n}b_{p,n}$ est central dans la borne L^p (2.13) donnée page 26. D'après le lemme 5 on a l'implication suivante :

$$\sum_{p=0}^n \prod_{k=p+1}^n b_k g_{k-1,n} < +\infty \implies \sum_{p=0}^n g_{p,n}b_{p,n} < +\infty.$$

Cela fournit une condition suffisante pour l'obtention d'une borne L^p uniforme en temps n . $g_{p,n}b_{p,n}$ est également impliqué dans les estimées de toutes les quantités définies dans la section 2.1.4 (page 25) comme r_n , $\overline{\beta}_n^2$, etc. De plus, d'après le lemme 1, on a la propriété de stabilité

$$\|\phi_{p,n}(\mu) - \phi_{p,n}(\nu)\|_{tv} \leq g_{p,n}b_{p,n} \|\mu - \nu\|_{tv}.$$

Cela montre que le produit $g_{p,n}b_{p,n}$ est central pour quantifier les propriétés de stabilité du semigroupe $\phi_{p,n}$.

2.2. ANALYSE NON ASYMPTOTIQUE GÉNÉRALE

2.2.2 Théorèmes de concentration uniformes

Dans le but d'obtenir des bornes uniformes par rapport à l'horizon temporel, le lemme 5 mène naturellement à une condition suffisante du type

$$\sup_{k \leq n} b_k \cdot g_{k-1,n} \leq a \quad \text{pour un certain } a \in]0, 1[$$

Dans cette situation, on montre que $g_{p,n} b_{p,n} \leq a^{n-p}$, et ainsi, à l'aide de (2.13),

$$\sup_{f \in \mathcal{B}_1(E)} \mathbb{E} \left(\left| \eta_n^N(f) - \eta_n(f) \right|^p \right)^{1/p} \leq \frac{B_p}{\sqrt{N}} \frac{1}{1-a},$$

avec les constantes B_p introduites en (2.2) page 21. On se fixe donc un paramètre $a \in]0, 1[$ et on cherche des conditions sur les b_p de façon à garantir $b_k g_{k-1,n} \leq a$. Ce paramètre a peut s'interpréter comme un degré de performance de l'approximation du modèle par les N particules. Afin d'explicitier des conditions pertinentes et surtout applicables, on étudie deux classes typiques de conditions de régularité sur les potentiels G_p . La première consiste à considérer que les coefficients g_p sont bornés (théorème 4). La seconde consiste à supposer que la suite de paramètres g_p tend vers 1 quand $p \rightarrow \infty$ (théorème 5).

Les inégalités de concentration développées dans le théorème 4 seront énoncées en fonction des paramètres (r_1^*, r_2^*) définis ci-après :

$$\begin{cases} r_1^* &= \frac{9}{2} \frac{(M(1+a))^2}{1-a} + \sqrt{\frac{8}{\sqrt{1-a^2}} + \frac{18(M(1+a))^2}{\sqrt{N}}} \\ r_2^* &= 18 \frac{(M(1+a))^2}{1-a} + \sqrt{\frac{8}{\sqrt{1-a^2}} + \frac{18(M(1+a))^2}{\sqrt{N}}} \end{cases} \quad (2.22)$$

Théorème 4. *On suppose*

$$\sup_{p \geq 1} g_p \leq M \quad \text{et} \quad \sup_{p \geq 1} b_p \leq \frac{a}{M(1+a)} \quad (2.23)$$

pour une constante finie $M \geq 1$. Dans cette situation, on a les estimées uniformes suivantes.

– La borne L^p :

$$\sup_{n \geq 0} d_p(\eta_n^N, \eta_n) \leq \frac{B_p}{2(1-a)\sqrt{N}} \quad (2.24)$$

– Pour tous $n \geq 0$, $N \geq 1$, $y \geq 0$ et $f \in \mathcal{B}_1(E)$, la probabilité de l'événement

$$\left| \eta_n^N(f) - \eta_n(f) \right| \leq \frac{r_1^* N + r_2^* y}{N^2} \quad (2.25)$$

est supérieure à $1 - e^{-y}$, avec les paramètres (r_1^*, r_2^*) définis en (2.22).

2.2. ANALYSE NON ASYMPTOTIQUE GÉNÉRALE

– Pour tous $n \geq 0$, $N \geq 1$, $\epsilon \in \{+1, -1\}$, $y \geq 0$ et $f \in \mathcal{B}_1(E)$, la probabilité de l'événement

$$\frac{\epsilon}{n} \log \left(\frac{\gamma_n^N(1)}{\gamma_n(1)} \right) \leq \frac{\tilde{r}_1}{N} h_0(y) + \tilde{r}_2 \cdot h_1 \left(\frac{y}{n \cdot N} \right) \quad (2.26)$$

est supérieure à $1 - e^{-y}$, avec les paramètres $\tilde{r}_1 = \frac{8M(M(1+a))^2}{1-a}$ et $\tilde{r}_2 = \frac{4M}{1-a}$.

Démonstration. Montrons tout d'abord les inégalités suivantes :

$$\begin{cases} g_{p,n} & \leq M(1+a) \\ b_p \cdot g_{p-1,n} & \leq a \end{cases} \quad (2.27)$$

Supposons que $b_p \leq b$ pour un certain $b < 1$. D'après le lemme 4 (page 29), on a

$$\begin{aligned} g_{p,n} - 1 & \leq \sum_{k=p+1}^n (M-1)(bM)^{k-p-1} \\ & = (M-1) \frac{1 - (bM)^{n-p}}{1 - bM} \\ & \leq \frac{M-1}{1 - bM}. \end{aligned}$$

Donc si $b = \frac{a}{M(1+a)}$ on obtient

$$g_{p,n} \leq \frac{M-1}{1 - \frac{a}{1+a}} + 1 = (1+a)(M-1) + 1 \leq (1+a)M.$$

D'autre part, on a

$$\begin{aligned} b_p g_{p-1,n} & \leq \frac{b(M-1)}{1 - \frac{a}{1+a}} + b \\ & = (1+a) \frac{a(M-1)}{M(1+a)} + \frac{a}{M(1+a)} \\ & = \frac{1}{M} \left(aM - a + \frac{a}{1+a} \right) \\ & \leq a \end{aligned}$$

et les inégalités (2.27) sont donc prouvées. Considérons maintenant les trois parties du théorème point par point.

- D'après le lemme 5 et (2.27) on a $g_{p,n} b_{p,n} \leq a^{n-p}$. En combinant cela avec (2.13), la borne L^p (2.24) est triviale.
- Montrons (2.25), qui est une conséquence de l'inégalité de concentration (2.14), page 27. En combinant les estimations de r_n , $\overline{\beta}_n^2$ et b_n^* données dans la section 2.1.4 avec (2.27) et $g_{p,n} b_{p,n} \leq a^{n-p}$, on obtient

2.2. ANALYSE NON ASYMPTOTIQUE GÉNÉRALE

$$r_n \leq \frac{4(M(1+a))^2}{1-a}, \quad \bar{\beta}_n^2 \leq \frac{4}{1-a^2} \quad \text{et} \quad b_n^* \leq 2.$$

En appliquant ces estimations dans (2.14), on obtient

$$-\log \mathbb{P} \left(|\eta_n^N(f) - \eta_n(f)| \geq \frac{r_1}{4N} + \varepsilon \right) \geq \frac{N\varepsilon^2}{r_2 + r_1(\frac{1}{\sqrt{N}} + \varepsilon)}$$

avec $r_1 = \frac{18M(1+a)}{1-a}$ et $r_2 = \frac{8}{\sqrt{1-a^2}}$. On pose $y = \frac{N\varepsilon^2}{r_2 + r_1(\frac{1}{\sqrt{N}} + \varepsilon)} > 0$. Etant donné un certain $y > 0$, on a

$$\varepsilon(y) = \frac{r_1 y + \sqrt{r_1^2 y^2 + 4N(r_2 + \frac{r_1}{\sqrt{N}})y}}{2N^2} \implies y = \frac{N\varepsilon(y)^2}{r_2 + r_1(\frac{1}{\sqrt{N}} + \varepsilon(y))}$$

et il devient alors clair que

$$\mathbb{P} \left(|\eta_n^N(f) - \eta_n(f)| \geq \frac{r_1}{4N} + \varepsilon(y) \right) \leq e^{-y}.$$

Après une série de majorations élémentaires, on obtient

$$\frac{r_1}{4N} + \varepsilon(y) \leq \frac{r_1^* N + r_2^* y}{N^2}.$$

- La dernière inégalité de concentration (2.26) est une conséquence de (2.14) (voir page 27) et de (2.27). En effet, à partir des estimations (2.17) et (2.18) page 27, on montre facilement que les quantités τ_n^* et $\bar{r}(n)$ vérifient

$$\tau_n^* \leq \frac{4M}{n(1-a)} \quad \text{et} \quad \bar{r}(n) \leq \frac{8.M^3(1+a)^2}{1-a}.$$

D'autre part, $\bar{\sigma}_n^2$ est clairement borné par n . On écrit alors

$$\frac{\bar{r}(n)}{N} h_0(y) + \tau_n^* \bar{\sigma}_n^2 h_1 \left(\frac{y}{N \cdot \bar{\sigma}_n^2} \right) = \frac{\bar{r}(n)}{N} h_0(y) + \frac{y \tau_n^*}{3N} + \sqrt{\frac{2y(\tau_n^* \bar{\sigma}_n^2)}{N}},$$

et le résultat (2.26) s'obtient en effectuant les remplacements adéquats. □

Considérons maintenant le cas où g_p décroît et converge vers 1 lorsque $p \rightarrow \infty$. L'idée de l'analyse qui suit est de trouver une condition sur les b_p qui garantisse que les $g_{p,n}$ sont uniformément bornés par rapport à n par $g_{p+1}^{1+\alpha}$, avec

$$\alpha = \frac{a}{1-a} > 0 \quad \left(\iff a = \frac{\alpha}{1+\alpha} \right).$$

2.2. ANALYSE NON ASYMPTOTIQUE GÉNÉRALE

Les inégalités de concentration développées dans le théorème 5 qui vient sont formulées à l'aide des paramètres $r_3^*(n)$, $r_4^*(n)$ et $\tilde{r}_3(n)$, \tilde{r}_4 , $\tilde{r}_5(n)$ définis ci-dessous :

$$\begin{cases} r_3^*(n) &= \frac{9 \cdot u_1(n)}{2(1-a)} + \sqrt{\frac{8}{\sqrt{1-a^2}} + \frac{18 \cdot u_1(n)}{\sqrt{N}}} \\ r_4^*(n) &= \frac{18 \cdot u_1(n)}{1-a} + \sqrt{\frac{8}{\sqrt{1-a^2}} + \frac{18 \cdot u_1(n)}{\sqrt{N}}} \end{cases} \quad (2.28)$$

$$\begin{cases} \tilde{r}_3(n) &= \frac{16 \cdot u_2(n)}{1-a} \\ \tilde{r}_4 &= \frac{4}{3} \sum_{n \geq 0} g_{n+1} \cdot a^n \leq \frac{4 \cdot g_1}{3(1-a)} \\ \tilde{r}_5(n) &= \frac{4\sqrt{2} \cdot u_3(n)}{1-a} \end{cases} \quad (2.29)$$

Les suites $u_1(n)$, $u_2(n)$ et $u_3(n)$ apparaissant dans les formules précédentes sont définies par

$$\begin{cases} u_1(n) &= (1-a) \sum_{p \geq 0} g_{n-p+1}^{2(1+\alpha)} a^p \xrightarrow{n \rightarrow \infty} 1 \\ u_2(n) &= \frac{1}{n} \sum_{p=1}^n g_p^{3+2\alpha} \xrightarrow{n \rightarrow \infty} 1 \\ u_3(n) &= \left(\frac{1}{n} \sum_{p=0}^{n-1} g_{p+1}^2 \right)^{1/2} \xrightarrow{n \rightarrow \infty} 1. \end{cases}$$

La suite $u_1(n)$ tend vers 1 par convergence dominée ; les suites $u_2(n)$ et $u_3(n)$ par théorème de Césaro.

Théorème 5. *On suppose que $g_p \downarrow 1$ quand $p \rightarrow \infty$ et que la suite b_p vérifie pour tout $p \geq 1$*

$$b_p \leq \frac{g_p^\alpha - 1}{g_p^{\alpha+1} - 1} \left(\xrightarrow{p \rightarrow +\infty} a \right) \quad \text{et} \quad b_p \leq \frac{a}{g_p^{\alpha+1}} \left(\xrightarrow{p \rightarrow +\infty} a \right). \quad (2.30)$$

Sous ces hypothèses, on a les estimées uniformes suivantes.

– La borne L^p

$$\sup_{n \geq 0} d_p(\eta_n^N, \eta_n) \leq \frac{B_p}{2(1-a)\sqrt{N}} \quad (2.31)$$

avec les constantes B_p introduites en (2.2) page 21.

– Pour tous $n \geq 0$, $N \geq 1$, $y \geq 0$ et $f \in \mathcal{B}_1(E)$, la probabilité de l'événement

$$|\eta_n^N(f) - \eta_n(f)| \leq \frac{r_3^*(n) \cdot N + r_4^*(n) \cdot y}{N^2} \quad (2.32)$$

est supérieure à $1 - e^{-y}$, avec les paramètres $r_3^(n)$, $r_4^*(n)$ définis en (2.28).*

– Pour tous $n \geq 0$, $N \geq 1$, $\epsilon \in \{+1, -1\}$, $y \geq 0$ et $f \in \mathcal{B}_1(E)$, la probabilité de l'événement

$$\frac{\epsilon}{n} \log \left(\frac{\gamma_n^N(1)}{\gamma_n(1)} \right) \leq \tilde{r}_3(n) \left(\frac{y + \sqrt{y}}{N} \right) + \tilde{r}_4 \left(\frac{y}{n \cdot N} \right) + \tilde{r}_5(n) \sqrt{\frac{y}{n \cdot N}} \quad (2.33)$$

est supérieure à $1 - e^{-y}$, avec les paramètres $\tilde{r}_3(n)$, \tilde{r}_4 , $\tilde{r}_5(n)$ définis en (2.29).

2.2. ANALYSE NON ASYMPTOTIQUE GÉNÉRALE

Démonstration. Commençons par montrer l'implication suivante :

$$(2.30) \implies \forall p \leq n, \begin{cases} g_{p,n} & \leq (g_{p+1})^{1+\alpha} \\ g_{p-1,n} \cdot b_p & \leq a \end{cases} \quad (2.34)$$

La preuve de la première inégalité est une simple récurrence à rebours sur p (avec n fixé), à l'aide de la majoration $g_{p-1,n} \leq g_p(1 + b_p(g_{p,n} - 1))$ (voir (2.21) page 29). Pour $p = n$, $g_{p,n}$ est clairement inférieur à $g_{p+1}^{1+\alpha}$ car $g_{n,n} = 1$. La seconde inégalité est immédiate. On suppose maintenant que $g_{p,n} \leq g_{p+1}^{1+\alpha}$. Dans ce cas, $g_{p-1,n} \leq g_p^{1+\alpha}$ est vérifié dès lors que

$$b_p \leq \frac{g_p^\alpha - 1}{g_{p+1}^{\alpha+1} - 1}.$$

On remarque que cette majoration est elle-même vraie dès que $b_p \leq \frac{g_p^\alpha - 1}{g_p^{\alpha+1} - 1}$, la suite $(g_p)_p$ étant supposée décroissante.

- Maintenant que nous avons montré (2.34) (qui implique $g_{p,n}b_{p,n} \leq a^{n-p}$ d'après le lemme 5 page 30), la borne L^p (2.24) s'obtient par une simple substitution dans le résultat L^p de référence (2.13) page 26.
- Pour démontrer (2.32), on se concentre sur les quantités $\bar{\beta}_n^2$, b_n^* et r_n apparaissant dans l'inégalité de concentration (2.14) page 27. A partir de (2.34) et $g_{p,n}b_{p,n} \leq a^{n-p}$, on vérifie facilement que

$$\bar{\beta}_n^2 \leq \frac{4}{1-a^2} \quad \text{et} \quad b_n^* \leq 2.$$

Le terme r_n peut être grossièrement majoré par $\frac{4}{1-a}g_1^{2(1+\alpha)}$, mais une manipulation plus fine fournit une estimation plus précise. En effet, en utilisant le fait que $b_{p,n}g_{p,n} \leq a^{n-p}$ et $g_{p,n} \leq g_{p+1}^{1+\alpha}$, on montre que

$$\begin{aligned} r_n &\leq 4 \sum_{p=0}^n g_{p,n}^2 a^{n-p} &\leq 4 \sum_{p=0}^n g_{p+1}^{2(1+\alpha)} a^{n-p} &\leq 4 \sum_{p=0}^n g_{n-p+1}^{2(1+\alpha)} a^p \\ &&&&&\leq 4 \sum_{p \geq 0} g_{n-p+1}^{2(1+\alpha)} a^p \\ &&&&&\leq \frac{4 \cdot u_1(n)}{1-a}. \end{aligned}$$

A partir de là, on montre (2.32) par les mêmes arguments que ceux donnés dans la preuve du théorème 4.

- Montrons la dernière inégalité de concentration (2.33). Il s'agit essentiellement d'une conséquence de l'inégalité (2.16) donnée page 27. En partant de la décomposition suivante

$$\frac{\bar{r}(n)}{N} h_0(y) + \tau_n^* \bar{\sigma}_n^2 h_1\left(\frac{y}{N \cdot \bar{\sigma}_n^2}\right) = \frac{2\bar{r}(n)}{N} (y + \sqrt{y}) + \frac{y\tau_n^*}{3N} + \sqrt{\frac{2y(\tau_n^* \bar{\sigma}_n^2)}{N}}, \quad (2.35)$$

on se doit de trouver des estimations pertinentes des quantités τ_n^* , \bar{r}_n et $(\tau_n^* \cdot \bar{\sigma}_n^2)$. Pour estimer τ_n^* , on remarque que $\forall q, g_{p+1} \leq g_{p-q+1}$, de sorte que

2.3. MÉTHODES SMC À SCHÉMA DE TEMPÉRATURE, RECUI TS SIMULÉS EN INTERACTION

$$\begin{aligned} \tau_{q,n} &\leq \frac{4}{n} \sum_{p=q}^{n-1} g_{p+1} a^{p-q} \leq \frac{4}{n} \sum_{p=q}^{n-1} g_{p-q+1} a^{p-q} \leq \frac{4}{n} \sum_{p=0}^{n-q-1} g_{p+1} a^p \\ &\leq \frac{4}{n} \sum_{p=0}^{n-1} g_{p+1} a^p. \end{aligned}$$

Finalement, on a $\tau_n^* \leq \frac{4.U_0}{n}$, où $U_0 = \sum_{p \geq 0} g_{p+1} a^p \leq \frac{g_1}{1-a}$.

On estime \bar{r}_n à l'aide des inégalités suivantes :

$$\begin{aligned} \bar{r}_n &\leq \frac{8}{n} \sum_{0 \leq q \leq p < n} g_{p+1} \cdot g_{p,q}^3 \cdot b_{p,q} \leq \frac{8}{n} \sum_{p=0}^{n-1} \sum_{q=p-n}^p g_{p+1}^{3+2\alpha} a^{p-q} \\ &\leq \frac{8}{1-a} \cdot \underbrace{\frac{1}{n} \sum_{p=1}^n g_p^{3+2\alpha}}_{=u_2(n)} \end{aligned}$$

Une dernière majoration nécessaire :

$$\begin{aligned} \tau_n^* \cdot \bar{\sigma}_n^2 &\leq \sum_{q=0}^{n-1} \tau_{q,n}^2 \leq \sum_{q=0}^{n-1} \left(\frac{4}{n} \sum_{p=0}^{n-q-1} g_{p+q+1} \cdot a^p \right)^2 \\ &\leq \sum_{q=0}^{n-1} \frac{16}{n^2} \cdot g_{q+1}^2 \cdot \frac{1}{(1-a)^2} \\ &\leq \frac{16}{n \cdot (1-a)^2} \times \underbrace{\frac{1}{n} \sum_{q=0}^{n-1} g_{q+1}^2}_{=(u_3(n))^2} \end{aligned}$$

A partir de ces estimations, le résultat énoncé (2.33) s'obtient en faisant les remplacements adéquats dans (2.35). □

2.3 Méthodes SMC à schéma de température, recuits simulés en interaction

2.3.1 Motivations

On considère la mesure de probabilité de Gibbs associée à un paramètre de température inverse $\beta \geq 0$ et à un potentiel donné $V \in \mathcal{B}(E)$, définie par

$$\mu_\beta(dx) = \frac{1}{Z_\beta} e^{-\beta \cdot V(x)} m(dx), \quad (2.36)$$

où m désigne une mesure de référence, et Z_β est une constante de normalisation. Soit β_n une suite réelle strictement croissante (qui tend éventuellement vers $+\infty$ quand $n \rightarrow \infty$).

2.3. MÉTHODES SMC À SCHÉMA DE TEMPÉRATURE, RECUITS SIMULÉS EN INTERACTION

Dans cette situation, les mesures $\eta_n = \mu_{\beta_n}$ peuvent être interprétées comme un flot de FK, associé aux potentiels $G_n = e^{-(\beta_n - \beta_{n-1})V}$ et aux transitions markoviennes M_n choisies comme étant des noyaux de MCMC vis-à-vis des lois cibles courantes (c'est-à-dire laissant stables les mesures μ_{β_n}). En effet, on a

$$\mu_{\beta_n}(dx) = \frac{e^{-\beta_n \cdot V(x)}}{Z_{\beta_n}} m(dx) = \frac{Z_{\beta_{n-1}}}{Z_{\beta_n}} \underbrace{e^{-(\beta_n - \beta_{n-1})V(x)}}_{=G_n(x)} \underbrace{\left(\frac{e^{-\beta_{n-1} \cdot V(x)}}{Z_{\beta_{n-1}}} m(dx) \right)}_{=\mu_{\beta_{n-1}}(dx)},$$

ce qui montre bien $\mu_{\beta_n} = \psi_{G_n}(\mu_{\beta_{n-1}})$. Soit ϕ_n la transformation de FK associée au potentiel G_n et à la transition de Markov M_n . On a

$$\phi_n(\mu_{\beta_{n-1}}) = \psi_{G_n}(\mu_{\beta_{n-1}}) = \mu_{\beta_n} \cdot M_n = \mu_{\beta_n}.$$

Savoir simuler ces lois (et calculer les constantes de normalisation $Z_{\beta_n} = \gamma_n(1)$) est un défi important dans de nombreux domaines d'application. Le plus simple d'entre eux consiste en la simulation d'une loi a posteriori complexe, sur un espace euclidien $E = \mathbb{R}^d$. Soit x la variable d'intérêt, associée à une densité a priori $p(x)$ (que l'on sait simuler) vis-à-vis de la mesure de Lebesgue dx sur E , et y un vecteur d'observations, associé à un modèle de vraisemblance calculable $p(y | x)$. Dans ce contexte, rappelons que $p(y | x)$ est la densité des observations sachant la variable d'intérêt. La densité $p(x | y)$ de la loi a posteriori η est donnée par la formule de Bayes

$$p(x | y) \propto p(x) \cdot p(y | x).$$

Dans le cas où η est fortement multimodale, il est difficile de simuler cette loi directement. Par exemple, les méthodes MCMC classiques ont tendance à rester piégées dans des modes locaux pendant des temps très longs. Par conséquent, elles ne convergent vers leur mesure limite η que pour des temps de calcul inconciliables avec une utilisation pratique. Pour pallier ce problème, une solution commune est d'approcher la loi cible η par une suite de mesures η_0, \dots, η_n qui ont pour densité

$$\eta_k(dx) \propto p(x) \cdot p(y | x)^{\beta_k} dx$$

où $(\beta_k)_{0 \leq k \leq n}$ est une suite réelle croissant de 0 à 1, de sorte que η_0 est la loi a priori de densité $p(x)$, facile à simuler, et la mesure finale η_n est la loi cible η (voir par exemple [7, 67, 69]). Si l'on pose $V := x \mapsto -\log(p(y | x))$ et $m(dx) := p(x)dx$, alors les η_k coïncident avec la mesure de Gibbs μ_{β_k} définie en (2.36). Dans ce contexte, les méthodes SMC s'avèrent être une approche pertinente, en particulier si η est multimodale, car l'utilisation d'un grand nombre de particules permet au nuage de couvrir plusieurs modes simultanément.

La constante de normalisation $\gamma_n(1)$ coïncide avec la vraisemblance marginale $p(y)$. Le calcul de cette constante est un autre problème tout à fait central en sélection de modèle, apparaissant naturellement dans les problèmes de chaînes de Markov cachées (voir la section 1.3.3 page 13) et de statistiques bayésiennes.

2.3. MÉTHODES SMC À SCHÉMA DE TEMPÉRATURE, RECUITS SIMULÉS EN INTERACTION

En outre, présentons ici une autre application importante du calcul des constantes $\gamma_n(1)$ en physique-chimie : l'estimation d'énergie libre. Le problème se pose via une densité non normalisée de la forme

$$q(\omega | T, \alpha) = \exp\left(-\frac{H(\omega, \alpha)}{k.T}\right)$$

où $H(\omega, \alpha)$ est l'énergie de l'état ω , k est la constante de Boltzmann, T est la température et α est un vecteur comprenant les caractéristiques du système. L'énergie libre F du système est définie par la quantité

$$F(T, \alpha) = -k.T. \log(z(T, \alpha)),$$

où $z(T, \alpha)$ est la constante de normalisation de la densité du système. Le lecteur pourra par exemple se référer à [15, 21, 45] pour une discussion approfondie de ces problèmes d'estimation de niveaux d'énergie.

Enfin, il est bien connu que la simulation des mesures de Gibbs est liée au problème de la minimisation du potentiel V . L'idée centrale est que μ_β tend à se concentrer sur les minimiseurs de V lorsque la température inverse β tend vers l'infini. Pour être plus précis sur cette idée, nous fournissons ici une inégalité de concentration exponentielle dans un prochain lemme (lemme 6).

Dans cette situation, les méthodes SMC peuvent s'interpréter comme l'évolution de N algorithmes de recuit simulé en interaction les uns avec les autres (voir la section 1.2.2 page 7 pour la description du recuit simulé simple). On abrégera ISA pour Interacting Simulated Annealing. D'autre part, dans la mesure où ils font intervenir une population de N individus soumis à des processus de type génétique (sélection, mutation), les méthodes ISA appartiennent aussi à la large famille des algorithmes évolutionnaires pour l'optimisation globale. Ces algorithmes consistent à explorer un espace d'état au moyen d'une population soumise à une stratégie d'évolution, plus précisément une évolution basée sur la sélection, la mutation, et le crossover. On pourra voir [49] ou [60] et leurs références pour un état de l'art. Comme ces algorithmes mettent en scène des stratégies complexes, éventuellement adaptatives, leur analyse est essentiellement heuristique, empirique, ou parfois asymptotique (voir [16] pour des résultats généraux de convergence des algorithmes génétiques). Le lecteur trouvera également dans [35] une preuve pour les méthodes ISA de la convergence presque sûre vers le minimum global dans le cas d'un espace d'état fini, lorsque le temps n tend vers l'infini, dès lors que la taille N de la population est supérieure à une constante critique.

Les résultats de la section précédente s'appliquent aux méthodes d'optimisation ISA. Notre approche est non asymptotique car nous estimons pour chaque instant n , et pour une taille de population N fixée la distance entre la mesure de Gibbs théorique η_n et son approximation empirique η_n^N . En un certain sens, lorsque $\beta_n \rightarrow +\infty$, η_n converge vers la mesure de Dirac δ_{x^*} où $x^* = \underset{x \in E}{\operatorname{argmin}} V(x)$, dès lors que le minimum global existe et est unique. On conjecture donc intuitivement que si cette distance admet une borne uniforme

2.3. MÉTHODES SMC À SCHÉMA DE TEMPÉRATURE, RECUITS SIMULÉS EN INTERACTION

en temps n , alors pour n grand on a $\eta_n^N \simeq \delta_{x^*}$.

Dans cette section, on se propose de transformer les conditions des théorèmes 4 et 5 en des conditions sur le schéma de température à utiliser et sur le nombres d'étapes de MCMC à chaque mutation pour garantir un certain degré de performance. Puis nous combinons les résultats de concentration de la section 2.2.2 avec le lemme 6 pour analyser la convergence du SMC en tant qu'algorithme d'optimisation globale.

2.3.2 Un modèle d'optimisation ISA

On se fixe un schéma de température inverse β_n et on pose

- $\eta_n(dx) = \mu_{\beta_n}(dx) = \frac{1}{Z_{\beta_n}} e^{-\beta_n V(x)} m(dx)$;
- $G_n(x) = e^{-\Delta_n \cdot V(x)}$;
- et donc $g_n = e^{\Delta_n \cdot \text{osc}(V)}$,

où les Δ_n désignent les incréments de température $\Delta_n = \beta_n - \beta_{n-1}$. Soit K_β le noyau de Markov propre au recuit simulé (voir la section 1.1.1 page 3), de mesure invariante μ_β et de noyau de proposition $K(x, dy)$, noyau que l'on suppose réversible par rapport à $m(dx)$. Rappelons que $K_\beta(x, dy)$ est défini par les formules suivantes :

$$K_\beta(x, dy) = K(x, dy) \cdot \min\left(1, e^{-\beta(V(y)-V(x))}\right) \quad \forall y \neq x$$

$$K_\beta(x, \{x\}) = 1 - \int_{y \neq x} K(x, dy) \cdot \min\left(1, e^{-\beta(V(y)-V(x))}\right)$$

Si l'on suppose qu'il existe une mesure ν et un réel $\delta > 0$ tels que pour tout x on a

$$K^{k_0}(x, \cdot) \geq \delta \nu(\cdot),$$

alors on peut montrer (voir par exemple [6]) que

$$\beta(K_\beta^{k_0}) \leq \left(1 - \delta e^{-\beta \overline{\Delta V}(k_0)}\right) \quad (2.37)$$

où $\overline{\Delta V}(k_0)$ désigne l'écart de potentiel maximum que l'on peut obtenir en faisant k_0 sauts élémentaires selon la transition markovienne K . Une solution algorithmique pour contrôler les propriétés de mélange du modèle ISA (et plus généralement des modèles SMC à schéma de température) consiste à considérer la transition markovienne $M_p = K_{\beta_p}^{k_0 \cdot m_p}$, c'est-à-dire le noyau du recuit simulé itéré $k_0 \cdot m_p$ fois. Dans ce cas, le concepteur de l'algorithme a un choix à faire sur le réglage de deux paramètres, à savoir le schéma de température β_p , et les nombres d'itération m_p . Remarquons dès maintenant que pour tout $b \in]0, 1[$, la condition $b_p \leq b$ se transforme en $\left(1 - \delta e^{-\beta_p \overline{\Delta V}(k_0)}\right)^{m_p} \leq b$, ce qui peut aussi s'écrire

$$m_p \geq \frac{\log\left(\frac{1}{b}\right) e^{\overline{\Delta V}(k_0) \cdot \beta_p}}{\delta}.$$

Démontrons maintenant un lemme technique qui sera utile dans la suite. Il concerne les propriétés de concentration des mesures de Gibbs.

2.3. MÉTHODES SMC À SCHÉMA DE TEMPÉRATURE, RECUITS SIMULÉS EN INTERACTION

Lemme 6. *Pour tout $\beta > 0$, et pour tous $0 < \varepsilon' < \varepsilon$, la mesure de Gibbs μ_β vérifie*

$$\mu_\beta(V \geq V_{\min} + \varepsilon) \leq \frac{e^{-\beta(\varepsilon - \varepsilon')}}{m_{\varepsilon'}},$$

où $m_{\varepsilon'} = m(V \leq V_{\min} + \varepsilon') > 0$.

Démonstration. La constante de normalisation Z_β de la définition (2.36) est nécessairement égale à $\int_E e^{-\beta V} dm$. On a donc

$$\begin{aligned} \mu_\beta(V \geq V_{\min} + \varepsilon) &= \frac{\int_{V \geq V_{\min} + \varepsilon} e^{-\beta V} dm}{\int_{V \geq V_{\min} + \varepsilon} e^{-\beta V} dm + \int_{V < V_{\min} + \varepsilon} e^{-\beta V} dm} \\ &\leq \underbrace{\left(\int_{V \geq V_{\min} + \varepsilon} e^{-\beta V} dm \right)}_{A_1} \underbrace{\left(\int_{V < V_{\min} + \varepsilon} e^{-\beta V} dm \right)^{-1}}_{A_2^{-1}}. \end{aligned}$$

Premièrement, il est clair que $A_1 \leq e^{-\beta(V_{\min} + \varepsilon)}$. Deuxièmement, $\varepsilon' < \varepsilon$ implique $\{V \leq V_{\min} + \varepsilon'\} \subset \{V < V_{\min} + \varepsilon\}$, donc on a

$$A_2 \geq \int_{V \leq V_{\min} + \varepsilon'} e^{-\beta V} dm \geq m(V \leq V_{\min} + \varepsilon') e^{-\beta(V_{\min} + \varepsilon')}.$$

On conclut la démonstration en effectuant les remplacements appropriés. \square

En combinant ce lemme 6, les théorèmes de la section 2.2.2 (avec la fonction indicatrice $f = \mathbf{1}_{\{V \geq V_{\min} + \varepsilon\}}$), et les estimées des coefficients de Dobrushin (2.37), on prouve aisément le théorème suivant :

Théorème 6. *Fixons $a \in]0, 1[$. Pour tout $\varepsilon > 0$, $n \geq 0$ et $N \geq 1$, soit $p_n^N(\varepsilon)$ la proportion des particules (ζ_n^i) telles que $V(\zeta_n^i) \geq V_{\min} + \varepsilon$. On suppose que le schéma de température inverse β_p et les nombres d'itération m_p vérifient l'une ou l'autre de ces deux conditions :*

1. $\sup_{p \geq 1} \Delta_p \leq \Delta < \infty$ (typiquement un schéma linéaire) et

$$m_p \geq \frac{\log\left(\frac{e^{\Delta \cdot \text{osc}(V)}(1+a)}{a}\right) e^{\overline{\Delta V}(k_0) \cdot \beta_p}}{\delta}.$$

2. $\Delta_p \downarrow 0$ (quand $p \rightarrow \infty$) et $m_p \geq \left(\text{osc}(V) \cdot \Delta_p + \log\left(\frac{1}{a}\right)\right) \frac{e^{\overline{\Delta V}(k_0) \cdot \beta_p}}{\delta}$.

2.3. MÉTHODES SMC À SCHÉMA DE TEMPÉRATURE, RECUITS SIMULÉS EN INTERACTION

Dans cette situation, pour tout $\varepsilon > 0$, $n \geq 0$, $N \geq 1$, $y \geq 0$, et $\varepsilon' \in (0, \varepsilon)$, la probabilité de l'événement

$$p_n^N(\varepsilon) \leq \frac{e^{-\beta_n(\varepsilon-\varepsilon')}}{m_{\varepsilon'}} + \frac{r_i^*N + r_j^*y}{N^2}$$

est supérieure à $1 - e^{-y}$, avec $(i, j) = (1, 2)$ (et $M = e^{\Delta \cdot \text{osc}(V)}$) dans le cas des Δ_p bornés, et $(i, j) = (3, 4)$ dans le second cas.

On distingue clairement deux termes d'erreur. Le premier, $\left(\frac{e^{-\beta_n(\varepsilon-\varepsilon')}}{m_{\varepsilon'}}\right)$, est lié à la concentration de la mesure de Gibbs autour de l'ensemble des minimiseurs globaux de V . Le second, $\left(\frac{r_i^*N + r_j^*y}{N^2}\right)$, est lié à la concentration de la mesure d'occupation autour de la mesure de Gibbs, sa mesure limite. En plus du fait que le théorème 6 fournisse des stratégies de réglage qui assurent la performance des modèles ISA, la dernière inégalité de concentration explicite l'importance relative de chaque paramètre, notamment la précision stochastique y , le seuil t sur la proportion de particules possiblement hors de la zone d'intérêt, la température finale β_n et la taille de la population N . Une équation simple, déduite de ce dernier théorème, comme par exemple $\left(\frac{e^{-\beta_n(\varepsilon-\varepsilon')}}{m_{\varepsilon'}} = \frac{r_i^*N + r_j^*y}{N^2} = \frac{t}{2}\right)$ peut s'appliquer à un réglage global du modèle ISA, qui constitue en pratique une tâche difficile.

Une façon naturelle de choisir Δ_p est de s'intéresser au nombre d'itérations n_1 nécessaires pour passer de la température β_p à β_n , et de comparer ce nombre à n_2 , le nombre d'itérations nécessaires pour passer de β_p à β_q , puis de β_q à β_n , avec $\beta_p < \beta_q < \beta_n$. Grossièrement parlant, on a vu que la condition de convergence était $b_p \leq \frac{a}{g_p}$, donc

$$\begin{aligned} - n_1 &\simeq \left(\text{osc}(V) \cdot \Delta_{p,n} + \log\left(\frac{1}{a}\right)\right) \frac{e^{\overline{\Delta V}(k_0) \cdot \beta_n}}{\delta} \\ - n_2 &\simeq \left(\text{osc}(V) \cdot \Delta_{p,q} + \log\left(\frac{1}{a}\right)\right) \frac{e^{\overline{\Delta V}(k_0) \cdot \beta_q}}{\delta} + \left(\text{osc}(V) \cdot \Delta_{q,n} + \log\left(\frac{1}{a}\right)\right) \frac{e^{\overline{\Delta V}(k_0) \cdot \beta_n}}{\delta} \end{aligned}$$

où $\Delta_{p,q} := \beta_p - \beta_q$ pour $p > q$. Après un calcul approché simple, il vient

$$n_1 \leq n_2 \iff \Delta_{p,q} \Delta_{q,n} \geq \frac{\log\left(\frac{1}{a}\right)}{\text{osc}(V) \cdot \overline{\Delta V}(k_0)}.$$

Cette condition n'est pas très instructive dans le cas où $\Delta_p \rightarrow 0$ sauf peut-être en ce qu'elle montre que la décomposition de l'erreur $\eta_n^N - \eta_n = \sum_{p=0}^n \phi_{p,n}(\eta_p^N) - \phi_{p,n} \phi_p(\eta_{p-1}^N)$, sous-jacente à toute notre analyse, est peu adaptée au cas où $\Delta_p \rightarrow 0$ (cas qui peut être comparé au cas du temps continu). En revanche, elle est très intéressante dans le cas de sauts de température inverse constants. Dans cette situation, le paramètre critique Δ_β est donné par

$$\Delta_\beta = \sqrt{\frac{\log\left(\frac{1}{a}\right)}{\text{osc}(V) \cdot \overline{\Delta V}(k_0)}}.$$

2.4. UNE MÉTHODE SMC À SCHÉMA DE TEMPÉRATURE ADAPTATIF

Plus précisément, au dessus de Δ_β , l'algorithme doit réaliser trop d'étapes de MCMC pour stabiliser le système. A l'inverse, lorsque la variation de température inverse est trop petite en deçà de cette valeur, l'algorithme piétine dans son évolution en faisant trop d'étapes inutiles, et il devient difficile d'atteindre la mesure cible.

2.4 Une méthode SMC à schéma de température adaptatif

Comme nous le mentionnions dans l'introduction de ce chapitre, les stratégies de réglage des paramètres du SMC développées dans la section 2.3.2 sont du même ordre que les schémas de refroidissement logarithmiques des méthodes SA traditionnelles (voir [6, 35]). Mais à la différence de ces modèles SA, il faut souligner que les performances des modèles ISA ne sont pas basées sur un paramètre critique de température initiale. Un autre avantage de l'algorithme ISA, et de façon générale des méthodes SMC à schéma de température, est de fournir à chaque itération temporelle une approximation de la mesure cible pour une température donnée. En d'autres termes, la distribution de la population reflète la distribution de la mesure de Gibbs à cet instant donné. Du point de vue de la mise en oeuvre informatique, l'incrémentement du paramètre de température Δ_p joue un rôle important. Par exemple, si Δ_p est trop grand, le processus de sélection est dominé par une minorité de particules bien placées, et la grande majorité des particules disparaît. La diversité du nuage de particules est donc perdue, alors même qu'il s'agit d'un des principaux avantages de la méthode SMC. A l'inverse, si Δ_p est trop petit, l'algorithme procède à une sélection non sélective, c'est-à-dire inutile. L'algorithme perd ensuite du temps à effectuer des transitions de MCMC alors que le nuage de particules a déjà atteint son équilibre. Pour éviter ces deux phénomènes néfastes, il est donc crucial de trouver un compromis intelligent entre le maintien de la diversité et le soucis d'efficience.

Elaborer un tel compromis à l'avance est en réalité presque aussi difficile que de connaître la fonction V à l'avance. Il est donc assez naturel d'implémenter des stratégies adaptatives, qui dépendent de la variabilité et de l'adaptation des particules. Ces techniques sont d'usage courant dans les applications (voir par exemple [52, 76], [22, 41, 67]). Dans le domaine des algorithmes évolutionnaires, l'élaboration de stratégies de sélection adaptative est une question cruciale (voir par exemple [5]) dans la mise au point d'algorithmes performants. Dans le cas des méthodes SMC d'échantillonnage de mesures de Gibbs, les moyens les plus courants pour choisir Δ_p sont basés sur des critères simples comme l'espérance du nombre de particules tuées (voir la section 2.4.2 page 44), ou la variance des poids (ESS pour Effective Sample Size). Tous ces critères sont basés sur la même idée intuitive, à savoir réaliser une sélection raisonnable. Il en résulte que tous ces modèles adaptatifs tendent à donner des performances sensiblement similaires.

On pourra trouver dans [38] une formalisation générale des algorithmes SMC adaptatifs.

2.4. UNE MÉTHODE SMC À SCHÉMA DE TEMPÉRATURE ADAPTATIF

L'idée qui y est développée est de définir l'adaptation comme le choix des temps n où la sélection s'opère. Ces temps sont choisis selon des critères adaptatifs qui dépendent du nuage de particules courant, ou plus généralement de tout le passé du processus. Sous des conditions assez faibles sur ces critères, il est montré comment le processus adaptatif converge asymptotiquement vers un processus statique faisant intervenir des temps de sélection déterministes, lorsque la taille de la population tend vers l'infini. En montrant l'égalité des temps aléatoires et déterministes sur un événement de probabilité $\xrightarrow{N \rightarrow \infty} 1$, un théorème limite central est obtenu pour une large classe d'algorithmes SMC adaptatifs. Cependant, le comportement en n des estimées est explosif, ce qui rend difficile un contrôle uniforme en temps.

L'approche que nous développons dans cette section est différente. Elle se focalise sur des résultats de convergence non asymptotiques pour l'algorithme SMC défini dans la section 2.4.2 page 44. L'adaptation consiste ici à choisir l'incrément de β , soit Δ_{n+1}^N , de sorte que

$$\eta_n^N(e^{-\Delta_{n+1}^N \cdot V}) = \varepsilon$$

où $\varepsilon > 0$ est une constante donnée, et cela à chaque itération n . Nous montrons que le processus stochastique associé peut s'interpréter comme une perturbation d'un flot de FK limite.

2.4.1 Représentation de Feynman-Kac

Soit $V \in \mathcal{B}(E)$. Pour simplifier l'analyse, et sans perdre de généralité, on suppose $V_{\min} = 0$. Fixons $\varepsilon > 0$. Pour toute mesure $\mu \in \mathcal{P}(E)$, on pose la fonction λ_μ de la façon suivante :

$$\lambda_\mu = \begin{array}{ccc} [0, +\infty[& \longrightarrow &]0, 1] \\ x & \mapsto & \mu(e^{-x \cdot V}) \end{array}$$

Il s'agit clairement d'une fonction décroissante, convexe, infiniment dérivable et vérifiant $\lambda_\mu(0) = 1$. De plus, si $\mu(\{V = 0\}) = 0$, alors $\lambda_\mu(x) \longrightarrow 0$ quand $x \rightarrow +\infty$. On est dans ce cas en situation de pouvoir définir sa fonction réciproque κ_μ :

$$\kappa_\mu = \begin{array}{ccc}]0, 1] & \longrightarrow & [0, +\infty[\\ \varepsilon & \mapsto & x \text{ tel que } \mu(e^{-x \cdot V}) = \varepsilon \end{array}$$

Cette fonction est également décroissante, convexe et infiniment dérivable. Elle vaut 0 en $\varepsilon = 1$, et vérifie $\kappa_\mu(\varepsilon) \longrightarrow +\infty$ quand $\varepsilon \rightarrow 0^+$.

On se fixe désormais une mesure de référence m sur E qui vérifie par hypothèse $m(\{V = 0\}) = 0$. Considérons la suite de températures inverses $(\beta_n)_n$ et les mesures de Gibbs qui lui sont associées $\eta_n = \mu_{\beta_n} \propto e^{-\beta_n V} \cdot m$, définies par récurrence via l'équation :

$$\Delta_{n+1} := (\beta_{n+1} - \beta_n) = \kappa_{\eta_n}(\varepsilon). \quad (2.38)$$

De façon équivalente, cette relation peut être formulée

2.4. UNE MÉTHODE SMC À SCHÉMA DE TEMPÉRATURE ADAPTATIF

$$\lambda_{\eta_n}(\Delta_{n+1}) = \varepsilon \quad \text{ou} \quad \eta_n(e^{-\Delta_{n+1} \cdot V}) = \varepsilon.$$

L'objectif principal de l'algorithme étudié dans cette section est d'approcher ces mesures cibles. Formellement, (η_n) admet la structure de FK décrite dans la section 2.3.1, avec les potentiels $G_n(x) = e^{-\Delta_n \cdot V(x)}$, et des noyaux de Markov MCMC M_n . On note g_n , b_n , les oscillations et coefficients de Dobrushin correspondants à ces objets, et on note ϕ_n les transformations de FK associées.

La résolution de l'équation $\eta_n(e^{-\Delta_{n+1} \cdot V}) = \varepsilon$ peut être interprétée comme un moyen d'imposer une sorte de régularité théorique au flot de FK. En effet, d'après la formule (2.9) page 23 (et la définition (2.36) page 36), il est équivalent de trouver $\Delta_{n+1} \geq 0$ tel que

$$\frac{\gamma_{n+1}(1)}{\gamma_n(1)} = \varepsilon \quad \left(= \frac{Z_{\beta_n + \Delta_{n+1}}}{Z_{\beta_n}} \right).$$

Autrement dit, la suite Δ_n est définie de sorte que les constantes de normalisation $\gamma_n(1)$ croissent géométriquement selon la raison $\gamma_{n+1}(1)/\gamma_n(1) = \varepsilon$. Néanmoins, il est très important de remarquer que ces incréments Δ_n , d'existence certes rigoureusement définie, restent des quantités purement théoriques car construites à partir d'objet en pratique inconnus. Il en résulte en particulier que les potentiels G_n qui leur sont associés ne sont pas explicitement connus ni calculables directement.

2.4.2 Un processus particulaire adaptatif

La suite de mesures de probabilité cibles η_n étant posée, décrivons maintenant un algorithme SMC spécifique pour l'approcher. Tout comme le SMC traditionnel, nous approchons ces mesures η_n en simulant un système de particules en interaction $(\zeta_n)_n = (\zeta_n^1, \dots, \zeta_n^N)_n$ de taille N de sorte que

$$\eta_n^N = \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\zeta_n^i} \xrightarrow{N \rightarrow \infty} \eta_n.$$

On initialise ce système par N simulations indépendantes de la loi η_0 , puis on alterne des étapes de sélection et de mutation, comme il est décrit dans la section 2.1.3 page 24 pour l'algorithme classique. Mais à la différence du modèle SMC classique, la fonction G_{n+1} qui intervient dans la sélection est inconnue. On commence donc l'étape de sélection par le calcul d'un incrément empirique Δ_{n+1}^N défini par

$$\Delta_{n+1}^N := \kappa_{\eta_n^N}(\varepsilon)$$

ou $\lambda_{\eta_n^N}(\Delta_{n+1}^N) = \eta_n^N(e^{-\Delta_{n+1}^N \cdot V}) = \varepsilon$. Etant donné que la quantité $\eta_n^N(e^{-\Delta \cdot V}) = \frac{1}{N} \sum_{1 \leq i \leq N} e^{-\Delta \cdot V(\zeta_n^i)}$

est facile à calculer pour tout $\Delta \geq 0$, on peut déterminer Δ_{n+1}^N en pratique assez facilement,

2.4. UNE MÉTHODE SMC À SCHEMA DE TEMPÉRATURE ADAPTATIF

par exemple via une procédure de dichotomie. Si l'on considère les potentiels stochastiques suivants :

$$G_{n+1}^N = e^{-\Delta_{n+1}^N \cdot V}$$

alors chaque particule ζ_n^i évolue en une nouvelle particule $\widehat{\zeta}_n^i$ choisie aléatoirement selon la transition de sélection stochastique suivante :

$$S_{n+1, \eta_n^N}^N(\zeta_n^i, dx) := G_{n+1}^N(\zeta_n^i) \delta_{\zeta_n^i}(dx) + \left(1 - G_{n+1}^N(\zeta_n^i)\right) \Psi_{G_{n+1}^N}(\eta_n^N)(dx).$$

Dans la formule précédente, $\Psi_{G_{n+1}^N}(\eta_n^N)$ désigne la mesure empirique "corrigée" (ou pondérée) définie par

$$\Psi_{G_{n+1}^N}(\eta_n^N) = \sum_{j=1}^N \frac{G_{n+1}^N(\zeta_n^j)}{\sum_{k=1}^N G_{n+1}^N(\zeta_n^k)} \delta_{\zeta_n^j}.$$

On remarque que $V_{\min} = 0$ garantit $0 < G_{n+1}^N \leq 1$. L'étape de mutation consiste à réaliser les transitions de Markov $M_{n+1}^N(\widehat{\zeta}_n^i, \cdot)$, définies comme $M_{n+1}(\widehat{\zeta}_n^i, \cdot)$ en remplaçant β_{n+1} par $\beta_{n+1}^N = \beta_n^N + \Delta_{n+1}^N$. Ainsi, conditionnellement au nuage de particules précédent ζ_n , la nouvelle population de particules ζ_{n+1} a pour loi

$$\begin{aligned} & \text{Loi} \left(\zeta_{n+1}^1, \dots, \zeta_{n+1}^N \mid \zeta_n^1, \dots, \zeta_n^N \right) \\ &= \left(\delta_{\zeta_n^1} \cdot S_{n+1, \eta_n^N}^N \cdot M_{n+1}^N \right) \otimes \dots \otimes \left(\delta_{\zeta_n^N} \cdot S_{n+1, \eta_n^N}^N \cdot M_{n+1}^N \right). \end{aligned} \quad (2.39)$$

La définition de Δ_{n+1}^N s'interprète évidemment comme une approximation naturelle de la relation théorique (2.38). Mais parallèlement à cela, elle admet aussi une interprétation purement algorithmique. En effet, conditionnellement à la n -ième génération de particules $(\zeta_n^1, \dots, \zeta_n^N)$, la probabilité qu'une particule quelconque ζ_n^i soit acceptée lors de la sélection, c'est-à-dire non affectée par le mécanisme de recyclage, est donnée par $G_{n+1}^N(\zeta_n^i) = e^{-\Delta_{n+1}^N \cdot V(\zeta_n^i)}$. Il vient alors que l'espérance du nombre de particules acceptées vaut $\sum_i e^{-\Delta_{n+1}^N \cdot V(\zeta_n^i)}$. Or cette quantité vaut exactement $N \times \eta_n^N(e^{-\Delta_{n+1}^N \cdot V})$, qui est justement égale à $N \cdot \varepsilon$ par définition de Δ_{n+1}^N . Ainsi, ε est une approximation de la proportion de particules qui restent sur place pendant l'étape de sélection. Autrement dit, à chaque génération n , l'incrément Δ_{n+1}^N est choisi précisément de sorte que la sélection élimine moins de $(1 - \varepsilon) \cdot N$ particules. Comme nous le disions dans l'introduction de ce chapitre, ce type de réglage est très important en pratique, car il permet d'éviter des comportements dégénérés de l'algorithme.

2.4.3 Analyse de perturbation

Cette section se focalise sur l'analyse de la convergence d'un modèle adaptatif simplifié. Plus précisément, on considère la situation où les noyaux de transition de (2.39) sont donnés par les transitions limites M_{n+1} . L'analyse du modèle adaptatif (2.39) est beaucoup plus complexe, et notre approche ne s'applique pas directement pour étudier sa convergence.

2.4. UNE MÉTHODE SMC À SCHÉMA DE TEMPÉRATURE ADAPTATIF

Malgré l'adaptation, la suite η_n^N peut être analysée comme une perturbation aléatoire de la suite théorique η_n . Fixons un instant n et une population ζ_n supposée être le nuage de particules au temps n . Si l'on note ϕ_{n+1}^N la transformation de FK associée au potentiel G_{n+1}^N et au noyau de Markov M_{n+1} , alors la mesure η_{n+1}^N est par construction proche de $\phi_{n+1}^N(\eta_n^N)$. En particulier, par les inégalités de type Khintchine présentées dans [28], on a

$$\forall f \in \mathcal{B}_1(E), \quad \mathbb{E} \left(\left| \eta_{n+1}^N(f) - \phi_{n+1}^N(\eta_n^N)(f) \right|^p \mid \zeta_n \right)^{1/p} \leq \frac{B_p}{\sqrt{N}}, \quad (2.40)$$

avec les constantes B_p introduites en (2.2) page 21. Une remarque à la fois simple et importante sur les transformations de Boltzmann-Gibbs (voir définition en (1.2) page 2) est que pour toute mesure μ et pour tout couple de fonctions strictement positives G et \tilde{G} , on a

$$\psi_{\tilde{G}}(\mu) = \psi_G \left(\psi_{\frac{\tilde{G}}{G}}(\mu) \right).$$

Par conséquent, si l'on pose $H_{n+1}^N := \frac{G_{n+1}^N}{G_{n+1}}$, les transformations perturbées ϕ_{n+1}^N peuvent s'écrire en fonction des transformations théoriques ϕ_{n+1} :

$$\phi_{n+1}^N = \phi_{n+1} \circ \psi_{H_{n+1}^N}. \quad (2.41)$$

Si l'on se projette à l'avance dans une approche par récurrence, on a affaire au problème suivant. Soit η une mesure déterministe (η_n dans notre analyse) et $\hat{\eta}$ une mesure aléatoire (η_n^N dans notre analyse), proche de η au sens de la distance d_p définie en (2.3) page 21 (hypothèse de récurrence). Soient ensuite un noyau de Markov M et les potentiels

$$G = e^{-\kappa_\eta(\varepsilon).V}, \quad \hat{G} = e^{-\kappa_{\hat{\eta}}(\varepsilon).V}, \quad \hat{H} = \frac{\hat{G}}{G} \quad (2.42)$$

et soit enfin ϕ (respectivement $\hat{\phi}$) la transformation de FK associée au potentiel G (respectivement \hat{G}). Comment peut-on estimer $d_p(\phi(\eta), \hat{\phi}(\hat{\eta}))$ en fonction de $d_p(\eta, \hat{\eta})$?

Pour répondre à cette question, on se propose de réaliser une estimation en deux étapes. Premièrement, on estime la distance entre $\hat{\eta}$ et $\psi_{\hat{H}}(\hat{\eta})$ (lemme 7). Deuxièmement, on analyse les propriétés de stabilité de la transformation ϕ (lemme 8). Cette stratégie est résumée par le diagramme synthétique suivant.

$$\begin{array}{ccc} \eta & \xrightarrow{\phi} & \phi(\eta) \\ \hat{\eta} & & \\ \downarrow & & \\ \psi_{\hat{H}}(\hat{\eta}) & \xrightarrow{\phi} & \hat{\phi}(\hat{\eta}) \end{array}$$

Inscrite dans une récurrence globale, elle mène naturellement au schéma d'analyse du modèle adaptatif représenté figure 2.2.

2.4. UNE MÉTHODE SMC À SCHEMA DE TEMPÉRATURE ADAPTATIF

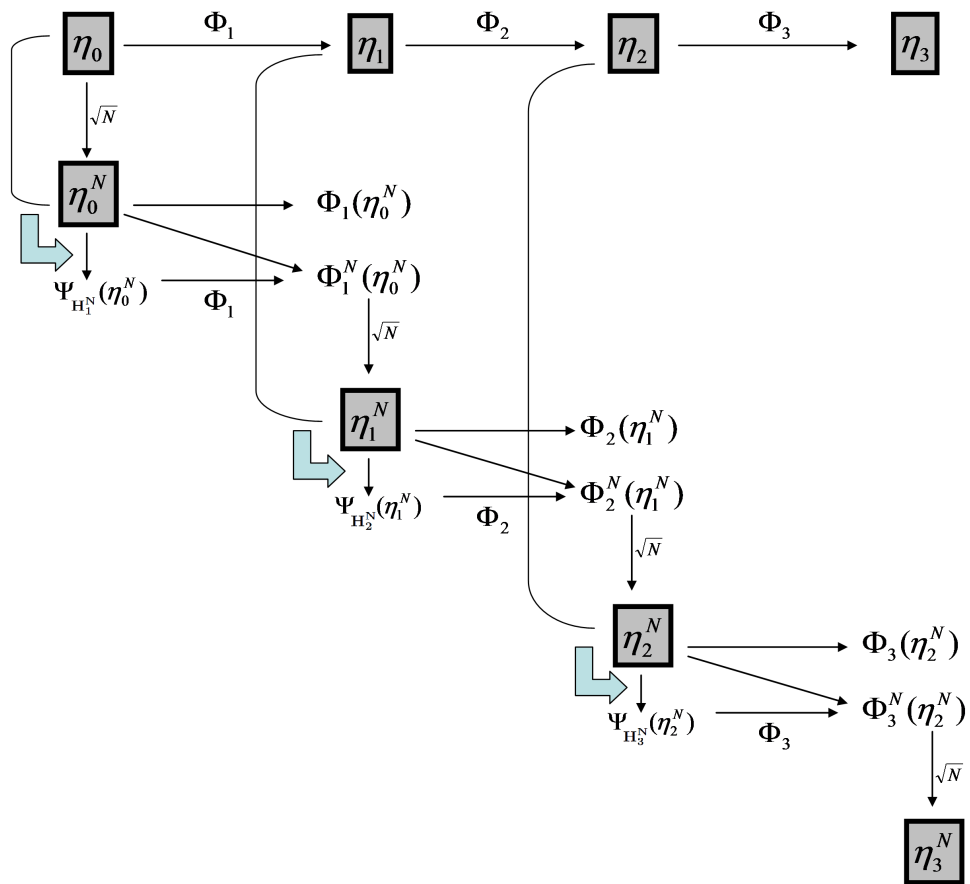


FIGURE 2.2 – Schéma d'analyse du modèle adaptatif

2.4. UNE MÉTHODE SMC À SCHÉMA DE TEMPÉRATURE ADAPTATIF

Lemme 7. Soit $\eta \in \mathcal{P}(E)$, $\hat{\eta} \in \mathcal{P}_\Omega(E)$ (ensemble des mesures aléatoires sur E) et soient G, \hat{G}, \hat{H} les fonctions strictement positives sur E définies par les équations (2.42). Si $\eta(\{V = 0\}) = \hat{\eta}(\{V = 0\}) = 0$ (p.s.), alors pour tout $\varepsilon > 0$, on a

$$d_p(\psi_{\hat{H}}(\hat{\eta}), \eta) \leq \frac{V_{\max} \cdot e^{\kappa_\eta(\varepsilon) \cdot V_{\max}}}{\eta(V)} \cdot d_p(\hat{\eta}, \eta).$$

Démonstration. Pour simplifier les notations, on pose $x := \kappa_\eta(\varepsilon)$ et $\hat{x} := \kappa_{\hat{\eta}}(\varepsilon)$. On commence par la décomposition suivante

$$\psi_{\hat{H}}(\hat{\eta})(f) - \hat{\eta}(f) = \frac{\hat{\eta}(\hat{H} \cdot f)}{\hat{\eta}(\hat{H})} - \hat{\eta}(f) = \frac{1}{\underbrace{\hat{\eta}(\hat{H})}_{A_1}} \underbrace{\hat{\eta} \left[\left(\hat{H} - \hat{\eta}(\hat{H}) \right) \cdot f \right]}_{A_2} \quad (2.43)$$

valable pour toute fonction $f \in \mathcal{B}(E)$. On remarque que $\hat{H} = \hat{G}/G = e^{(x-\hat{x}) \cdot V}$, ce qui permet la minoration suivante :

$$\hat{\eta}(\hat{H}) = \hat{\eta} \left(e^{(x-\hat{x}) \cdot V} \right) \geq \hat{\eta} \left(e^{-\hat{x} \cdot V} \right) = \varepsilon,$$

la dernière égalité venant de la définition de \hat{x} . On a donc montré : $|A_1| \leq \varepsilon^{-1}$. D'autre part, on a $\text{osc}(\hat{H}) = \left| e^{(x-\hat{x}) \cdot V_{\max}} - 1 \right|$, de sorte que

$$\begin{aligned} |A_2| &\leq \hat{\eta} \left(\left| \hat{H} - \hat{\eta}(\hat{H}) \right| \right) \cdot \|f\|_\infty \leq \text{osc}(\hat{H}) \cdot \|f\|_\infty \\ &\leq \left| e^{(x-\hat{x}) \cdot V_{\max}} - 1 \right| \cdot \|f\|_\infty. \end{aligned} \quad (2.44)$$

Intuitivement, la quantité $\hat{u} := \left(e^{(x-\hat{x}) \cdot V_{\max}} - 1 \right)$ est petite. Son analyse requiert un travail sur les fonctions λ_η et $\lambda_{\hat{\eta}}$. Fixons $\omega \in \Omega$. Si $x \geq \hat{x}$, alors on écrit

$$\left(\lambda_{\hat{\eta}}(x) - \underbrace{\lambda_\eta(x)}_{=\varepsilon = \lambda_{\hat{\eta}}(\hat{x})} \right) = \left(\lambda_{\hat{\eta}}(x) - \lambda_{\hat{\eta}}(\hat{x}) \right) = \int_{\hat{x}}^x -\lambda'_{\hat{\eta}}(s) ds.$$

Or, pour toute mesure $\mu \in \mathcal{P}(E)$ et pour tout $s \geq 0$, on a

$$\begin{aligned} -\lambda'_\mu(s) = \mu \left(V \cdot e^{-sV} \right) &\geq \mu \left(V \cdot e^{-sV_{\max}} \right) \\ &\geq \mu(V) \cdot e^{-sV_{\max}}. \end{aligned}$$

Donc

$$\begin{aligned} \left(\lambda_{\hat{\eta}}(x) - \lambda_\eta(x) \right) &\geq \hat{\eta}(V) \frac{-1}{V_{\max}} \left[e^{-sV_{\max}} \right]_{\hat{x}}^x \\ &= \hat{\eta}(V) \frac{e^{-xV_{\max}}}{V_{\max}} \left(e^{(x-\hat{x})V_{\max}} - 1 \right) \\ &= \hat{\eta}(V) \frac{e^{-xV_{\max}}}{V_{\max}} \hat{u}. \end{aligned}$$

2.4. UNE MÉTHODE SMC À SCHEMA DE TEMPÉRATURE ADAPTATIF

De façon symétrique, on montre

$$x \leq \hat{x} \implies (\lambda_\eta(x) - \lambda_{\hat{\eta}}(x)) \geq \hat{\eta}(V) \frac{e^{-xV_{\max}}}{V_{\max}} (-\hat{u}).$$

D'où la majoration presque sûre suivante :

$$|\hat{u}| \cdot \hat{\eta}(V) \frac{e^{-xV_{\max}}}{V_{\max}} \leq |\lambda_\eta(x) - \lambda_{\hat{\eta}}(x)|.$$

En utilisant la décomposition $\hat{\eta}(V) = \eta(V) + (\hat{\eta}(V) - \eta(V))$, et par des manipulations simples, on montre

$$|\hat{u}| \leq \frac{V_{\max} e^{xV_{\max}}}{\eta(V)} \underbrace{|\lambda_\eta(x) - \lambda_{\hat{\eta}}(x)|}_{A_3} + \underbrace{\frac{(\hat{\eta}(V) - \eta(V))}{\eta(V)}}_{A_4} \cdot \underbrace{|\hat{u}|}_{A_5}.$$

En considérant la norme L^p de la partie droite de cette inégalité, on constate les éléments suivants :

- $A_3 = (\eta - \hat{\eta})(e^{-x \cdot V})$. Donc comme $\text{osc}(e^{-x \cdot V}) \leq 1$, on a $\|A_3\|_p \leq d_p(\hat{\eta}, \eta)$;
- comme $\text{osc}(V) = V_{\max}$, on a $\|A_4\|_p \leq \frac{V_{\max}}{\eta(V)} \cdot d_p(\hat{\eta}, \eta)$;
- $A_5 = |\hat{u}| = e^{xV_{\max}} |e^{-\hat{x}V_{\max}} - e^{-xV_{\max}}| \leq e^{xV_{\max}}$.

En faisant les remplacements appropriés, on obtient

$$\|\hat{u}\|_p \leq \frac{2V_{\max} e^{xV_{\max}}}{\eta(V)} \cdot d_p(\hat{\eta}, \eta). \quad (2.45)$$

On combine alors ce résultat avec (2.43) et (2.44) pour obtenir

$$\|\psi_{\hat{H}}(\hat{\eta})(f) - \hat{\eta}(f)\|_p \leq \frac{2V_{\max} e^{xV_{\max}}}{\varepsilon \cdot \eta(V)} \cdot d_p(\hat{\eta}, \eta) \cdot \|f\|_\infty$$

Finalement, on passe de $\|f\|_\infty$ à $\frac{\text{osc}(f)}{2}$ en remarquant que $\psi_{\hat{H}}(\hat{\eta})(f) - \hat{\eta}(f)$ vaut 0 pour toutes les fonctions f constantes, et en considérant l'inégalité précédente avec $\tilde{f} = f - \frac{f_{\max} + f_{\min}}{2}$, fonction qui vérifie $\|\tilde{f}\|_\infty = \frac{\text{osc}(f)}{2}$.

Le lemme est alors bien démontré. \square

Lemme 8. Soit $\eta \in \mathcal{P}(E)$, $\hat{\eta} \in \mathcal{P}_\Omega(E)$, et soit ϕ la transformation de FK associée à la fonction strictement positive G et au noyau de Markov M . Si on pose $g := \sup_{x,y \in E} G(x)/G(y)$

et $b := \beta(M)$, alors on a

$$d_p(\phi(\hat{\eta}), \phi(\eta)) \leq g \cdot b \cdot d_p(\hat{\eta}, \eta).$$

2.4. UNE MÉTHODE SMC À SCHÉMA DE TEMPÉRATURE ADAPTATIF

Démonstration. On fixe $f \in \mathcal{B}(E)$. Partons de l'identité suivante :

$$\begin{aligned}\phi(\hat{\eta})(f) - \phi(\eta)(f) &= \frac{\hat{\eta}(G \times M.f)}{\hat{\eta}(G)} - \phi(\eta)(f) \\ &= \frac{\hat{\eta}(G \times [M.(f - \phi(\eta)(f))])}{\hat{\eta}(G)}.\end{aligned}$$

Soit $\tilde{f} = M.(f - \phi(\eta)(f))$. D'après la propriété (2.4) donnée page 22, \tilde{f} vérifie $\text{osc}(\tilde{f}) = \text{osc}(M.f) \leq b \cdot \text{osc}(f)$.

De plus, on a $\eta(G \times \tilde{f}) = \eta(G \times M.f) - \eta\left(G \times \frac{\eta(G \times M.f)}{\eta(G)}\right) = 0$. On écrit donc

$$\begin{aligned}\phi(\hat{\eta})(f) - \phi(\eta)(f) &= \frac{\hat{\eta}(G \times \tilde{f})}{\hat{\eta}(G)} - \underbrace{\frac{\eta(G \times \tilde{f})}{\eta(G)}}_{=0} = \frac{1}{\hat{\eta}(G)}(\hat{\eta} - \eta)(G \times \tilde{f}) \\ &= \frac{G_{\max}}{\hat{\eta}(G)}(\hat{\eta} - \eta)\left(\frac{G}{G_{\max}} \times \tilde{f}\right).\end{aligned}$$

Tout d'abord, on remarque que $\left| \frac{G_{\max}}{\hat{\eta}(G)} \right| \leq \frac{G_{\max}}{G_{\min}} \leq g$. Ensuite, la fonction \tilde{f} peut s'écrire

$$\tilde{f} = M.(f - \psi_G(\eta)(M.f)) = (M.f) - \psi_G(\eta)(M.f).$$

Sous cette forme (différence entre une fonction et sa moyenne), il est clair que $\tilde{f}_{\max} \geq 0$ et $\tilde{f}_{\min} \leq 0$. Dans ces conditions, $\text{osc}\left(\frac{G}{G_{\max}} \times \tilde{f}\right) \leq \text{osc}(\tilde{f}) \leq b \cdot \text{osc}(f)$. On conclut alors

$$\begin{aligned}\left(E \left| \frac{G_{\max}}{\hat{\eta}(G)}(\hat{\eta} - \eta)\left(\frac{G}{G_{\max}} \times \tilde{f}\right) \right|^p \right)^{1/p} &\leq g \cdot E \left[E \left| (\hat{\eta} - \eta)\left(\frac{G}{G_{\max}} \times \tilde{f}\right) \right|^p \right]^{1/p} \\ &\leq g \cdot b \cdot d_p(\hat{\eta}, \eta) \cdot \text{osc}(f).\end{aligned}$$

ce qui clôt la démonstration. □

2.4.4 Résultats de convergence non asymptotiques

Cette section est essentiellement dédiée à la preuve du théorème 3 énoncé page 20. On déduit également des inégalités de concentration pour le modèle de SMC adaptatif. Commençons donc avec la preuve du théorème 3.

Preuve du théorème 3 :

On fixe $p \geq 1$ et on pose $\tilde{e}_n = \sum_{k=0}^n \prod_{i=k+1}^n b_i g_i (1 + c_i)$. Remarquons que cette suite peut aussi être définie par l'initialisation $\tilde{e}_0 = 1$ et la relation de récurrence

$$\tilde{e}_{n+1} = 1 + g_{n+1} b_{n+1} (1 + c_{n+1}) \cdot \tilde{e}_n$$

2.4. UNE MÉTHODE SMC À SCHEMA DE TEMPÉRATURE ADAPTATIF

On considère également pour tout n la quantité suivante :

$$e_n := \frac{\sqrt{N}}{2B_p} \cdot \sup_{f \in \mathcal{O}_1(E)} \|\eta_n^N(f) - \eta_n(f)\|_p.$$

On raisonne par récurrence (voir schéma de la figure 2.3) pour montrer que la proposition $\mathbf{HR}(n) = \{e_n \leq \tilde{e}_n\}$ est vraie pour tout n . Etant donné que la première mesure empirique η_0^N est obtenue par N simulations indépendantes de la loi η_0 , $\mathbf{HR}(0)$ est directement donnée par les inégalités de Khintchine. Supposons maintenant que $\mathbf{HR}(n)$ est vérifiée.

D'après l'identité (2.41) établie page 46, on peut écrire la décomposition suivante.

$$\begin{aligned} \eta_{n+1}^N - \eta_{n+1} &= (\eta_{n+1}^N - \phi_{n+1}^N(\eta_n^N)) + (\phi_{n+1}^N(\eta_n^N) - \eta_{n+1}) \\ &= \underbrace{(\eta_{n+1}^N - \phi_{n+1}^N(\eta_n^N))}_{A_1} + \underbrace{(\phi_{n+1}^N(\psi_{H_{n+1}^N}(\eta_n^N)) - \phi_{n+1}(\eta_n))}_{A_2} \end{aligned} \quad (2.46)$$

Conditionnellement à ζ_n , et en utilisant l'inégalité (2.40) rappelée page 46, on sait que

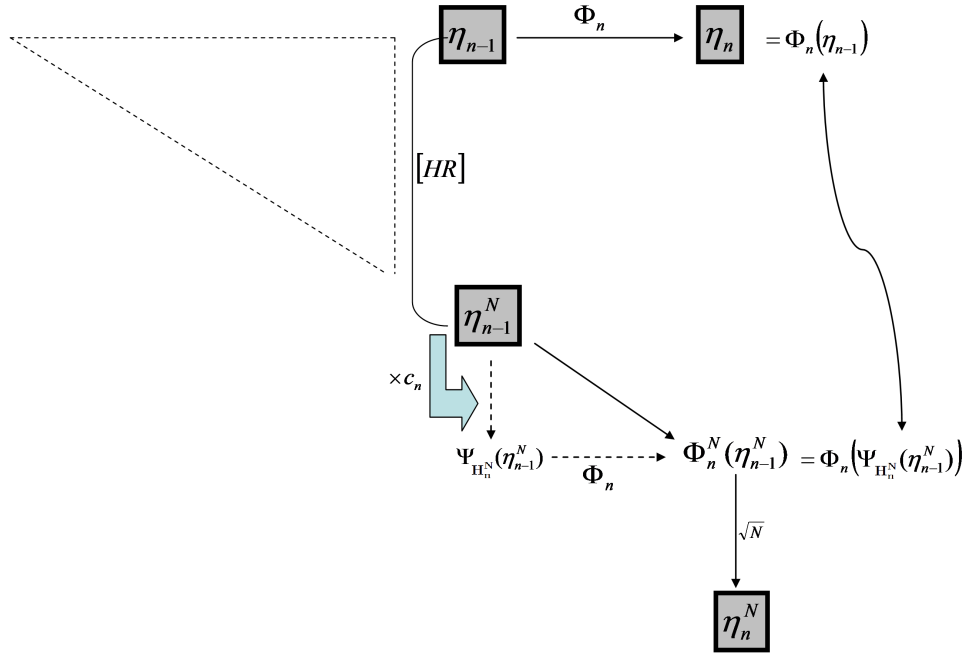


FIGURE 2.3 – Schéma de l'étape d'héritité

pour toute fonction $f \in \mathcal{O}_1(E)$, on a $\frac{\sqrt{N}}{2B_p} \|A_1(f)\|_p \leq 1$. Pour estimer A_2 , on commence par décomposer $(\psi_{H_{n+1}^N}(\eta_n^N) - \eta_n)$ de la façon suivante :

$$\psi_{H_{n+1}^N}(\eta_n^N) - \eta_n = \underbrace{(\psi_{H_{n+1}^N}(\eta_n^N) - \eta_n^N)}_{Q_1} + \underbrace{(\eta_n^N - \eta_n)}_{Q_2}.$$

2.4. UNE MÉTHODE SMC À SCHÉMA DE TEMPÉRATURE ADAPTATIF

On a par hypothèse de récurrence que $\frac{\sqrt{N}}{2B_p} \cdot \sup_{f \in \mathcal{O}_1(E)} \|Q_2(f)\|_p \leq \tilde{e}_n$. Cette inégalité, combinée avec le lemme 7, donne

$$\frac{\sqrt{N}}{2B_p} \cdot \sup_{f \in \mathcal{O}_1(E)} \|Q_1(f)\|_p \leq c_{n+1} \cdot \tilde{e}_n.$$

Ainsi, les mesures $\psi_{H_{n+1}^N}(\eta_n^N)$ et η_n vérifient

$$\frac{\sqrt{N}}{2B_p} \cdot \sup_{f \in \mathcal{O}_1(E)} \left\| \psi_{H_{n+1}^N}(\eta_n^N)(f) - \eta_n(f) \right\|_p \leq (1 + c_{n+1}) \cdot \tilde{e}_n.$$

Par application du lemme 8 on a également

$$\frac{\sqrt{N}}{2B_p} \cdot \sup_{f \in \mathcal{O}_1(E)} \|A_2(f)\|_p \leq g_{n+1} b_{n+1} (1 + c_{n+1}) \cdot \tilde{e}_n.$$

En revenant à la décomposition (2.46), on conclut que $e_{n+1} \leq 1 + g_{n+1} b_{n+1} (1 + c_{n+1}) \cdot \tilde{e}_n = \tilde{e}_{n+1}$, ce qui achève la démonstration du théorème. \square

Nous sommes maintenant en position d'obtenir une condition suffisante pour des résultats de concentration et de bornes L^p uniformes en temps n .

Corollaire 1. *Si la condition $b_n g_n (1 + c_n) \leq a$ est vérifiée pour un certain $a < 1$ et ce pour tout n , alors on a les bornes d'erreur suivantes*

$$d_p(\eta_n^N, \eta_n) \leq \frac{B_p}{2(1-a)\sqrt{N}}. \quad (2.47)$$

pour tout p , avec les constantes B_p introduites en (2.2) page 21. De plus, pour toute fonction $f \in \mathcal{B}_1(E)$, on a les inégalités de concentration suivantes :

$$\forall s \geq 0, \quad \mathbb{P}\left(|\eta_n^N(f) - \eta_n(f)| \geq s\right) \leq r_1(\sqrt{N} \cdot s) e^{-r_2 N s^2} \quad (2.48)$$

et

$$\forall y \geq 1, \quad \mathbb{P}\left(|\eta_n^N(f) - \eta_n(f)| \geq \frac{r(1 + \sqrt{y})}{\sqrt{N}}\right) \leq e^{-y}, \quad (2.49)$$

avec les paramètres

$$\begin{cases} r_1 &= e^{1/2}(1-a) \\ r_2 &= \frac{1}{2}(1-a)^2 \\ r &= \frac{2}{1-a}. \end{cases}$$

2.4. UNE MÉTHODE SMC À SCHEMA DE TEMPÉRATURE ADAPTATIF

Démonstration. L'inégalité (2.47) est une conséquence directe du théorème 3. Fixons $n, f \in \mathcal{B}_1(E)$ et posons

$$X := |\eta_n^N(f) - \eta_n(f)| \quad \text{et} \quad \epsilon_N := \frac{1}{(1-a)\sqrt{N}}.$$

Sous ces notations, on a $\|X\|_p \leq B_p \cdot \epsilon_N$ pour tout $p \geq 1$. Fixons $s \geq 0$. D'après l'inégalité de Markov, pour tout $t \geq 0$ on a

$$\mathbb{P}(X \geq s) = \mathbb{P}(e^{tX} \geq e^{ts}) \leq e^{-st} \mathbb{E}(e^{tX}). \quad (2.50)$$

En utilisant les formules explicites (2.2) données page 21, on estime la transformée de Laplace $\mathbb{E}(e^{tX})$.

$$\begin{aligned} \mathbb{E}(e^{tX}) &= \sum_{p \geq 0} \mathbb{E}\left(\frac{t^p X^p}{p!}\right) \\ &\leq \sum_{p \geq 0} \frac{t^{2p} \epsilon_N^{2p} (2p)!}{(2p)! 2^p \cdot p!} + \sum_{p \geq 0} \frac{t^{2p+1} \epsilon_N^{2p+1} (2p+1)!}{(2p+1)! 2^p \cdot p! \sqrt{2p+1}} \\ &\leq (1 + t\epsilon_N) e^{\frac{t^2 \epsilon_N^2}{2}} \end{aligned}$$

En prenant l'inégalité (2.50) avec $t = \frac{1}{\epsilon_N} \left(\frac{s}{\epsilon_N} - 1\right)$, on obtient

$$\mathbb{P}(X \geq s) \leq \frac{s}{\epsilon_N} e^{-\frac{1}{2} \left[\left(\frac{s}{\epsilon_N}\right)^2 - 1\right]},$$

qui est équivalent à la première inégalité de concentration (2.48). Pour la seconde, on utilise l'inégalité

$$ue^{-\frac{u^2}{2}} = e^{-(\frac{u^2}{2} - \log u)} \leq e^{-(\frac{u^2}{2} - u + 1)}$$

avec $u = \frac{s}{\epsilon_N}$. Il est alors aisé de montrer que pour tout $y \geq 1$ et $u \geq 2$, l'équation $y = \frac{-u^2}{2} - u + 1$ est équivalente à $u = 1 + \sqrt{1 + 4(y-1)} \leq 2(1 + \sqrt{y})$. Ainsi, $\forall y \geq 1$:

$$\mathbb{P}(X \geq 2\epsilon_N(1 + \sqrt{y})) \leq e^{-y},$$

ce qui est équivalent à (2.49), d'où la fin de la démonstration. \square

Conclusion de la partie théorique et remarques

Ce chapitre constitue une contribution aux travaux d'analyse des méthodes SMC. D'un point de vue global, il comporte essentiellement deux idées nouvelles.

2.4. UNE MÉTHODE SMC À SCHÉMA DE TEMPÉRATURE ADAPTATIF

La première est qu'il est possible de contrôler les quantités de stabilité des transformations des semigroupes de Feynman-Kac $\phi_{p,n}$ à partir d'informations sur les transformations locales ϕ_p . A ce titre, le lemme 4, page 29, est particulièrement central. L'essentiel des résultats des sections 2.2 et 2.3 est obtenu par ses conséquences, ou sa combinaison avec d'autres résultats. Cette approche de "dépliage" du semigroupe est une approche parmi d'autres pour quantifier sa stabilité, mais elle s'avère particulièrement adaptée pour l'étude des modèles à schéma de température, et plus généralement aux modèles de flots de mesures du type

$$\eta_n(dx) \propto H_n(x)dx = \left(\prod_{k=1}^n G_k(x) \right) dx$$

où les G_k sont les potentiels de sélection, et où les noyaux de Markov sont choisis de façon à laisser stables ces mesures et explorer un ensemble E . Nous verrons dans les chapitres suivants que ces modèles sont d'une très grande utilité, notamment dans l'échantillonnage de loi a posteriori complexes.

La seconde concerne l'adaptation. En substance, la section 2.4 propose une méthodologie assez générale pour l'analyser, dès lors que la fonction de sélection peut se mettre sous la forme d'une fonction déterministe de la population courante :

$$G_{n+1}^N = f_n(\eta_n^N).$$

Le point sensible, et surtout spécifique à notre approche, est le lemme 7, page 48, qui quantifie la perturbation selon la norme L^p dans le cas des modèles à schéma de température. Plusieurs remarques peuvent néanmoins être faites sur ce travail :

- Premièrement, la constante intervenant dans le lemme 7 n'est à mon sens pas satisfaisante : il me paraît clair intuitivement que cette constante doit tendre vers 0 lorsque le pas de température théorique tend vers 0. C'est un problème d'autant plus important qu'on comprend dans la suite que cette constante est centrale, car elle mène à la définition des c_n qui sont les véritables quantifieurs de l'erreur dans les bornes finales.
- Deuxièmement, il est dommage de ne pas avoir réussi à obtenir une formulation du théorème 3 en fonction des quantités du semigroupe $(b_{p,n}, g_{p,n})$. La raison principale est l'absence (à ma connaissance) d'un résultat du type

$$d_p(\phi(\hat{\mu}_1), \phi(\hat{\mu}_2)) \leq g \cdot b \cdot d_p(\hat{\mu}_1, \hat{\mu}_2) \tag{2.51}$$

pour des mesures $\hat{\mu}_1, \hat{\mu}_2$ toutes deux aléatoires (ce qui est le cas de η_n^N et $\psi_{H_{n+1}^N}(\eta_n^N)$).

- Troisièmement, on utilise l'hypothèse très contestable selon laquelle on est capable pour chaque génération n de simuler un noyau de Markov laissant stable η_n , c'est-à-dire la mesure de Gibbs associée à β_n . En réalité l'utilisateur n'a pas accès à β_n mais seulement à son estimateur β_n^N . Il y a ici une approximation par rapport à l'implémentation pratique, intuitivement acceptable, mais qui mériterait un approfondissement théorique. Une parade technique consisterait par exemple à discrétiser les valeurs permises à β_n et β_n^N . En effet, en combinant l'estimation L^p (2.45) page 49 avec la technique (très générale)

2.4. UNE MÉTHODE SMC À SCHÉMA DE TEMPÉRATURE ADAPTATIF

par transformée de Laplace développée page 53, on obtient facilement une inégalité de concentration sur $(\Delta_n^N - \Delta_n)$ en e^{-N} . Or, sur un espace de valeurs discret, cette inégalité se transforme en une égalité $\beta_n^N = \beta_n$ pour un événement de probabilité $\xrightarrow{N \rightarrow \infty} 1$, ce qui signifie bien que l'algorithme étudié coïncide avec l'algorithme implémenté. Cette logique, proche de celle de [38], ouvre la voie à un TLC, mais souffre du même inconvénient dans le cas non asymptotique : le comportement explosif en n .

Rappelons enfin que toute notre étude repose sur des hypothèses de contraction au sens de la distance en variation totale, autrement dit au sens de la norme infinie. Ces modèles de convergence forte ont le désavantage logique de mener à des conditions restrictives pour assurer la bonne performance de l'algorithme, notamment sur le coefficient de Dobrushin des transitions markoviennes, et par suite, sur les schémas d'itération pour les modèles à schéma de température. Mais symétriquement, leur simplicité et la force de leurs hypothèses permettent l'obtention de résultats plus généraux et surtout plus ambitieux. On l'a vu dans ce chapitre via les bornes L^p et inégalités de concentration non asymptotiques, y compris dans le cas adaptatif, encore très peu traité dans la littérature, ainsi que leurs versions uniformes en temps. Ainsi, bien que ces hypothèses fortes les éloignent des applications pratiques, ces modèles forment de façon générale un cadre simple, permettant de mettre en évidence des quantités typiques, des dépendances vis-à-vis de tel ou tel paramètre, des structures et des schémas d'analyse des modèles particuliers. En plus des résultats obtenus, leur développement constitue une base de référence en vue d'éventuelles approches plus subtiles ou plus spécifiques à une application donnée.

2.4. UNE MÉTHODE SMC À SCHÉMA DE TEMPÉRATURE ADAPTATIF

Chapitre 3

Théorie et pratique des méthodes SMC : illustrations

Introduction

L'analyse théorique des méthodes SMC développée dans le chapitre 2 met en évidence deux paramètres théoriques antagonistes. Le premier,

$$g_n := \sup_{x,y \in E} \frac{G_n(x)}{G_n(y)}$$

concerne les fonctions de sélection G_n de l'algorithme (voir l'introduction aux méthodes SMC section 1.3 page 10 pour les notations), et quantifie l'instabilité du système, soit intuitivement la distance entre les mesures de probabilité η_{n-1} et $\psi_{G_n}(\eta_{n-1})$. Plus la valeur de g_n est élevée, plus l'étape de sélection est "hasardeuse" (variance élevée), en tant qu'approximation particulière de la pondération par le potentiel G_n . Le second paramètre,

$$b_n = \beta(M_n) := \sup_{\substack{x,y \in E \\ A \subset E}} |M_n(x, A) - M_n(y, A)|$$

est le coefficient de Dobrushin du noyau de Markov M_n . C'est au contraire un quantificateur de la stabilité du système dans la mesure où il estime la contraction de l'application $\mu \rightarrow \mu.M_n$ et donc la capacité de l'algorithme à oublier ses erreurs. Plus la valeur de b_n est petite, plus la distance entre la mesure empirique courante $\hat{\eta}_{n-1}$ (mesure empirique associée aux particules sélectionnées $\hat{\zeta}_{n-1}^1, \dots, \hat{\zeta}_{n-1}^N$) et la mesure théorique $\psi_{G_n}(\eta_{n-1})$ se trouve réduite.

Les résultats de convergence proposés exigent sans surprise un compromis entre ces deux logiques antagonistes. On a vu, notamment sur l'exemple de la simulation des mesures de Gibbs (voir par exemple le théorème 6 page 40), que les conditions pouvaient s'avérer très lourdes (schémas d'itération exponentiels) et donc difficilement acceptables en pratique. D'autre part, ces quantités sont loin d'être triviales à calculer. Il n'existe par exemple à

notre connaissance aucune méthode générale pour calculer ou estimer le coefficient de Dobrushin d'un noyau de Markov. De même, le calcul de l'oscillation au sens multiplicatif d'une fonction que l'on sait seulement calculer ponctuellement constitue en tant que tel un problème d'exploration aveugle à part entière. Dans le cas particulier des mesures de Gibbs liées à un potentiel V , on a vu dans la section 2.3 (notamment via (2.37) page 39) que ces estimations étaient possibles, mais nécessitaient une connaissance sur le comportement du noyau de proposition utilisé dans le Metropolis-Hastings et sur la quantité $osc(V)$.

Pour toutes ces raisons, les outils théoriques développés se montrent en général insuffisants pour élaborer une méthode SMC efficace. Ce chapitre est donc dédié à cette problématique, par le biais d'études de cas précis. Dans la section 3.1, nous proposons deux exemples artificiels pour analyser le lien entre théorie et pratique : la simulation d'une mesure de Gibbs sur un espace fini ($|E| = 10$), et la simulation d'un mélange dynamique de gaussiennes dans un espace borné ($\dim(E)=2$). Au sein de ce cadre contrôlé, où toutes les quantités relatives à l'analyse du SMC sont calculables ou estimables directement, nous montrons la pertinence de nos résultats, ainsi que leurs limites. A l'inverse de cette démarche d'expérience théorique, nous abordons ensuite dans la section 3.2 l'algorithme SMC appliqué à la résolution d'un premier problème inverse concret en électromagnétisme. Il consiste à déterminer des propriétés électromagnétiques d'un matériau multicouche 1D à partir de mesures de coefficients de réflexion. Dans le cadre d'une approche bayésienne, le SMC s'avère efficace pour échantillonner une loi a posteriori fortement multimodale sur un ensemble multidimensionnel. Cependant, dans la mesure où la modélisation directe du problème fait intervenir une fonction déterministe f_{MD} très difficile à analyser a priori, on est placé dans une situation d'exploration aveugle d'un ensemble E . Autrement dit, les connaissances a priori pour l'application d'une méthode SMC se limitent essentiellement aux trois hypothèses de base que nous donnions en (1.5) page 11 : savoir simuler la loi initiale η_0 , savoir calculer ponctuellement les potentiels $G_n(x)$, $\forall x \in E$, et savoir simuler les noyaux de Markov $M_n(x, \cdot)$, $\forall x \in E$. Dans cette situation, les critères théoriques sont inutilisables, et les bonnes performances ne sont pas rigoureusement garanties. Nous montrons cependant comment des stratégies pratiques de réglages du SMC, basées notamment sur des critères adaptatifs dans les étapes de sélection et mutation, permettent d'aboutir à un comportement stable de l'algorithme et à des performances satisfaisantes dans un certain nombre de cas.

3.1 Exemples "jouets"

Nous consacrons cette section à l'illustration des résultats théoriques du chapitre 2, basés sur l'oscillation des potentiels et sur le coefficient de Dobrushin des transitions markoviennes. En construisant des exemples artificiels simples, on se met dans une situation où tout est calculable, ce qui permet d'estimer leur pertinence et leurs limites.

3.1. EXEMPLES "JOUETS"

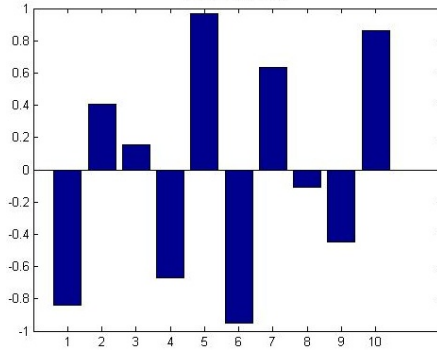


FIGURE 3.1 – Potentiel V

3.1.1 Illustration sur un espace fini, limites de l’approche Dobrushin

En général, le coefficient de Dobrushin est une quantité délicate à calculer, ou même à estimer. Mais dans le cas d’un espace d’état E fini, où les noyaux de Markov M_n peuvent être représentés par des matrices de transition K_n , ce coefficient se calcule de façon exacte par la formule :

$$\beta(M_n) = \frac{1}{2} \sup_{i,j} \sum_{l=1}^d |K_n(i,l) - K_n(j,l)|$$

où d est le cardinal de E .

On se propose donc d’illustrer nos théorèmes de réglage sur l’exemple simple de la simulation d’une mesure de Gibbs sur un espace à 10 éléments, par exemple $E = \{1, 2, \dots, 10\}$. On se fixe le potentiel $V : x \mapsto \cos(10x)$, représenté sur les entiers sur la figure 3.1 et une “température” finale $\beta_{final} = 6$. La suite β_n augmente linéairement de 0 à β_{final} en 6 itérations. Dans cette situation, rappelons que les mesures de Gibbs associées à ces objets sont les

$$\eta_n(dx) = \frac{1}{Z_{\beta_n}} e^{-\beta_n V(x)} dx$$

où les Z_{β_n} sont les constantes de normalisation adéquates pour garantir une masse totale égale à 1. Cette suite peut être mise sous la forme d’un flot de FK (voir la section 2.1.2 page 22) associé aux potentiels $G_n = e^{-(\beta_n - \beta_{n-1})V}$ et à des transitions markoviennes M_n choisies comme étant des noyaux de MCMC vis-à-vis des lois cibles courantes, c’est-à-dire laissant stables les mesures η_n (voir page 37). Le paramètre g_n est totalement explicite :

$$g_n = e^{(\beta_n - \beta_{n-1}) \cdot \text{osc}(V)} \quad \text{où} \quad \text{osc}(V) := V_{\max} - V_{\min}.$$

Les noyaux de Markov utilisés sont les noyaux du recuit simulé (voir la définition (1.4) page 4) associés au potentiel et températures précédentes, ainsi qu’au noyau de proposition suivant : $i \rightarrow i + 1$ avec probabilité $\frac{1}{2}$, $i \rightarrow i - 1$ avec probabilité $\frac{1}{2}$, tous les entiers étant ici considérés modulo 10.

3.1. EXEMPLES "JOUETS"

Comme les incréments de température sont constants, les g_n définis précédemment, sont constants. Plus précisément : $g_n = 6.803$. Dans ces conditions, et pouvant à loisir calculer les coefficients de Dobrushin des noyaux concernés, on choisit les m_n , nombres d'itérations du noyau de Metropolis-Hastings, *de sorte à avoir* $b_n g_n < 1$. Sont obtenus ainsi, pour chacune des 6 itérations, les nombres suivants :

$$\begin{array}{lll} m_1 = 30 & m_2 = 141 & m_3 = 740 \\ m_4 = 4054 & m_5 = 22597 & m_6 = 126773 \end{array}$$

réglage qui donne les produits $b_n g_n$ décrits sur la figure 3.2.

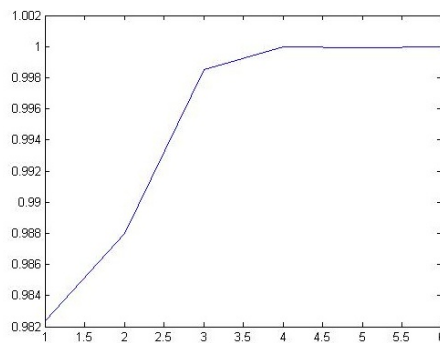


FIGURE 3.2 – Produits $b_n g_n$

Avec $N = 1000$ particules, l'algorithme donne le résultat de la figure 3.3(a), les bâtons bleus représentant la mesure empirique fournie par l'algorithme, et les rouges représentant la mesure théorique. Sans surprise, le résultat est satisfaisant, et la différence entre les deux mesures est de l'ordre de grandeur d'une erreur d'échantillonnage avec $N = 1000$. Cependant, on comprend les limites de cette approche en prenant le réglage constant $m_n = 10$ qui donne des résultats très similaires, présentés sur la figure 3.3(b).

Nos théorèmes de réglages donnent des conditions suffisantes de convergence, et non pas des conditions nécessaires, le coefficient de Dobrushin étant *une* façon d'estimer le mélange d'un noyau de Markov. D'autre part, on s'aperçoit, même sur un cas simple, que le nombre d'itérations demandé peut prendre de très grandes valeurs, incompatibles avec des objectifs de temps de calcul raisonnables.

3.1.2 Illustration sur un exemple en 2D

On se propose ici de construire un exemple simple où l'approche Dobrushin, et les réglages proposés trouvent toute leur pertinence. Soit E le rectangle $[0, 3] \times [0, 4]$ vu comme un tore (le point $(3 - \varepsilon, 2)$ étant voisin du point $(0 + \varepsilon, 2)$, par exemple). On définit les trois points suivants :

$$x_1 = (2, 3) \quad x_2 = (1, 1) \quad x_3 = (1, 3)$$

3.1. EXEMPLES "JOUETS"

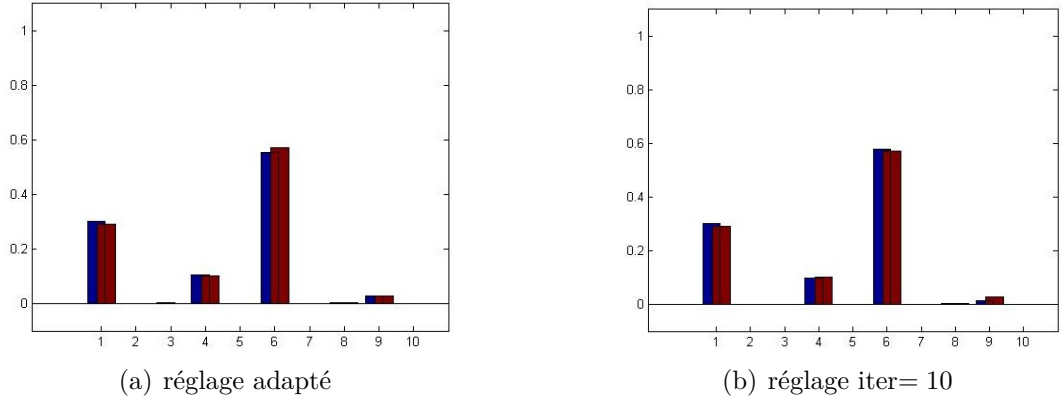


FIGURE 3.3 – mesures empirique/théorique

Les densités cibles η_n sont posées comme étant des mélanges de trois gaussiennes centrées respectivement en x_1 , x_2 et x_3 , et d'écart type commun $\sigma = 0.4$, soit

$$\eta_n(dx) \propto \left[A_1(n)e^{-\frac{\|x-x_1\|_2^2}{2\sigma^2}} + A_2(n)e^{-\frac{\|x-x_2\|_2^2}{2\sigma^2}} + A_3(n)e^{-\frac{\|x-x_3\|_2^2}{2\sigma^2}} \right] dx,$$

où le symbole \propto signifie la proportionnalité (une mesure de probabilité peut être définie à une constante multiplicative près sans ambiguïté). Tout l'intérêt réside dans les variations des pondérations de chacune des trois gaussiennes A_1 , A_2 et A_3 en fonction de n . On pose en effet, pour n variant de 1 à 150, $A_1(n)$ l'application qui évolue linéairement de 1 à 0.1, A_2 de 0.1 à 1, et A_3 constante à 1, pour obtenir une suite de densités (η_n) dont les étapes 1, 50, 100 et 150 sont représentées sur la figure 3.4.

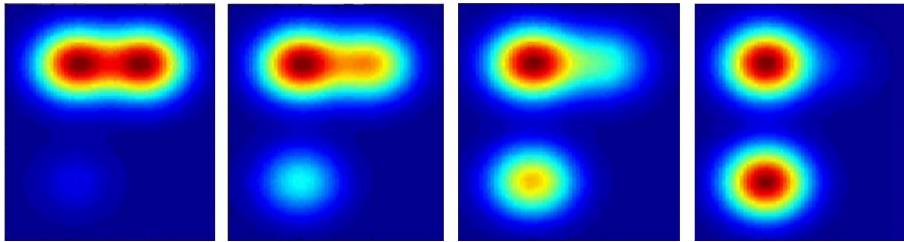


FIGURE 3.4 – Densités théoriques pour $n = 1, 50, 100$ et 150 . Une gaussienne reste constante autour de x_3 , une disparaît autour de x_1 et une apparaît autour de x_2 .

Comme toute suite de mesures à densités strictement positives, (η_n) peut être artificiellement mise sous la forme d'un flot de Feynman-Kac. En posant les fonctions

$$f_n(x) := A_1(n)e^{-\frac{\|x-x_1\|_2^2}{2\sigma^2}} + A_2(n)e^{-\frac{\|x-x_2\|_2^2}{2\sigma^2}} + A_3(n)e^{-\frac{\|x-x_3\|_2^2}{2\sigma^2}},$$

3.1. EXEMPLES "JOUETS"

pour avoir $\eta_n(dx) \propto f_n(x)dx$, on remarque que si l'on définit les fonctions de sélection G_n par : $G_n := \frac{f_{n+1}}{f_n}$, on vérifie bien la relation $\eta_{n+1} = \Psi_{G_n}(\eta_n)$. D'autre part, si l'on définit des noyaux de Markov M_n comme étant des noyaux de Metropolis-Hastings associés aux fonctions f_n (voir la définition (1.3) page 3), ces noyaux laissent invariant les η_n au sens où $\eta_n \cdot M_n = \eta_n$. La suite (η_n) suit donc le flot de Feynman-Kac associé aux fonctions de sélection G_n et aux noyaux M_n , itérés m_n fois. Comme noyau de proposition $K(x, dy)$, à l'intérieur des Metropolis-Hastings, on choisit des gaussiennes¹ centrées en x et de variance $\sigma_{prop}^2 \cdot Id$.

Ce contexte étant posé, on cherche à estimer le coefficient de Dobrushin de nos noyaux M_n . Plutôt que d'utiliser la majoration théorique de [6] que nous rappelions en (2.37) page 39, nous avons préféré un calcul approché obtenu par discrétisation de l'espace d'état (quadrillage simple), et application de la formule dans le cas d'un ensemble E fini à d éléments :

$$\beta(M_n) = \frac{1}{2} \sup_{i,j} \sum_{l=1}^d |K_n(i, l) - K_n(j, l)|$$

où K_n est la matrice de transition associée à M_n . Ces coefficients dépendent du σ_{prop} choisi pour les propositions gaussiennes, et du nombre d'itérations m_n . Pour illustrer les résultats théoriques, nous proposons deux réglages différents pour cet algorithme, les évaluations étant à chaque fois réalisées avec un nombre de particules $N = 800$.

Le premier réglage consiste à ne pas itérer les noyaux de Markov, soit $m_n = 1$, et à prendre des propositions gaussiennes "étroites", soit $\sigma_{prop} = 0.1$. En calculant les oscillations g_n des G_n et les coefficients de Dobrushin b_n approchés, on obtient les produits $b_n \cdot g_n$ représentés sur la figure 3.5,

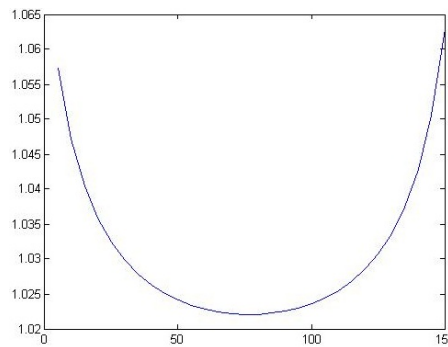


FIGURE 3.5 – produits $b_n g_n$ cas 1

le point remarquable étant que ces produits sont strictement supérieurs à 1. Les résultats de l'algorithme aux étapes 1, 50, 100 et 150 sont donnés sur la figure 3.6.

1. en fait, "pseudo-gaussiennes", c'est-à-dire ramenées sur le tore

3.1. EXEMPLES "JOUETS"

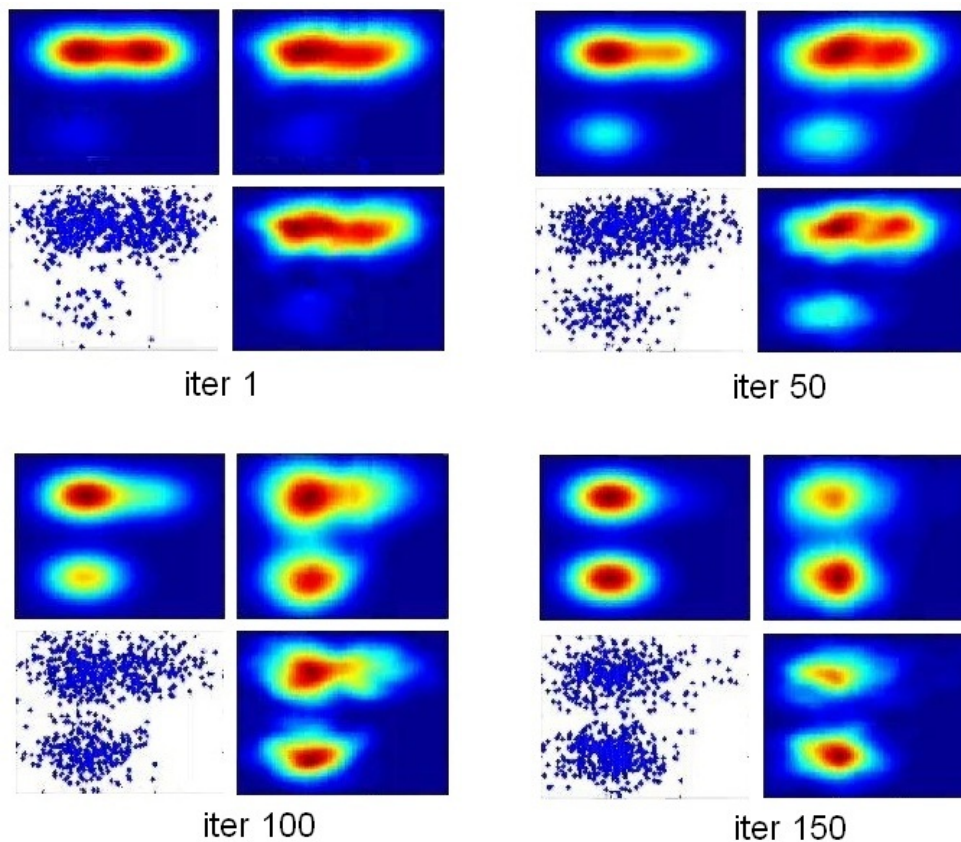


FIGURE 3.6 – Densités théorique/empiriques, cas 1. Le cadre supérieur gauche représente la densité théorique cible, le cadre inférieur gauche représente le nuage de particules, et les deux derniers cadres représentent le nuage de particules sous la forme d’approximations de densités par noyaux gaussiens (noyau constant en bas à droite, adaptatif en haut à droite).

Il est alors remarquable que le nuage a décroché de la densité cible, les noyaux n’étant pas assez mélangeants, les migrations des particules d’un mode à l’autre ne se font pas de façon satisfaisante.

Le second réglage consiste à adopter le schéma d’itération de la figure 3.7(a) (nombre d’itération m_n en fonction de n), et un σ_{prop} plus large, soit $\sigma_{prop} = 0.8$, ce qui donne les produits $b_n \cdot g_n$ de la figure 3.7(b), tous strictement inférieurs à 1. L’algorithme donne alors les résultats représentés sur la figure 3.8, aux étapes 1, 50, 100 et 150.

Notre densité empirique “colle” alors à la densité cible. Pour une évaluation plus quantitative que l’appréciation visuelle, on représente figure 3.9 les proportions des particules du nuage final, d’une part autour du point x_2 , d’autre part autour du point x_3 , comparées aux proportions théoriques de la mesure cible.

Dans cet exemple, l’algorithme SMC est face à un réel problème d’exploration. En effet,

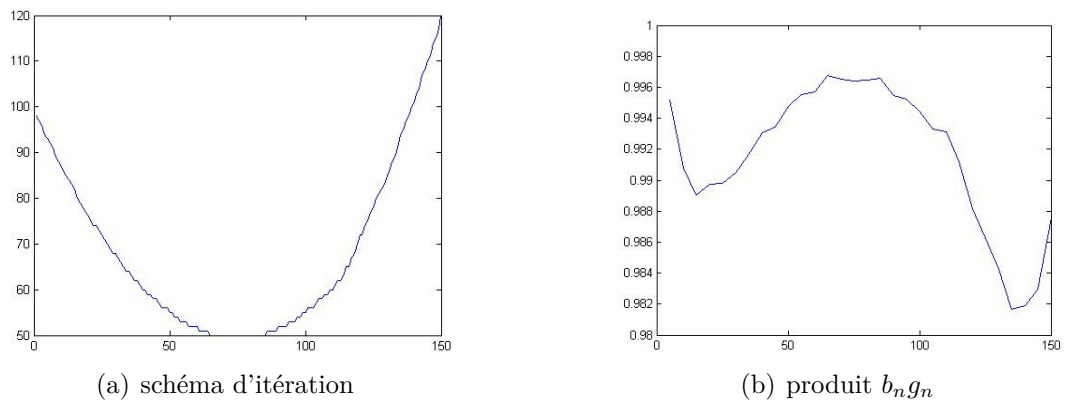


FIGURE 3.7 – cas 2

les densités cibles chargent une zone qui ne contient à l'origine aucune particule, et dont l'accès passe nécessairement par le franchissement d'une barrière de potentiel. Il importe alors que les particules soient capables de se déplacer dans tout l'espace au moyen des étapes de mutation. D'abord pour accéder à cette zone, la repérer, et y "transvaser" des particules par le jeu des sélections/duplications. Ensuite pour compenser les erreurs d'échantillonnage dues à la variance de ces sélections. Dans ce contexte, les critères théoriques de convergence, focalisés justement sur les capacités de mélange des noyaux de Markov des mutations et sur l'instabilité des sélections, trouvent toute leur pertinence.

3.1. EXEMPLES "JOUETS"

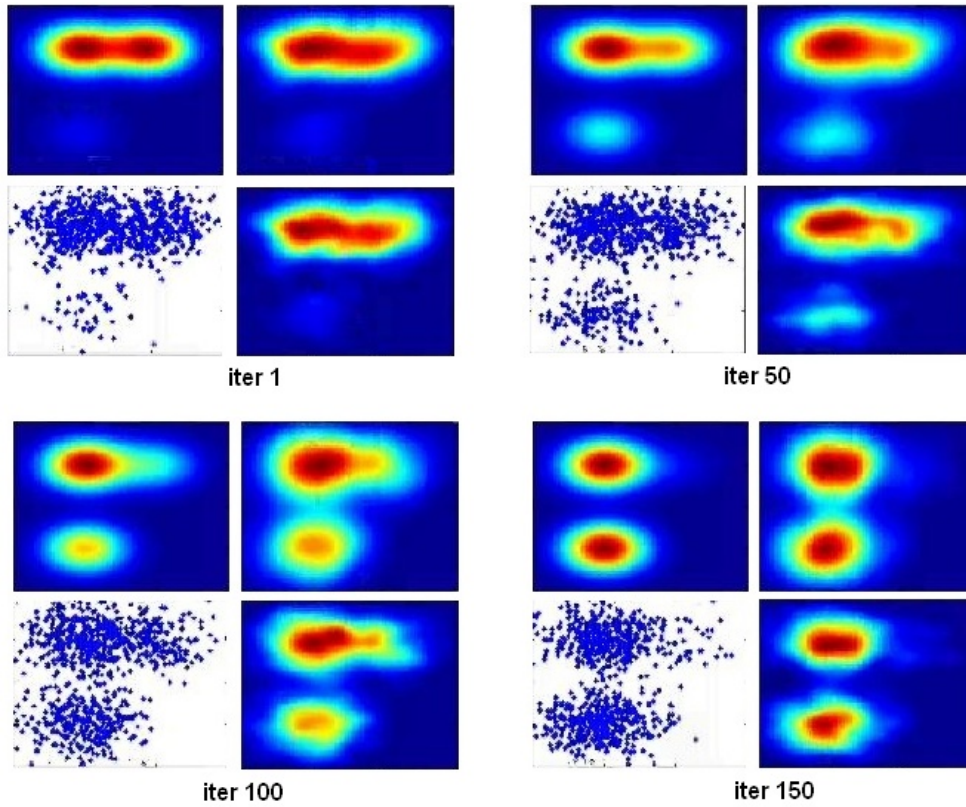


FIGURE 3.8 – Densités théorique/empiriques, cas 2.

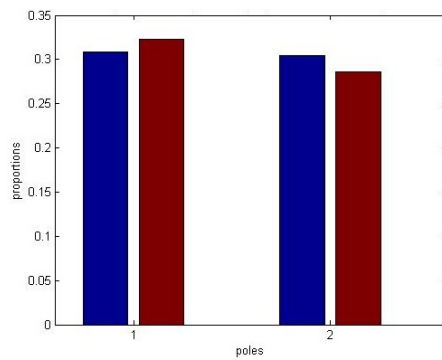


FIGURE 3.9 – proportions empiriques/théoriques des particules autour des 2 principaux pôles

3.2 Problème inverse multicouche 1D

Après ces exemples théoriques, nous nous sommes intéressés à une application pratique des méthodes SMC en électromagnétisme, et plus précisément à leur utilisation pour la résolution d'un problème inverse. Le contexte est le suivant : une onde plane monochromatique est envoyée sur un conducteur parfait recouvert de plusieurs couches de matériaux différents. Chaque matériau Mat i est supposé homogène, et caractérisé par sa permittivité ε_i et sa perméabilité μ_i , deux quantités complexes dépendant de la fréquence de l'onde. Le problème est abordé ici en 1 dimension (voir schéma de la figure 3.10), au sens où les

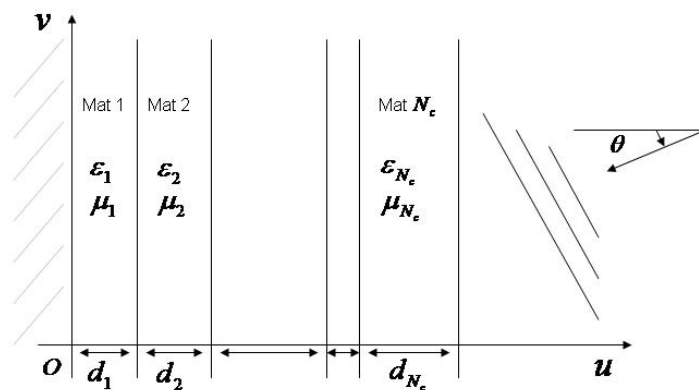


FIGURE 3.10 – Schéma matériau multicouche

couches sont infinies en v et w (direction orthogonale au plan du schéma), et que toutes les grandeurs qui nous intéressent ne dépendent que de u (régime stationnaire en temps). En réponse à l'émission de l'onde plane monostatique de fréquence f et d'angle d'incidence θ , une partie de l'énergie revient à la source de l'onde sous la forme d'une onde de même fréquence ; il s'agit de l'écho radar de l'objet multicouche. Cet écho est (entre autres) caractérisé par le coefficient de réflexion (ou coefficient de rétrodiffusion), quantité complexe définie par $R = \frac{\mathbf{E}_{\text{scat}}}{\mathbf{E}_{\text{inc}}}$, où \mathbf{E}_{inc} désigne le champ incident, et \mathbf{E}_{scat} le champ rétrodiffusé. Ce coefficient R dépend de la fréquence d'émission, de l'angle d'incidence, ainsi que de la polarisation de l'onde HH (pour *émission Horizontale réception Horizontale*) ou VV (pour *émission Verticale réception Verticale*). On pourra se reporter à la section 4.1 page 85 pour une description plus détaillée de cette quantité et de la façon de la mesurer. Dans ce contexte, on suppose qu'une expérience fournit, pour différentes fréquences f et différents angles θ , une famille de coefficients de réflexion $R^{HH}(f, \theta)$, $R^{VV}(f, \theta)$ entachés d'une erreur de mesure. À partir de ces observations et connaissant parfaitement les épaisseurs des différentes couches de matériaux, on cherche à estimer les $\mu_i(f)$, $\varepsilon_i(f)$ de chacun des matériaux Mat i .

Pour formaliser ce problème inverse, on développe tout d'abord dans la section 3.2.1 une approche paramétrique pour définir l'espace d'état du système. On propose ensuite sec-

3.2. PROBLÈME INVERSE MULTICOUCHE 1D

tion 3.2.2 une modélisation probabiliste du problème, au moyen notamment d'un modèle direct partiellement aléatoire. On décrit enfin section 3.2.3 une méthode d'inversion bayésienne par algorithme SMC. La section 3.2.4 présente et analyse nos résultats dans diverses situations.

3.2.1 Paramétrisation du problème, définition de l'espace d'état E

Soit N_c le nombre de couches du matériau multicouche. A chaque matériau i sont associées quatre courbes :

$\varepsilon'_i(f)$: la partie réelle de ε_i en fonction de la fréquence d'émission ;

$\varepsilon''_i(f)$: la partie imaginaire de ε_i ;

$\mu'_i(f)$: la partie réelle de μ_i ;

$\mu''_i(f)$: la partie imaginaire de μ_i .

On cherche donc à estimer $4 \times N_c$ courbes. On paramétrise alors de la façon suivante : on se donne pour chaque matériau i quatre courbes de référence $\varepsilon'_{i,ref}(f)$, $\varepsilon''_{i,ref}(f)$, $\mu'_{i,ref}(f)$, $\mu''_{i,ref}(f)$, et on fait l'hypothèse que les courbes recherchées sont de la forme :

$$\varepsilon'_i(f) = \lambda_i^1 \times \varepsilon'_{i,ref}(f) \quad \varepsilon''_i(f) = \lambda_i^2 \times \varepsilon''_{i,ref}(f)$$

$$\mu'_i(f) = \lambda_i^3 \times \mu'_{i,ref}(f) \quad \mu''_i(f) = \lambda_i^4 \times \mu''_{i,ref}(f)$$

où λ_i^1 , λ_i^2 , λ_i^3 et λ_i^4 sont quatre constantes positives indépendantes de la fréquence. Autrement dit, on connaît les courbes à une constante multiplicative près, ce qui correspond à un modèle d'a priori biaisé en fréquence.

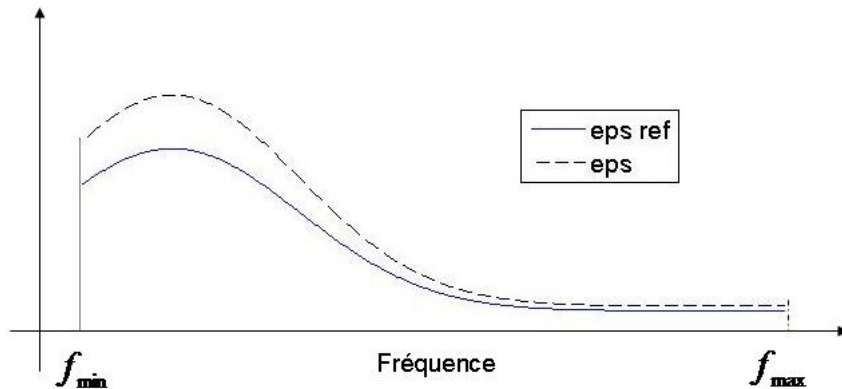


FIGURE 3.11 – Modèle multiplicatif

Par ce modèle multiplicatif simple, chaque courbe, par exemple $\varepsilon'_i(f)$, se déduit de la seule connaissance de λ_i^1 (voir figure 3.11 pour $\lambda_i^1 = 1.2$), d'où un espace d'état E assimilable à

3.2. PROBLÈME INVERSE MULTICOUCHE 1D

$\mathbb{R}_+^{4N_c}$. Autrement dit, chaque élément $x \in E$ est la donnée de $4N_c$ coefficients multiplicatifs $(\lambda_i^{1,2,3,4})_{1 \leq i \leq N_c}$.

3.2.2 Modélisation probabiliste

Pour formuler mathématiquement le problème d'inversion, on définit ici le lien entre l'inconnue (le jeu de coefficients multiplicatifs $(\lambda_i^{1,2,3,4})_{1 \leq i \leq N_c}$, noté x) et l'observation (l'ensemble des coefficients de réflexion mesurés, noté y). Ce lien est fait en considérant un modèle analytique explicite, provenant de la théorie de la réflexion, auquel s'ajoute un modèle de bruit aléatoire additif gaussien pour représenter les incertitudes de mesure. En probabilisant l'inconnue x , on formule alors la solution du problème inverse par une mesure de probabilité sur E donnée par la formule de Bayes.

3.2.2.1 Modèle d'observation

Fixons-nous une fréquence f , un angle θ , un nombre de couches N_c , des épaisseurs de matériaux d_1, \dots, d_{N_c} et des matériaux définis par des coefficients ε_i et μ_i . Le calcul des coefficients de réflexion du matériau multicouche R_{surf}^{HH} et R_{surf}^{VV} (pour chacune des polarisations HH et VV) passe par le calcul de *l'impédance* Z (qui est une fonction de u définie pour chacune des polarisations) à la surface de chacune des couches. Par commodité, on note ε_i , μ_i et $Z(u)$ les permittivités, perméabilités et impédances "normalisées", soit :

$$\begin{cases} \varepsilon_i & \leftarrow \frac{\varepsilon_i}{\varepsilon_0} \\ \mu_i & \leftarrow \frac{\mu_i}{\mu_0} \\ Z & \leftarrow Z \times \sqrt{\varepsilon_0 \mu_0} \end{cases}$$

de sorte que ces trois quantités soient sans unité. L'hypothèse conducteur parfait pour $u < 0$ se traduit en particulier par $Z(0) = 0$.

On admet ici (voir par exemple [66, 19]) qu'aux incertitudes de mesure près, les $Z_i^{HH} := Z(d_i)$ en polarisation HH et $Z_i^{VV} := Z(d_i)$ en polarisation VV sont calculables par le modèle récursif suivant :

$$\begin{cases} Z_0^{HH} := Z^{HH}(0) = 0 \\ Z_0^{VV} := Z^{VV}(0) = 0 \end{cases}$$

et pour chaque $i \geq 1$, connaissant Z_{i-1}^{HH} et Z_{i-1}^{VV} , on a Z_i^{HH} et Z_i^{VV} par :

$$\begin{aligned} \nu_i &= \sqrt{\varepsilon_i \cdot \mu_i} & \eta_i &= \sqrt{\frac{\mu_i}{\varepsilon_i}} & u_i &= \sqrt{1 - \left(\frac{\sin \theta}{\nu_i}\right)^2} \\ A_i^{HH} &= \frac{\eta_i \cdot u_i - Z_{i-1}^{HH}}{\eta_i \cdot u_i + Z_{i-1}^{HH}} & A_i^{VV} &= \frac{Z_{i-1}^{VV} \cdot u_i - \eta_i}{Z_{i-1}^{VV} \cdot u_i + \eta_i} \end{aligned}$$

3.2. PROBLÈME INVERSE MULTICOUCHE 1D

$$\begin{cases} Z_i^{HH} = \eta_i \cdot u_i \cdot \frac{D_i - \frac{A_i^{HH}}{D_i}}{D_i + \frac{A_i^{HH}}{D_i}} \\ Z_i^{VV} = \frac{\eta_i}{u_i} \cdot \frac{D_i + \frac{A_i^{VV}}{D_i}}{D_i - \frac{A_i^{VV}}{D_i}} \end{cases}$$

où $D_i = e^{j\frac{2\pi f}{C}\nu_i u_i d_i}$.

Enfin, les coefficients de réflexion sont définis par :

$$R_{surf}^{HH} = \frac{Z_{N_c}^{HH} - \cos \theta}{Z_{N_c}^{HH} + \cos \theta} \quad R_{surf}^{VV} = \frac{Z_{N_c}^{VV} \cos \theta - 1}{Z_{N_c}^T \cos \theta - 1}$$

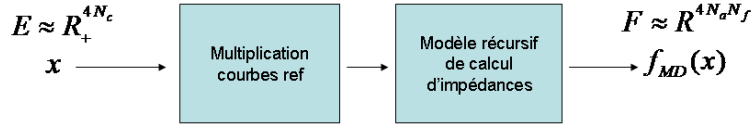


FIGURE 3.12 – Modèle direct "parfait"

Ainsi, en se fixant un nombre de matériaux N_c , des épaisseurs d_1, \dots, d_{N_c} , des fréquences f_1, \dots, f_{N_f} , des angles $\theta_1, \dots, \theta_{N_a}$ et des courbes de référence $\varepsilon'_{i,ref}(f)$, $\varepsilon''_{i,ref}(f)$, $\mu'_{i,ref}(f)$, $\mu''_{i,ref}(f)$, on obtient le modèle direct "parfait" (i.e. sans bruit de mesure) schématisé sur la figure 3.12, où $x = (\lambda_i^{1,2,3,4})_{1 \leq i \leq N_c} \in E$ désigne l'état du système et $f_{MD}(x)$ constitue le vecteur des coefficients de réflexion obtenus pour chaque couple fréquence-angle, et pour chaque polarisation. En séparant les parties réelle et imaginaire des $N_f \times N_a$ coefficients de réflexion pour chacune des polarisation HH et VV, on obtient que $f_{MD}(x)$ est de dimension $4N_f \cdot N_a$. L'ensemble F , assimilable à $\mathbb{R}^{4N_f N_a}$, est appelé espace image.

Ce modèle déterministe étant défini, on formalise le modèle direct d'observation complet (c'est-à-dire avec bruit de mesure) selon l'équation suivante :

$$y = f_{MD}(x) + b, \quad (3.1)$$

où b est un bruit aléatoire gaussien de dimension (réelle) $4N_f \cdot N_a$, centré et de matrice de covariance $\sigma_{bruit}^2 \cdot Id$ (Id désigne ici la matrice identité en dimension $4N_f \cdot N_a$). Cela correspond à une hypothèse de termes de bruit de mesure i.i.d pour chacune des composantes de y , cela conditionnellement à x .

3.2.2.2 Approche et inversion bayésiennes

Afin de prendre en considération les informations que l'on a sur x et ses composantes (positivité, ordre de grandeur), on choisit de probabiliser cette variable, plus précisément

3.2. PROBLÈME INVERSE MULTICOUCHE 1D

de lui associer une densité a priori $p(x)$. Dans un premier temps, on associe à chacune des composantes x_i du vecteur x une densité marginale a priori $p(x_i)$ sous la forme d'une gaussienne de variance σ_{ap}^2 , tronquée sur les valeurs positives, soit

$$p(x_i) \propto \exp\left(\frac{-x_i^2}{2\sigma_{ap}^2}\right) \mathbf{1}_{x_i \geq 0}.$$

Dans un second temps, en partant du principe que les x_i sont a priori indépendants entre eux, on obtient la densité jointe $p(x)$ suivante :

$$p(x) \propto \prod_{i=1}^{4N_c} p(x_i) \propto \exp\left(\frac{-\|x\|_2^2}{2\sigma_{ap}^2}\right) \mathbf{1}_{x \geq 0},$$

$x \geq 0$ signifiant que toutes les composantes x_i de x sont positives.

Toujours dans une logique bayésienne, la densité de vraisemblance $p(y|x)$ se déduit directement de la formulation (3.1) du modèle direct :

$$p(y|x) \propto \exp\left(\frac{-\|f_{MD}(x) - y\|_2^2}{2\sigma_{bruit}^2}\right).$$

Sous ce formalisme, et d'après la formule de Bayes, la densité a posteriori de x sachant y vaut :

$$p(x|y) \propto \underbrace{\exp\left(\frac{-\|f_{MD}(x) - y\|_2^2}{2\sigma_{bruit}^2} - \frac{\|x\|_2^2}{2\sigma_{ap}^2}\right)}_{=:h(x)} \mathbf{1}_{x \geq 0} \quad (3.2)$$

On vient d'expliciter la mesure de probabilité sur E solution de notre problème inverse, mais cette solution purement formelle n'est en l'état d'aucune utilité. La difficulté est de savoir l'exploiter. De ce point de vue, il paraît clair qu'aucune solution analytique n'est à espérer. Pour tout $x \in E$, on sait calculer la quantité $h(x)$, ce qui ouvre la possibilité d'une évaluation systématique de cette fonction sur un maillage couvrant l'ensemble E . Cependant, cette solution est irréalisable compte tenu de la dimension de l'espace et de l'irrégularité de la fonction f_{MD} (comme on le constate a posteriori), qui nécessite un maillage trop fin et donc des temps de calcul trop longs. Dans cette situation, la solution que l'on adopte est donc de *simuler* cette loi a posteriori. Son échantillonnage permet de répondre à nos objectifs concrets : déterminer les zones de E où se concentre sa masse, autrement dit expliciter numériquement les x qui sont cohérents avec les observations et nos a priori.

3.2.3 Echantillonnage par algorithme SMC

De fait, la fonction f_{MD} est très irrégulière. Il en résulte que la loi de probabilité $p(x|y)dx$ sur E est fortement multimodale, en plus d'être définie sur un espace multidimensionnel,

3.2. PROBLÈME INVERSE MULTICOUCHE 1D

ce qui la rend impossible à simuler par des méthodes directes comme la méthode d'acceptation-rejet ou l'échantillonnage d'importance. Au contraire, les méthodes SMC sont particulièrement appropriées pour ce genre de situation. L'utilisation d'un grand nombre de particules permet au nuage de couvrir plusieurs modes à la fois, et la formulation séquentielle dans l'espace des mesures de probabilité sur E permet une introduction progressive de la difficulté d'échantillonnage. Ce dernier point, particulièrement important, requiert l'élaboration d'une stratégie d'interpolation.

3.2.3.1 Interpolation dans l'espace des mesures

Dans la perspective d'approcher la mesure de probabilité $p(x|y)dx$ par un système dynamique de particules, on pose la famille de mesures suivantes :

$$\eta_\alpha(dx) \propto \exp\left(-\alpha \frac{\|f_{MD}(x) - y\|_2^2}{2\sigma_{bruit}^2} - \frac{\|x\|_2^2}{2\sigma_{ap}^2}\right) \mathbf{1}_{x \geq 0} \cdot dx,$$

où α est un paramètre vivant dans l'ensemble $[0, 1]$ de sorte que :

- η_1 est la mesure de probabilité recherchée ;
- η_0 est une simple gaussienne tronquée évidente à simuler ;
- pour toute suite croissante α_n , $n \in \{0, \dots, n_{max}\}$ où $\alpha_0 = 0$ et $\alpha_{n_{max}} = 1$, la suite de mesures $(\eta_{\alpha_n})_n$ suit le flot de Feynman-Kac associé à la mesure initiale η_0 , aux potentiels

$$G_n(x) = \exp\left((\alpha_n - \alpha_{n-1}) \frac{\|f_{MD}(x) - y\|_2^2}{2\sigma_{bruit}^2}\right)$$

et aux noyaux de Markov laissant invariants les η_{α_n} .

Si l'on fixe une telle suite α_n , on est alors en situation d'appliquer un algorithme SMC (voir la section 1.3 page 10) initialisé en η_0 , en utilisant les G_n comme fonctions de sélection et les M_n comme noyaux de mutation, de façon à approcher la suite de mesures η_{α_n} . Si le processus particulaire est suffisamment stable et robuste, la dernière génération de N particules $\zeta_{n_{max}} = (\zeta_{n_{max}}^1, \dots, \zeta_{n_{max}}^N)$ fournit une bonne approximation $\eta_{\alpha_{n_{max}}}^N$

$$\eta_{\alpha_{n_{max}}}^N = \frac{1}{N} \sum_{i=1}^N \delta_{\zeta_{n_{max}}^i}$$

de la mesure de probabilité $\eta_{\alpha_{n_{max}}} = \eta_1$ qui est exactement la mesure de probabilité associée à la densité $p(x|y)$. En d'autres termes, la localisation de ces particules dans l'espace E est la solution du problème inverse. Malgré cette formulation très générale du SMC, de nombreux paramètres restent à régler dans l'application pratique de l'algorithme.

3.2.3.2 Etapes de sélection

Les fonctions de sélection G_n sont dépendantes du pas $\Delta\alpha := \alpha_n - \alpha_{n-1}$ à chaque itération. Plusieurs suites (α_n) ont été testées. Pour des raisons de stabilité et d'efficacité algorithmiques (voir l'introduction de la section 2.4 page 42), nous avons retenu une méthode dynamique d'incrémentement du paramètre α , quasi identique au modèle décrit section 2.4.2

3.2. PROBLÈME INVERSE MULTICOUCHE 1D

page 44. Le raisonnement est le suivant : on part du principe qu'une sélection raisonnable est une sélection qui opère sur environ 30% des particules. Si $\omega_1, \dots, \omega_N$ sont les poids des N particules, la probabilité que la particule i soit redistribuée est $\frac{\omega_i}{\max(\omega_j)}$, ce qui signifie que l'espérance du nombre de particules qui vont bouger est :

$$\sum_i \frac{\omega_i}{\max(\omega_j)} = \frac{\|\omega\|_1}{\|\omega\|_\infty}$$

où $\omega = (\omega_1, \dots, \omega_N)$. C'est cette quantité que l'on veut proche de 30% de N . Or, d'après la définition des G_n , les ω_i sont de la forme $v_i^{\Delta\alpha}$, donc

$$\frac{\|\omega\|_1}{\|\omega\|_\infty} = \sum_i \left(\frac{v_i}{\max(v_j)} \right)^{\Delta\alpha},$$

ce qui met en évidence une fonction décroissante de $\Delta\alpha$, valant N pour $\Delta\alpha = 0$ et 1 pour $\Delta\alpha \rightarrow +\infty$. Comme cette fonction est aisément calculable, on détermine le $\Delta\alpha$ de telle façon qu'elle vaille environ $30\%.N$, ce par une dichotomie.

3.2.3.3 Etapes de mutation

Le choix du noyau de Markov employé pour la mutation représente en pratique une large liberté pour l'utilisateur. Ce noyau de Markov a pour but à la fois d'assurer un mélange suffisant pour explorer convenablement tout l'espace d'état, et de faire bouger un nombre suffisant de particules sur les N présentes. Pour le type de noyau de Markov, le choix que nous avons fait dans notre problème est d'utiliser les noyaux de Metropolis-Hastings $MH_n(\sigma_{prop})$ associés aux fonctions $h_n = e^{-V_n}$ où :

$$V_n(x) = \alpha_n \frac{\|f_{MD}(x) - y\|_2^2}{2\sigma_{bruit}^2} + \frac{\|x\|_2^2}{2\sigma_{ap}^2} + Inf_-(x)$$

avec $Inf_-(x) := 0$ si $x \geq 0$ et $+\infty$ sinon (voir la définition (1.3) page 3). Le noyau de proposition $K(x, \cdot)$ utilisé dans $MH_n(\sigma_{prop})$ est le noyau gaussien centré en x de matrice de covariance $\sigma_{prop}^2 Id$, Id désignant ici la matrice identité en dimension $4N_c$.

Si l'on utilise à chaque étape un seul de ces noyaux, les deux objectifs cités au-dessus s'avèrent antagonistes. Un σ_{prop} élevé donnera de bonnes propriétés de mélange, mais très vite, les zones d'intérêt (zones où se concentre la mesure de probabilité cible) sont plus restreintes, de sorte que la grande majorité des particules restent sur place, la proposition de MH étant trop éloignée des zones d'intérêt pour être acceptée. A l'inverse, un σ_{prop} petit assure des propositions proches des particules d'origine, donc plus probablement intéressantes et donc acceptées. En revanche, de tels noyaux ne permettent pas d'explorer tout l'espace au sens où pour passer une barrière de potentiel, par exemple pour sortir d'un minimum local, une particule devra effectuer de nombreux petits sauts défavorables.

Le choix réalisé est la composition de plusieurs $MH_n(\sigma_{prop})$, pour des σ_{prop} décroissants. Encore une fois, cette suite de σ_{prop} est choisie de façon dynamique (voir [22, 67] pour

3.2. PROBLÈME INVERSE MULTICOUCHE 1D

des méthodes similaires), partant d'un $\sigma_{initial}$ large permettant d'explorer l'espace et diminuant jusqu'à l'obtention d'un nombre total de particules bougées de 90% environ. Plus précisément, à chaque itération, on commence par utiliser un $\sigma_{prop} := \sigma_{initial}$ fixe, de l'ordre du σ utilisé pour l'a priori. On compte le nombre de particules qui ont bougé. Si ce nombre ne dépasse pas 90%, on décrémente σ en le multipliant par un $\lambda \in]0, 1[$ fixe (typiquement, $\lambda = 0.9$). On recompte le nombre de particules qui ont bougé (depuis le début), pour le comparer à $90\%.N$, et ainsi de suite, jusqu'à ce que 90% des particules aient bougé. A chaque itération, même lorsque le nuage de particules est très resserré, il est effectué des propositions avec σ_{prop} large. A noter que pour limiter le nombre trop grand de propositions inutiles, nous avons également opté pour une décrémentation de σ_{prop} plus dynamique, où le λ s'adapte pour passer de $\sigma_{initial}$ au dernier σ_{prop} de l'itération précédente par exemple en maximum 30 décrémentations.

3.2.4 Résultats

Commençons par quelques précisions numériques. Le modèle direct probabiliste (3.1) et la formule de Bayes (3.2) font intervenir plusieurs quantités caractéristiques qui traduisent notre degré de connaissance a priori et nos hypothèses sur le bruit de mesure. Du point de vue du processus d'inversion par SMC et de sa mise en oeuvre informatique, elles interviennent naturellement dans les étapes d'initialisation, de sélection et de mutation.

- A priori : les x_i , composantes du vecteur x , sont d'après notre paramétrage des coefficients multiplicatifs que l'on applique aux courbes de référence, et sont donc typiquement de l'ordre de l'unité. On pose donc notre σ_{ap} autour de 5 ou 10 pour cadrer une zone raisonnable tout en restant non informatif dans la zone d'intérêt.
- Vraisemblance : la variance de notre modèle de vraisemblance est liée aux hypothèses faites sur les incertitudes de mesure. On se place dans le contexte (simulé) d'un bruit de mesure faible, de l'ordre de $\sigma_{bruit} = 10^{-4}$ ou 10^{-5} selon les cas. Paradoxalement, ce choix complique l'inversion, car il implique une attache aux données très rigoureuse, ce qui se traduit par une loi a posteriori extrêmement localisée spatialement, et donc un processus d'exploration délicat à mener.

De façon générale, la qualité des résultats obtenus dépend essentiellement du nombre de couches N_c , qui définit la dimension de l'espace d'état $4N_c$. Beaucoup plus que de N , le nombre de particules, pris généralement entre 500 et 3000 pour avoir à la fin un nuage assez dense pour pouvoir représenter une densité, et des calculs à un processeur ne dépassant pas l'ordre de grandeur de l'heure. Plusieurs tests ont été faits avec $N_c = 1$ ou 2, et tous donnent des résultats très satisfaisants. Nous en présentons deux exemples, l'un sur un cas de matériau relativement quelconque (section 3.2.4.1), l'autre sur un cas de matériau optimisé pour être très absorbant (section 3.2.4.2). Nous nous penchons ensuite sur le cas $N_c \geq 3$, plus problématique (section 3.2.4.3).

3.2. PROBLÈME INVERSE MULTICOUCHE 1D

3.2.4.1 Exemple d'un matériau quelconque

Nous prenons pour premier exemple un matériau à deux couches. Les deux matériaux considérés, donnés dans [73], sont notés **Mat 1** et **Mat 2**. Leurs courbes de référence sont représentées sur la figure 3.13.

On prend dans cet exemple un nombre de fréquences $N_f = 4$, fréquences prises entre 0.1 et 18 GHz. Trois angles sont utilisés : $\{-40, 0, 40\}$ (en degrés), 0 degré correspondant à une onde arrivant sur le matériau multicouche de face. Les épaisseurs sont les suivantes (en mm) :

- **Mat 1** : 2.12
- **Mat 2** : 0.96

L'état réel x_{cible} , c'est-à-dire les $4N_c = 8$ coefficients multiplicatifs utilisés pour simuler les mesures est défini numériquement par

$$\begin{aligned} \lambda_1^1 &= 0.6 & \lambda_1^2 &= 0.2 & \lambda_1^3 &= 2.1 & \lambda_1^4 &= 1.2 \\ \lambda_2^1 &= 1.5 & \lambda_2^2 &= 0.3 & \lambda_2^3 &= 0.9 & \lambda_2^4 &= 2.5 \end{aligned}$$

soit $x_{cible} = [0.6 \ 0.2 \ 2.1 \ 1.2 \ 1.5 \ 0.3 \ 0.9 \ 2.5]$. On choisit $\sigma_{ap} = 5$, $\sigma_{bruit} = 10^{-4}$. Les résultats obtenus avec un système de $N = 1000$ particules sont présentés figure 3.14.

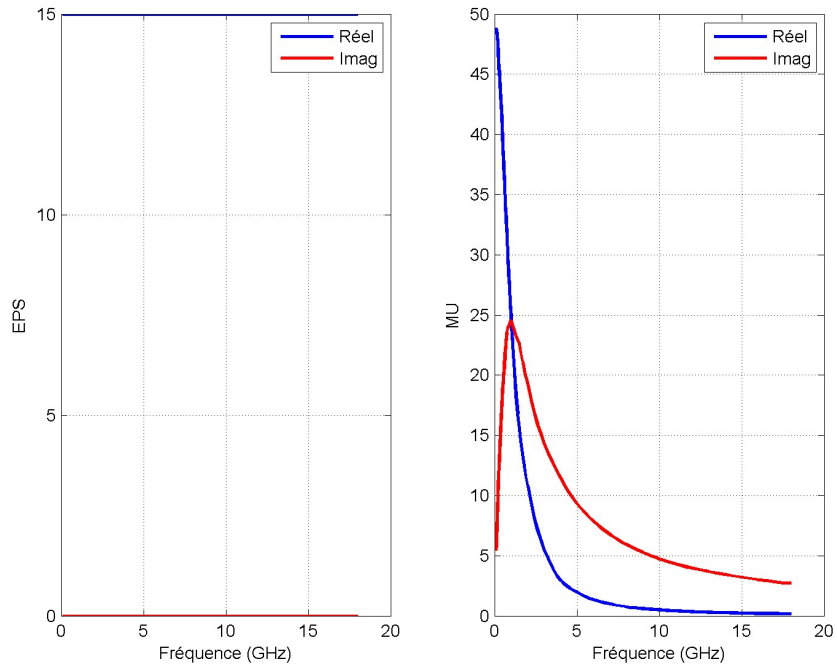
Les 2 histogrammes de la première ligne représentent la répartition des particules, dans 2 des 8 dimensions du problème, le point rouge indiquant dans chacune la position de la particule cible. Il est à noter que ce ne sont pas les composantes elles-mêmes des particules qui sont représentées, mais ces composantes multipliées par les valeurs des ε et μ de référence correspondants à la fréquence centrale du domaine utilisé (préférées pour leur sens physique). Numériquement, on multiplie donc les x et x_{cible} par

$$[15.0 \ 0 \ 4.94 \ 13.72 \ 9.79 \ 12.79 \ 1 \ 0]$$

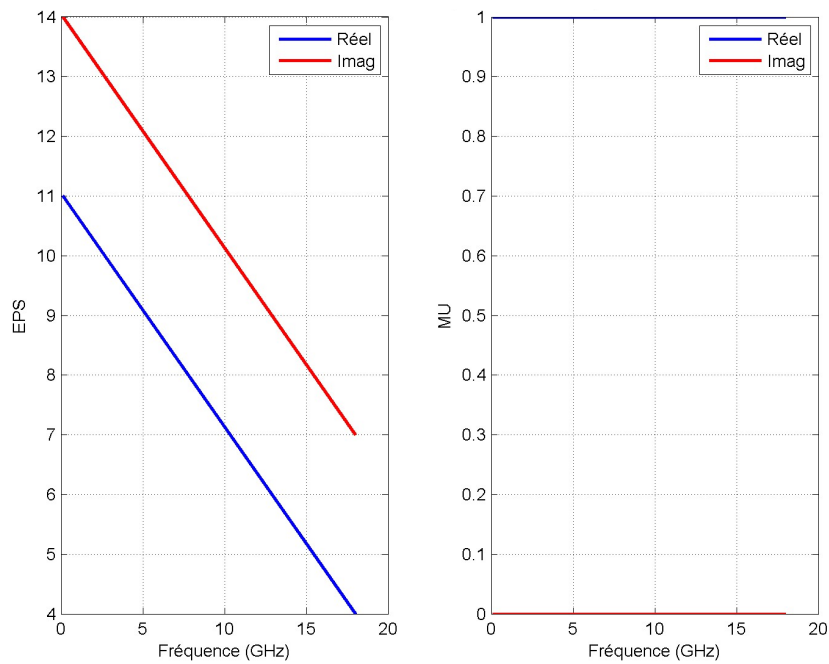
terme à terme. Indépendamment de cette multiplication, les densités obtenues autour de ces valeurs cibles ont des variances sensiblement différentes, reflétant la sensibilité différente du modèle direct selon ces dimensions. En d'autres termes, certaines composantes sont plus observables que d'autres. La figure en bas à droite représente une projection des particules (nuage de particules fourni par l'algorithme en bleu, particule cible en rouge) sur un espace 2D choisi aléatoirement. C'est une représentation tout-à-fait non rigoureuse du problème, mais plus intuitive.

L'espace image, c'est-à-dire l'espace des coefficients de réflexion, est de dimension (dans \mathbb{R}) $4N_a \times N_f$, car pour chaque polarisation HH ou VV, et chaque couple angle-fréquence, on dispose d'un coefficient de réflexion qui a une partie réelle et une partie imaginaire. Sont représentés en figure 3.15 les histogrammes des f_{MD} des particules, obtenus pour 4 des 48 dimensions.

3.2. PROBLÈME INVERSE MULTICOUCHE 1D



(a) Mat 1



(b) Mat 2

FIGURE 3.13 – Permittivité/Perméabilité matériaux

3.2. PROBLÈME INVERSE MULTICOUCHE 1D

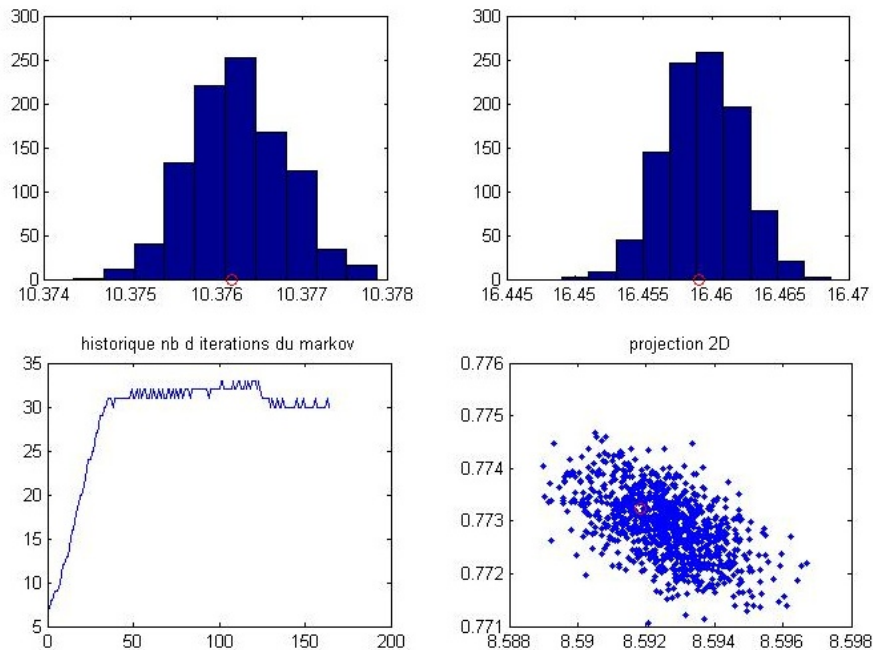


FIGURE 3.14 – Espace d'état, cas $N_c = 2$. En haut, histogrammes des particules selon les composantes λ_1^3 et λ_1^4 . En bas, nombre de mutations par génération de particules et projection du nuage de particules sur un plan aléatoire de l'espace E .

Cette fois, l'ordre de grandeur des variances des densités obtenues ne dépend sensiblement plus de la dimension observée. C'est un résultat que l'on peut expliquer. Théoriquement, une gaussienne centrée sur le vecteur des coefficients mesurés, et de matrice de covariance $\sigma_{bruit}^2 \cdot Id$ serait la densité de $f_{MD}(x)$ sachant y dans le cas d'un a priori non informatif (c'est-à-dire uniforme) sur $f_{MD}(x)$. Cette dernière hypothèse n'est pas vérifiée, mais trouve tout de même une certaine pertinence sous une forme plus faible : premièrement, nous avons posé un a priori non informatif (c'est-à-dire proche de la loi uniforme sur les zones d'intérêt) sur x , et deuxièmement, la mesure cible (et donc les particules) se concentrant sur un voisinage proche de la particule cible, f_{MD} peut être approchée par son développement au premier ordre, une application affine, c'est-à-dire une application qui envoie des mesures uniformes sur des mesures uniformes. A cette approximation près sur f_{MD} , la loi a priori de $f_{MD}(x)$ est donc une loi uniforme sur le sous-espace engendré par l'application affine en question (le sous-espace tangent de f_{MD}). La dimension de l'espace d'état étant strictement inférieure à celle de l'espace image, ce sous-espace est un sous-espace strict, de dimension au maximum $4N_c$. En revanche, le terme de bruit, lui, charge de la même façon chacune des $4N_a N_f$ dimensions. A l'approximation affine près, la loi de $f_{MD}(x)|y$ n'est pas la gaussienne évoquée plus haut, mais une projection de celle-ci sur l'espace tangent à f . On peut relever deux conséquences notoires, que l'on comprend visuellement sur le schéma de la figure 3.16.

3.2. PROBLÈME INVERSE MULTICOUCHE 1D

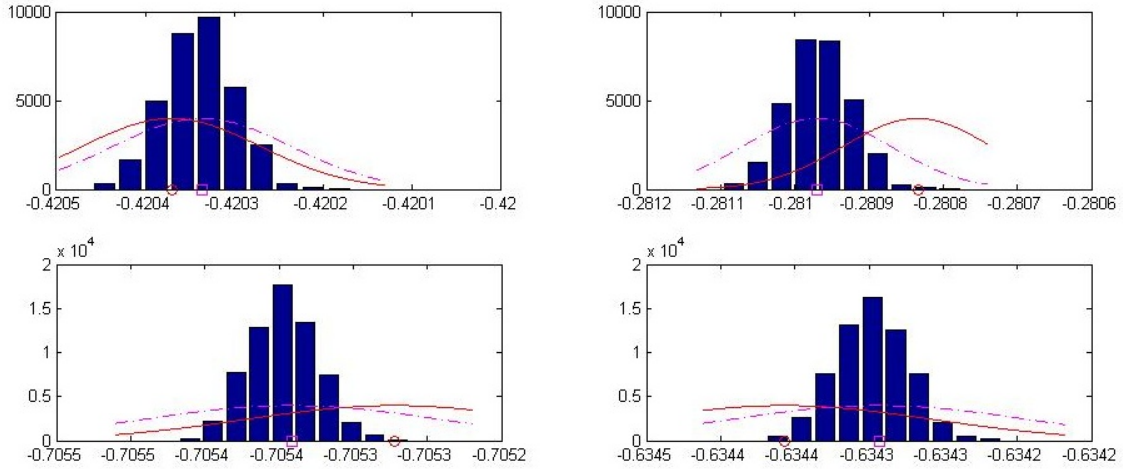


FIGURE 3.15 – 4 histogrammes des particules dans l’espace image, cas $N_c = 2$. Les points rouges représentent ici les y_i , coefficients de réflexion mesurés (ici simulés), et les carrés magenta représentent les “coefficients parfaits”, c’est-à-dire les coefficients sans terme de bruit (inconnus dans un contexte non simulé). Sont représentées les gaussiennes de variance σ_{bruit}^2 , centrées sur les coefficients mesurés (en rouge), et sur les coefficients parfaits (pointillés magenta).

Premièrement, la variance de cette loi peut être plus faible que σ_{bruit}^2 selon les dimensions. Sur le schéma, on voit par exemple que la variance selon le vecteur unitaire \vec{j} sera beaucoup plus faible, ce parce-qu’il est presque orthogonal à l’espace tangent, alors que la variance selon \vec{i} sera proche de σ_{bruit}^2 . Ce phénomène se lit dans nos résultats où la différence de dimension est très forte (8 et 48), la plupart des distributions étant beaucoup plus resserrées que les gaussiennes de variance σ_{bruit}^2 représentées à côté.

Deuxièmement, le centre de cette loi n’est plus y , mais son projeté, qui est beaucoup plus proche des coefficients parfaits $f_{MD}(x_{cible})$ que des coefficients mesurés y . C’est le cas sur le schéma dans la direction \vec{j} . En des termes plus approximatifs, on peut remonter aux termes de bruit dans certaines dimensions, grosso modo celles orthogonales à l’espace tangent. On peut ainsi observer sur nos résultats des distributions plus proches du carré magenta que du point rouge, le carré n’étant pourtant pas observable par l’algorithme.

3.2.4.2 Cas d’un matériau "optimisé" (selon [73])

Nous avons jugé opportun de réaliser un test sur le cas d’un empilement de couches optimisé pour l’obtention de coefficients de réflexion très faibles. Avec des matériaux donnés dans [73] et assemblés avec les épaisseurs 3.187mm et 0.668mm, les coefficients de réflexion obtenus pour $\theta = 0$ et $f \in [0.2, 2]$ (en GHz) sont inférieurs à 10^{-1} pour toutes les fréquences entre 0.2 et 2 GHz (voir [73]). Bien qu’il se formalise de la même manière que le précédent, ce cas présente une différence qualitative : au lieu de rechercher une zone de l’espace E qui s’apparente grosso modo à une ligne de niveau de la fonction f_{MD} , l’algorithme doit

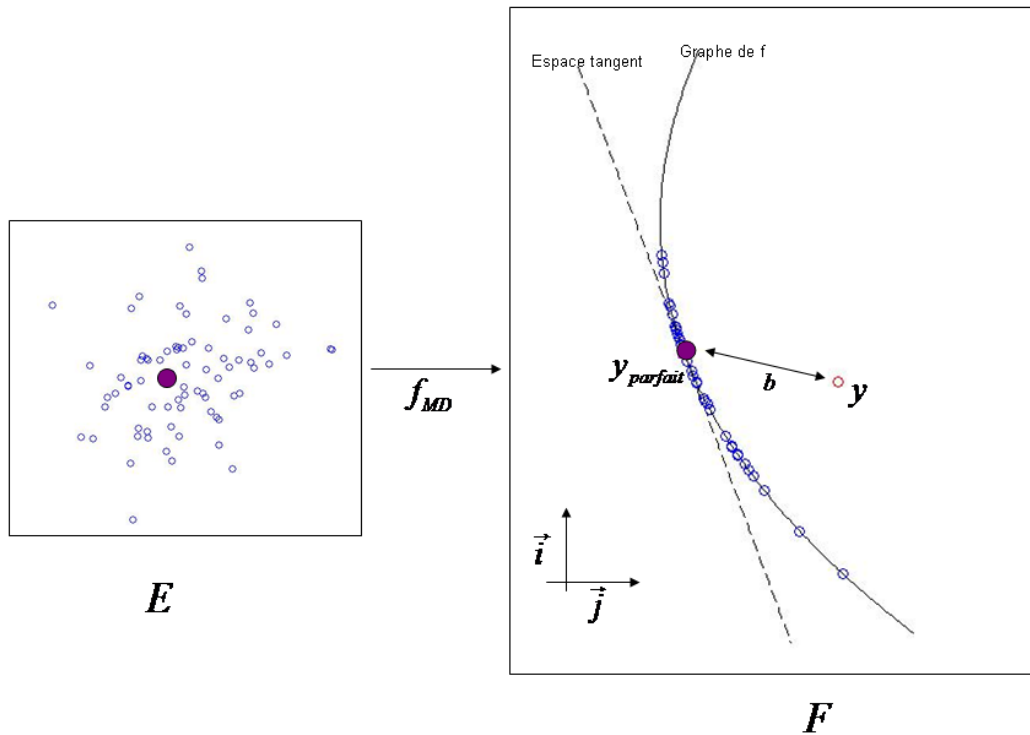


FIGURE 3.16 – Sensibilité dimensionnelle

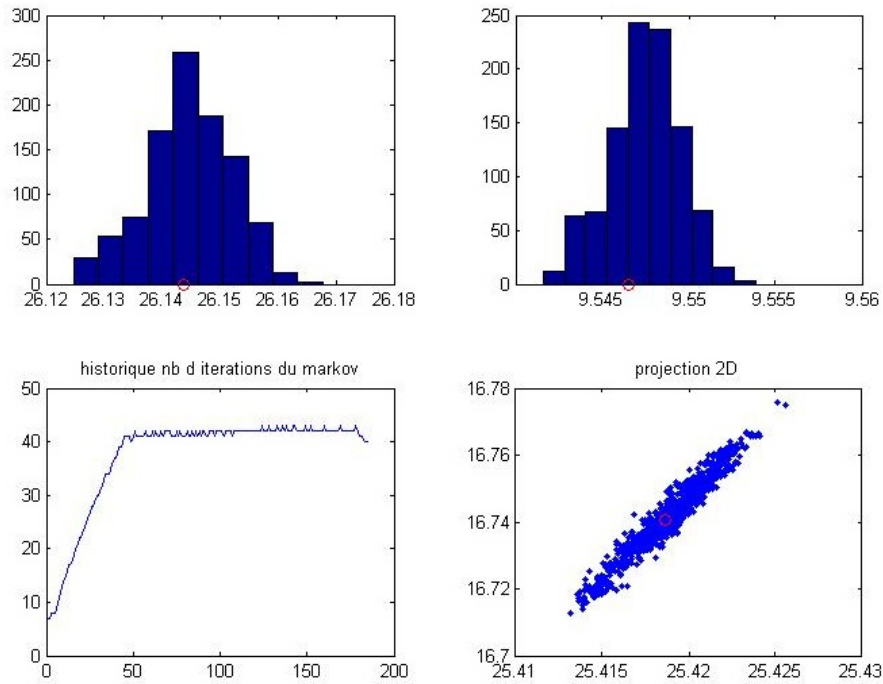
trouver quelque chose d'intuitivement proche d'un *minimiseur* de cette fonction, ce qui est potentiellement très différent d'un point de vue géométrique. Les résultats tendent néanmoins à montrer que l'algorithme se comporte de façon similaire. Nous les représentons figure 3.17 dans l'espace d'état et l'espace image, en utilisant la même légende que pour le cas précédent.

3.2.4.3 Analyse d'un cas $N_c \geq 3$

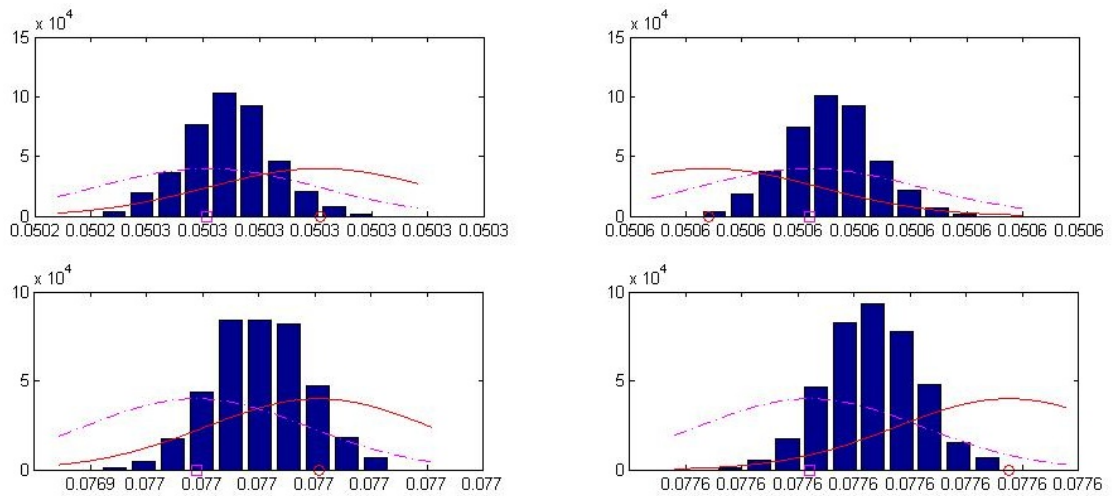
Les tests faits pour $N_c \geq 3$ ne donnent pas de résultats satisfaisants, soit d'un point de vue performance, soit d'un point de vue temps de calcul. La raison essentielle est la géométrie de la densité cible. Deux types de problème ont été identifiés :

- Premier type : le terme de vraisemblance dans l'expression de la densité cible charge un ensemble multidimensionnel large ou non borné. Dans ce cas, les bornes du support de la mesure cible ne sont dues qu'à l'a priori, et l'ensemble à couvrir est trop grand pour être représenté par un nuage de particules. Les particules ont tendance à effectuer des sortes de marche aléatoire au sein de cet ensemble, sans pouvoir se distinguer les unes des autres. Il en résulte que la mesure empirique définie par le nuage de particules couvre un sous-domaine du domaine d'intérêt. D'un point de vue problème inverse, l'algorithme fournit *des* solutions possibles, mais pas toutes, et sans quantification probabiliste pertinente.
- Deuxième type, qualitativement différent mais en pratique proche du premier : le terme

3.2. PROBLÈME INVERSE MULTICOUCHE 1D



(a) espace d'état



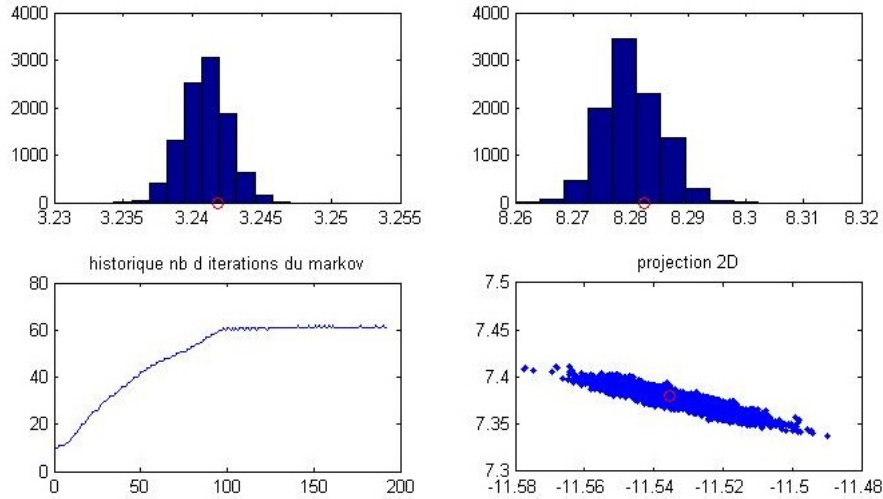
(b) espace image

FIGURE 3.17 – Cas matériau optimisé

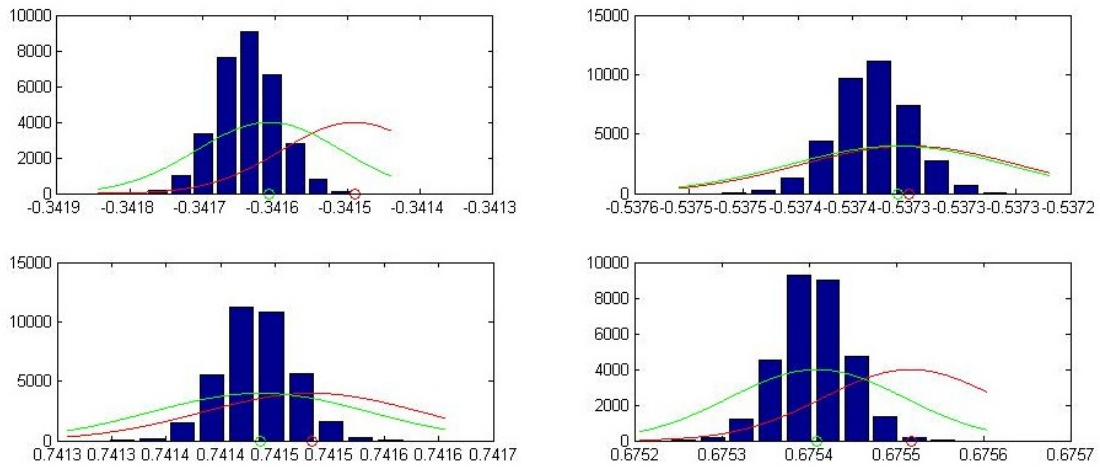
de vraisemblance $\|f_{MD}(x) - y\|_2^2$ admet des minima locaux, possiblement nombreux et multidimensionnels, donnant des résultats très proches les uns des autres. Dans ce cas, les particules sont facilement piégées dans des minima locaux, la conséquence pratique étant un nombre de particules N très grand et de nombreux sauts markoviens à effectuer pour

3.2. PROBLÈME INVERSE MULTICOUCHE 1D

pallier cette tendance. Les résultats de la figure 3.18, obtenus pour $N_c = 3$, satisfaisants d'un point de vue performance, ont été réalisés avec $N = 10000$, pour un temps de calcul total de l'ordre de la journée.



(a) espace d'état



(b) espace image

FIGURE 3.18 – Cas $N_c = 3$

Conclusion du chapitre

A l'évidence, les hypothèses des théorèmes du chapitre 2 expriment des conditions suffisantes, mais pas nécessaires. Au demeurant, la section 3.1.2 illustre une importante vérité : l'absence d'une capacité à explorer l'espace d'état E , en plus du risque de laisser des zones

3.2. PROBLÈME INVERSE MULTICOUCHE 1D

importantes totalement désertes de particules, empêche l'algorithme de corriger ses erreurs d'échantillonnage. Il fournit alors un résultat qualitativement satisfaisant (les zones d'intérêt sont occupées par des particules), mais quantitativement erroné, au sens où le *nombre* de particules n'est pas représentatif de la masse de probabilité réelle qu'il est censé représenter. Même s'il paraît cohérent, le résultat d'un algorithme SMC n'est fiable que s'il découle d'un processus stable.

En pratique, cette stabilité est difficile voire impossible à exprimer en fonction des quantités théoriques propres au flot de Feynman-Kac, mais elle peut se contrôler différemment. Tout d'abord, par un contrôle sur la vitesse d'évolution du processus, notamment par une approche adaptative de la sélection. Avancer trop vite dans l'espace des mesures mène à des erreurs d'échantillonnage brutales, impossibles à compenser. Imposer une sélection raisonnable semble donc être un moyen simple pour à la fois éviter un décrochage, et garantir un rythme minimum d'évolution de l'algorithme. Ensuite, par un contrôle du "mélange". Les mutations sont très délicates à définir car dans la mesure où elles représentent l'unique moyen de déplacement des particules, elles doivent assurer deux missions :

- l'exploration de E , qui nécessite des mouvements larges tout au long de l'algorithme ;
- la convergence du nuage vers les zones d'intérêt, qui nécessite des mouvements sur une distance de même ordre de grandeur que ces zones.

Face à ces objectifs contradictoires, nous avons fait le choix dans l'application présentée section 3.2 (comme d'ailleurs dans le chapitre 4) d'une stratégie prudente. A chaque génération, les mutations sont opérées par composition de noyaux de Metropolis-Hastings aux noyaux de proposition de taille très variée, et sujettes à des impératifs chiffrés de bon mélange. Du point de vue de l'implémentation pratique, cela implique de très nombreuses propositions rejetées, et donc un coût important en temps de calcul. Mais l'avantage est la fiabilité du résultat. A ce titre, l'exemple de la section 3.2 est très parlant. Il montre comment une modélisation probabiliste, couplée à une méthode d'inversion SMC réglée de façon intelligente, permet une résolution fiable et précise d'un problème inverse mal posé et très irrégulier. Sans parallélisation, le temps de calcul reste la principale barrière, notamment en dimension élevée où un grand nombre de particules est nécessaire pour couvrir l'espace d'état de façon satisfaisante.

3.2. PROBLÈME INVERSE MULTICOUCHE 1D

Chapitre 4

Diffraction inverse

Introduction

La diffraction inverse est un sujet d'importance majeure. Il comprend des applications très variées [51, 25] en acoustique, optique et en électromagnétisme, par exemple en imagerie médicale, tomographie ou SAR (Synthetic Aperture Radar, ou radar à ouverture synthétique). En électromagnétisme (EM), le problème direct de diffraction consiste en la détermination du champ rétrodiffusé suite à la diffraction d'une onde incidente dans un milieu inhomogène, et cela en connaissant la géométrie et les propriétés physiques du diffracteur. Symétriquement, la diffraction inverse consiste à déduire de l'information sur les inhomogénéités à partir de mesures des champs lointains (voir [25]). Il s'agit donc d'un problème inverse. Dans ce chapitre, nous nous intéressons à un problème plus spécifique de diffraction inverse EM. L'objectif général est d'estimer les propriétés électromagnétiques de matériaux à partir de mesures globales de champs rétrodiffusés à haute fréquence. On touche ici une gamme d'applications que l'on peut situer à l'intersection du contrôle non-destructif, du contrôle de qualité et de la caractérisation de matériaux. De nombreuses techniques de caractérisation EM de matériaux ont été développées dans les domaines des matériaux agroalimentaires [79], des absorbants radar [55], etc. La plupart de ces techniques, de la méthode en ligne de transmission à la méthode en cavité, requièrent des échantillons de matériaux de petite échelle. Il faut ainsi positionner les échantillons sur le porte-échantillon à l'intérieur des lignes de transmission ou guides d'onde. Bien que les propriétés EM (i.e. perméabilité et permittivité) peuvent être mesurées, elles diffèrent sensiblement de celles du produit final, lorsque les matériaux sont assemblés et placés sur l'objet à grande échelle [55]. Les méthodes dites de mesures de SER en espace libre permettent de surmonter cette difficulté. Elles consistent à mesurer la réflectivité monostatique d'un échantillon plan de grande taille. Cet échantillon est situé à l'intérieur d'une chambre anéchoïque, dans le champ lointain d'antennes émettrices et réceptrices. La réflectivité est mesurée pour des ondes incidentes arrivant selon plusieurs angles. Dans ce chapitre, nous nous focalisons sur le problème de diffraction inverse suivant : le contrôle et l'évaluation des propriétés EM d'un objet à partir des mesures de réflectivité globale dans un dispositif

de type SER en espace libre. En d'autres termes, on cherche à déterminer la perméabilité et la permittivité des matériaux le long de l'objet.

Il y a environ cinquante ans, un problème comparable a été souligné dans [2]. Il s'agissait d'appliquer une optimisation de type moindres carrés pour déterminer les constantes diélectriques qui faisaient coïncider les SER calculées analytiquement avec les mesures. Cette problématique revint plus tard au devant de la scène sous une autre forme dans [23] : l'évaluation des coefficients complexes de permittivité et de perméabilité d'un matériau stratifié. Plus récemment, [46] considère la reconstruction par tomographie micro-ondes des propriétés diélectriques d'un objet très inhomogène par un algorithme stochastique d'optimisation globale de type recuit simulé. De façon comparable, [13] développe un algorithme de pseudo-inversion pour l'imagerie 2D, ayant pour objectif la localisation et l'estimation des permittivités diélectriques d'objets diélectriques cylindriques inhomogènes. De façon générale, la diffraction inverse est largement considérée comme un problème inverse mal posé. Comme pour la reconstruction d'image ou d'autres problèmes inverses en imagerie ([74, 40]), cela nécessite d'une façon ou d'une autre une composante de régularisation. Son rôle est de limiter les oscillations artificielles dues au caractère mal posé du problème. Selon [25], les procédures de régularisation se divisent en deux classes : les schémas d'optimisation non-linéaires, et les schémas d'inversion linéaires reposant sur des approximations, telles que l'optique physique ou l'approximation de Born. En outre, il faut mentionner la méthode de "linear sampling" (voir [24, 25]), apte à la reconstruction 3D d'obstacles résultant d'inhomogénéités locales.

Dans ce chapitre, nous développons une approche statistique globale pour formaliser la diffraction inverse de type "SER en champ libre", et résoudre le problème inverse mal posé qui lui correspond. D'une certaine manière, notre approche a sa place dans l'une et l'autre des deux classes de procédures de régularisation que nous avons mentionnées. Elle fait intervenir une méthode d'approximation : l'utilisation intensive d'un solveur des équations de Maxwell, réalisé à l'aide d'une grille de calcul haute performance, permet l'élaboration d'un méta-modèle de vraisemblance approchée. C'est le point de départ d'un modèle statistique complet. Ce modèle mène assez naturellement à un schéma d'inférence proche de l'optimisation, inspiré principalement de la théorie des processus stochastiques et des méthodes de Monte Carlo. Une inférence bayésienne est réalisée par un algorithme SMC.

Depuis une célèbre étude de Geman et Geman [47], les méthodes stochastiques sont couramment utilisées en diffraction inverse, et, plus généralement, en problème inverse en imagerie : recuit simulé pour la reconstruction d'image [75], algorithme EM (EM pour Expectation-Maximization) pour l'imagerie radar [57], etc. En imagerie micro-ondes, [72] évoque les algorithmes génétiques et de façon générale les heuristiques stochastiques, comme les méthodes évolutionnaires, les algorithmes mémétiques, les méthodes PSO et AC (particle swarm optimization, ant colonies), etc. En résumé, de nombreuses tentatives ont été réalisées en électromagnétisme pour appliquer des méthodes stochastiques à la résolution de

4.1. PROBLÈME DE DIFFRACTION INVERSE

problèmes inverses difficiles, et d'optimisation non convexe (citons par exemple le problème de l'absorbance des ondes radar par un matériau multicouche [66, 19]). Malgré une efficacité incontestable, les méthodes d'inversion stochastiques sont souvent mises en difficulté en grande dimension. En conséquence, l'approche que nous proposons prend soin de tirer avantage de la structure spécifique du problème. Elle s'appuie en particulier sur une méthode de réduction de variance des algorithmes de Monte Carlo, appelée Rao-Blackwellisation [61, 30], permettant de réaliser une partie de l'estimation stochastique par des calculs analytiques exacts et rapides.

Le chapitre est organisé de la façon suivante. Dans la section 4.1, nous introduisons les mesures de SER en champ libre, et développons la problématique de diffraction inverse. Dans la section 4.2, nous décrivons une modélisation probabiliste complète : modèle de vraisemblance des observations, modèle a priori sur la solution via une hiérarchie bayésienne et une chaîne de Markov cachée. Enfin, l'algorithme d'inversion est détaillé dans la section 4.3, et ses performances sont analysées d'un point de vue statistique dans la section 4.4.

4.1 Problème de diffraction inverse

4.1.1 Mesures de rétrodiffusion électromagnétique

Les mesures de rétrodiffusion EM datent de l'invention du radar elle-même [55]. Brièvement parlant, la diffraction EM est le phénomène standard qui se produit lorsqu'un objet est exposé à une onde EM, et disperse l'énergie incidente dans toutes les directions, définissant ainsi une distribution spatiale d'énergie. En particulier, une partie de l'énergie revient à la source de l'onde ; il s'agit de l'écho radar de l'objet, dont l'intensité est donnée par la surface équivalente radar (SER) de l'objet. Plus précisément, la SER est définie par :

$$\sigma_s = \lim_{R \rightarrow +\infty} 4\pi R^2 \frac{|\mathbf{E}_{\text{scat}}|^2}{|\mathbf{E}_{\text{inc}}|^2} \quad (4.1)$$

Elle quantifie le pouvoir de rétrodiffusion d'un objet, i.e. le ratio entre la densité de puissance de l'onde réceptionnée \mathbf{E}_{scat} et la densité de puissance de l'onde émise sur l'objet (R désignant la distance radar-objet). Elle est fonction de l'angle d'incidence, de la polarisation de l'onde émise, ainsi que de sa fréquence. Le terme $4\pi R^2$ prend en compte la sphéricité des fronts d'onde. Implicitement, (4.1) est valable lorsque l'onde incidente est plane ($R \rightarrow +\infty$). En pratique, il est possible de mesurer la SER à des distances limitées avec une précision tout à fait acceptable. Ces mesures sont couramment réalisées dans des chambres intérieures de tests SER, appelées chambres anéchoïques, dans lesquelles les interférences peuvent être limitées par des matériaux absorbant les hautes fréquences (voir figure 4.1).

Dans ce chapitre, nous considérons le cas d'un objet illuminé par un radar, c'est-à-dire une antenne unique ou un réseau d'antennes (dont un exemple est donné figure 4.1),

4.1. PROBLÈME DE DIFFRACTION INVERSE

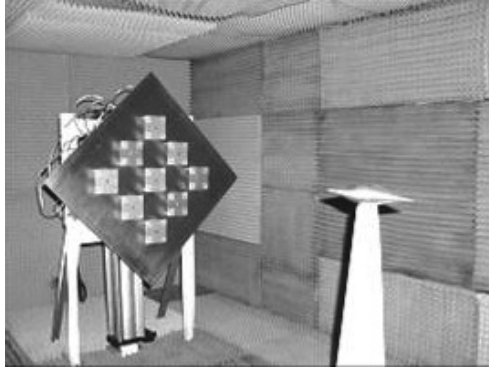


FIGURE 4.1 – Mesure de SER en chambre anéchoïque

satisfaisant à une certaine directivité et à des conditions de champ lointain [68]. Nous nous limitons au cas d'un dispositif radar monostatique, ce qui signifie que l'émetteur et le récepteur sont confondus (voir figure 4.2). En supposant que le radar illumine l'objet à un certain angle d'incidence, avec une onde continue (OC) monochromatique quasi-plane de fréquence f (champ électrique incident \mathbf{E}_{inc}), l'objet rétrodiffuse en retour une OC au radar (champ électrique rétrodiffusé \mathbf{E}_{scat}) à la même fréquence. Au moyen d'un système d'instrumentation approprié (radar, analyseur de réseau, etc.) et d'un processus de calibration, il est possible de mesurer le coefficient de rétrodiffusion, soit la quantité complexe définie par $\mathcal{S} = \frac{\mathbf{E}_{\text{scat}}}{\mathbf{E}_{\text{inc}}}$. Cette quantité résume la diffraction EM en indiquant à la fois les variations d'intensité et de phase de l'onde. \mathcal{S} est lié de près à la SER, via $\sigma_s = |\mathcal{S}|^2$. Il est important de remarquer que le coefficient de rétrodiffusion quantifie une caractéristique globale du comportement électromagnétique de l'objet, et ce dans des conditions spécifiques (incidence, fréquence, polarisations d'émission et de réception).

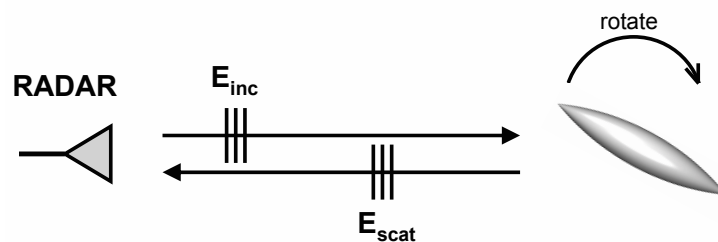


FIGURE 4.2 – Principe de mesure monostatique de rétrodiffusion

Considérons à présent le mode d'acquisition de SER suivant, largement utilisé en imagerie radar ISAR (Inverse Synthetic Aperture Radar). Il consiste à mesurer des coefficients de rétrodiffusion \mathcal{S} :

- à différentes fréquences : $f \in \{f_1, f_2, \dots, f_{K_f}\}$, pour K_f fréquences discrètes successives. On émet pour cela des impulsions à bande étroite (voir méthode SFCW, Stepped

4.1. PROBLÈME DE DIFFRACTION INVERSE

Frequency Continuous Wave, [80]);

- à différents angles : $\theta \in \{\theta_1, \theta_2, \dots, \theta_{K_\theta}\}$, pour K_θ angles d'incidence différents (objet mis en rotation via un support motorisé);
- à différentes polarisations linéaires d'émission et de réception : $\text{pol} \in \{HH, VV\}$, dénotant respectivement les polarisations horizontale et verticale pour à la fois l'émission et la réception.

Notons \mathcal{M} l'ensemble des mesures réalisées, composé des $2 \cdot K_f \cdot K_\theta$ coefficients complexes de rétrodiffusion : $\mathcal{M} = \{\mathcal{S}^{f,\theta,\text{pol}}\}$, pour $f \in \{f_1, \dots, f_{K_f}\}$, $\theta \in \{\theta_1, \dots, \theta_{K_\theta}\}$ et $\text{pol} \in \{HH, VV\}$.

4.1.2 Contrôle non destructif

Le travail présenté dans ce chapitre s'intéresse au contrôle dans un cadre industriel, assimilable au contrôle non destructif (CND). A la différence des techniques usuelles de caractérisation EM de matériaux [55], il s'agit ici de déterminer des propriétés radioélectriques (i.e. la permittivité diélectrique et la perméabilité magnétique relatives) de matériaux qui sont assemblés et placés sur l'objet ou système pleine échelle. Est-ce possible à partir de l'ensemble des mesures \mathcal{M} ? et est-il possible d'extraire des informations locales sur les propriétés des matériaux situés le long de l'objet à partir des mesures globales de rétrodiffusion?

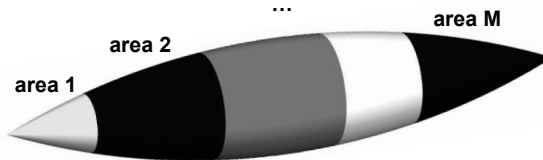


FIGURE 4.3 – Objet recouvert par N_a régions de matériaux

Afin de délimiter l'étude, le chapitre se restreint à un objet métallique axisymétrique, recouvert par N_a régions de matériaux, chaque région correspondant à un matériau relativement homogène, dont les propriétés radioélectriques isotropes varient faiblement dans l'espace. La figure 4.3 donne une illustration avec un objet de forme ogivale inspiré du benchmark [82]. Par conséquent, à partir des mesures globales de rétrodiffusion \mathcal{M} , l'objectif est de déterminer les propriétés EM isotropes locales $(\epsilon_1, \mu_1), (\epsilon_2, \mu_2), \dots, (\epsilon_N, \mu_N)$ le long de l'objet, N désignant le nombre total de zones élémentaires de l'objet (cf. figure 4.4).

4.1.3 Un problème inverse aux équations de Maxwell

Il est assez clair qu'il n'y a pas de modèle direct capable de déterminer les propriétés radioélectriques à partir des informations globales de rétrodiffusion. Au contraire, les modèles

4.1. PROBLÈME DE DIFFRACTION INVERSE

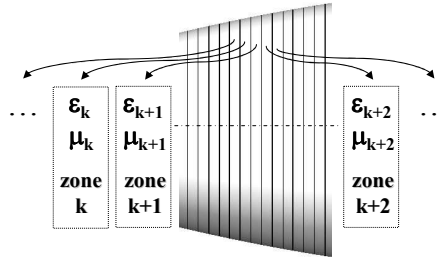


FIGURE 4.4 – Découpage en zones élémentaires

de diffraction permettent l'opération inverse, c'est-à-dire le calcul des coefficients de rétrodiffusion sachant les propriétés EM, la géométrie de l'objet et les conditions d'acquisition (i.e. la fréquence de l'onde, l'angle d'incidence, etc.). Ce calcul repose sur la résolution des équations de Maxwell en régime harmonique, c'est-à-dire des équations aux dérivées partielles qui représentent le processus de diffraction électromagnétique au sein d'un obstacle inhomogène. Il est réalisé à l'aide d'un solveur massivement parallèle, selon une méthode couplée éléments finis-équation intégrale tirant parti de la symétrie de révolution de l'objet. De façon générale, l'étape de discrétisation pose des problèmes de taille de données, notamment dans le cas des hautes fréquences. De plus, comme on le verra par la suite, le processus d'inversion implique que le solveur soit appliqué de nombreuses fois. Il est donc nécessaire d'avoir recours à des calculs haute performance (HPC pour High Performance Computing) : une grille de calcul d'environ vingt mille processeurs, pour une performance supérieure au pétaflop (un million de milliard d'opérations par seconde).

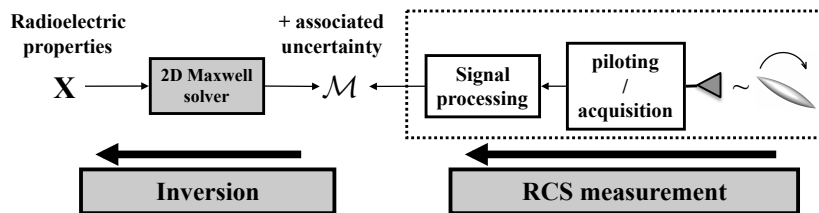


FIGURE 4.5 – Le problème de diffraction inverse

La figure 4.5 résume la problématique de diffraction inverse. D'une part, le processus de mesure de SER, incluant l'acquisition, le traitement du signal, la calibration, etc., fournit les coefficients de rétrodiffusion \mathcal{M} , avec des incertitudes. D'autre part, la nécessité d'inverser le solveur de Maxwell, de façon à déterminer les propriétés radioélectriques inconnues, notées \mathbf{x} . Même à l'aide du HPC, il n'existe pas de moyen direct de réaliser cette inversion, car il s'agit d'un problème mal posé et de surcroît de grande dimension, comparable au problèmes inverses rencontrés en imagerie [40]. Dans la suite, nous proposons une approche statistique globale, capable à la fois de modéliser le problème en tenant compte de toutes

4.2. FORMULATION STATISTIQUE DU PROBLÈME

les informations a priori, et de réaliser l'inversion. De façon similaire à la régularisation de Tikhonov, cette approche tend à éliminer les oscillations artificielles dues au caractère mal posé du problème.

4.2 Formulation statistique du problème

L'approche statistique globale est introduite en deux étapes : d'abord sa formulation à une fréquence fixe f_k , puis le modèle entier pour toutes les fréquences f_1, f_2, \dots, f_{K_f} .

4.2.1 Modélisation à une fréquence fixée f_k

Soit f_k la fréquence d'une impulsion SFCW donnée. Nous définissons ici les trois composantes principales du modèle à la fréquence f_k : l'état du système \mathbf{x}_k , l'observation \mathbf{y}_k , et le lien probabiliste entre eux, i.e. le modèle de vraisemblance $p(\mathbf{y}_k|\mathbf{x}_k)$. Pour alléger les notations, ils sont notés respectivement \mathbf{x} , \mathbf{y} et $p(\mathbf{y}|\mathbf{x})$ dans cette section.

4.2.1.1 Etat du système

Le vecteur $\mathbf{x} = [\underline{\epsilon}' \quad \underline{\epsilon}'' \quad \underline{\mu}' \quad \underline{\mu}'']^T$ inclut les composantes de permittivité et perméabilité relatives des N zones élémentaires ; ' et '' désignant respectivement les parties réelle et imaginaire.¹ (à la fréquence f_k). Les quatre composantes peuvent être décomposées de la façon suivante : $\underline{\epsilon}' = [\epsilon'_1 \dots \epsilon'_N]^T$, $\underline{\epsilon}'' = [\epsilon''_1 \dots \epsilon''_N]^T$, $\underline{\mu}' = [\mu'_1 \dots \mu'_N]^T$ et $\underline{\mu}'' = [\mu''_1 \dots \mu''_N]^T$. L'état du système \mathbf{x} est donc de dimension $4N$; il regroupe tous les paramètres inconnus que l'on souhaite estimer.

4.2.1.2 Observation

$\mathbf{y} = [\Re(\mathcal{S}_{\mathbf{HH}}) \quad \Im(\mathcal{S}_{\mathbf{HH}}) \quad \Re(\mathcal{S}_{\mathbf{VV}}) \quad \Im(\mathcal{S}_{\mathbf{VV}})]^T$ contient les parties réelle ($\Re(\cdot)$) et imaginaire ($\Im(\cdot)$) des coefficients de rétrodiffusion $\mathcal{S}_{\mathbf{HH}}$ et $\mathcal{S}_{\mathbf{VV}}$ mesurés aux K_θ angles $\theta_1, \dots, \theta_{K_\theta}$ (à la fréquence f_k). Les deux quantités complexes $\mathcal{S}_{\mathbf{HH}}$ et $\mathcal{S}_{\mathbf{VV}}$ peuvent être détaillées : $\mathcal{S}_{\mathbf{HH}} = [\mathcal{S}^{f_k, \theta_1, \mathbf{HH}} \quad \mathcal{S}^{f_k, \theta_2, \mathbf{HH}} \quad \dots \quad \mathcal{S}^{f_k, \theta_{K_\theta}, \mathbf{HH}}]^T$ et $\mathcal{S}_{\mathbf{VV}} = [\mathcal{S}^{f_k, \theta_1, \mathbf{VV}} \quad \mathcal{S}^{f_k, \theta_2, \mathbf{VV}} \quad \dots \quad \mathcal{S}^{f_k, \theta_{K_\theta}, \mathbf{VV}}]^T$. La dimension de l'espace d'observation est donc $4 \cdot K_\theta$.

4.2.1.3 Modèle de vraisemblance

$p(\mathbf{y}|\mathbf{x})$ décrit la relation probabiliste entre l'état du système \mathbf{x} et l'observation \mathbf{y} (à la fréquence f_k). En d'autres termes, il décrit la densité de probabilité de l'observation \mathbf{y} sachant un état \mathbf{x} donné. C'est un élément clef de la connaissance à prendre en compte. L'objectif de notre inférence est en quelque sorte d'inverser cette relation. Le modèle de vraisemblance peut être exprimé comme la loi d'un vecteur gaussien, d'espérance $\mathcal{F}_{\text{Maxwell}}(\mathbf{x})$ et

1. En d'autres termes, $\epsilon = \epsilon' + j\epsilon''$ et $\mu = \mu' + j\mu''$ (avec la convention de dépendance en temps $e^{j\omega t}$).

4.2. FORMULATION STATISTIQUE DU PROBLÈME

de matrice de covariance \mathbf{R}_m :

$$\mathbf{y}|\mathbf{x} \sim \mathcal{N}(\mathcal{F}_{\text{Maxwell}}(\mathbf{x}), \mathbf{R}_m) \quad (4.2)$$

où $\mathcal{F}_{\text{Maxwell}}$ désigne le modèle direct, de l'espace d'état dans l'espace d'observation, associé au solveur des équations de Maxwell mentionné ci-avant. En tenant compte des incertitudes de mesure, le modèle de vraisemblance résulte des considérations suivantes.

- Le solveur, basé sur une méthode directe, est exact, c'est-à-dire extrêmement précis. On suppose donc que $\mathcal{F}_{\text{Maxwell}}$ calcule "l'observation parfaite", soit vierge de tout bruit de mesure, biais, etc. Implicitement, on suppose que la forme de l'objet est parfaitement connue, et que conditionnellement aux propriétés radioélectriques, l'incertitude ne provient que de la mesure.
- D'après de précédentes études sur les incertitudes de mesure (voir principes de métrologie [1]), il est démontré que les incertitudes de mesure peuvent être raisonnablement modélisées par un bruit gaussien additif ($\mathbf{y} = \mathcal{F}_{\text{Maxwell}}(\mathbf{x}) + \mathbf{v}_m$, $\mathbf{v}_m \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_m)$) de matrice de covariance \mathbf{R}_m .

Par conséquent, le modèle de vraisemblance peut être écrit de la façon suivante (avec $\nu = 4 \cdot K_\theta$) :

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{\frac{\nu}{2}} \sqrt{\det \mathbf{R}_m}} e^{-\frac{1}{2}(\mathbf{y} - \mathcal{F}_{\text{Maxwell}}(\mathbf{x}))^T \mathbf{R}_m^{-1} (\mathbf{y} - \mathcal{F}_{\text{Maxwell}}(\mathbf{x}))} \quad (4.3)$$

A première vue, en considérant une unique fréquence f_k , la résolution du problème inverse requiert de nombreuses évaluations de $p(\mathbf{y}|\mathbf{x})$, c'est-à-dire du solveur $\mathcal{F}_{\text{Maxwell}}(\mathbf{x})$. Or, ces évaluations sont beaucoup trop coûteuses en temps de calcul, même avec le calcul haute performance. Pour éviter une utilisation trop lourde de $\mathcal{F}_{\text{Maxwell}}$, nous avons opté pour une approche de type apprentissage statistique. Son principe de base est de construire un modèle de remplacement, soit une approximation de $\mathcal{F}_{\text{Maxwell}}$ qui soit acceptable dans le domaine limité de l'espace d'état qui nous intéresse. En un sens, ce modèle est lié aux méthodes par approximation linéaire décrites dans [25] en diffraction inverse, et plus précisément l'approximation de Born [51, 25], couramment utilisée. Ici, la linéarisation statistique n'est pas construite comme une troncature des interactions physiques, mais comme une approximation en aval du modèle entier des équations de Maxwell, qui prend en compte les multiples interactions comme les ondes rampantes, etc. Le système, c'est-à-dire l'espace d'état de \mathbf{x} , de grande dimension, et les réponses associées $\mathcal{F}_{\text{Maxwell}}(\mathbf{x})$, est exploré via un échantillonnage aléatoire, conformément à la connaissance a priori que l'on a sur les propriétés radioélectriques recherchées (densité a priori $p(\mathbf{x})$). Les calculs sont massivement distribués sur un grille de processeurs HPC, chaque calcul faisant intervenir le solveur de Maxwell parallélisé. Le nombre de calculs dépend essentiellement de la dimension de l'espace d'état. Le processus de simulation de Monte Carlo conduit à la base d'apprentissage suivante :

$$\mathcal{B} = \{(\mathbf{x}^{(1)}, \mathbf{y}^{(1)}), (\mathbf{x}^{(2)}, \mathbf{y}^{(2)}), \dots, (\mathbf{x}^{(N_S)}, \mathbf{y}^{(N_S)})\} \quad (4.4)$$

4.2. FORMULATION STATISTIQUE DU PROBLÈME

où $\mathbf{x}^{(k)} \sim p(\mathbf{x})$ et $\mathbf{y}^{(k)} = \mathcal{F}_{\text{Maxwell}}(\mathbf{x}^{(k)})$ (pour $k = 1, \dots, N_S$), N_S désignant la taille de l'échantillon. La régression linéaire multidimensionnelle fournit un moyen simple et efficace de construire un modèle linéaire $\mathbf{y} = f(\mathbf{x}) + \mathbf{v}_1$ (\mathbf{v}_1 étant un terme d'erreur de linéarisation) avec

$$f(\mathbf{x}) = \mathbf{A} \cdot \mathbf{x} + \mathbf{y}^0 \text{ or } f(\mathbf{x}) = \mathbf{A}^* \cdot [1 \quad \mathbf{x}], \quad \mathbf{A}^* = [\mathbf{y}^0 \quad \mathbf{A}]. \quad (4.5)$$

\mathbf{A}^* , l'estimateur des moindres carrés de la matrice de paramètres, est donné par la relation

$$\mathbf{A}^* = (\mathcal{X}_{\mathcal{B}}^T \cdot \mathcal{X}_{\mathcal{B}})^{-1} \mathcal{X}_{\mathcal{B}}^T \mathcal{Y}_{\mathcal{B}} \quad \text{avec} \quad \mathcal{X}_{\mathcal{B}} = \begin{bmatrix} 1 & \mathbf{x}^{(1)} \\ 1 & \mathbf{x}^{(2)} \\ \dots & \dots \\ 1 & \mathbf{x}^{(N_S)} \end{bmatrix}, \quad \mathcal{Y}_{\mathcal{B}} = \begin{bmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \\ \dots \\ \mathbf{y}^{(N_S)} \end{bmatrix}. \quad (4.6)$$

Pour des raisons de stabilité numérique, une décomposition QR de $\mathcal{X}_{\mathcal{B}}$ est mise en oeuvre. Par analyse des résidus, il est alors possible d'évaluer la qualité d'ajustement du modèle linéaire, c'est-à-dire de déterminer l'écart entre les données et le modèle dans le domaine d'intérêt. En principe, la quantification de la matrice de covariance (\mathbf{R}_l), associée à l'erreur de linéarisation \mathbf{v}_1 , requiert un jeu de données supplémentaires, voire l'application de méthodes de validation croisée. Il faut remarquer qu'une analyse statistique peut être menée en complément pour extraire des modèles réduits et supprimer les variables explicatives inutiles, telles que des composantes de perméabilité ou permittivité dans certains sous-ensembles de zones. Cette réduction de dimension dépend des phénomènes physiques impliqués, c'est-à-dire de l'interaction onde électromagnétique-objet, qui diffèrent selon la bande fréquentielle d'excursion.

Si l'on revient au modèle de vraisemblance (4.2), cela aboutit à un terme d'erreur global $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_m$, de matrice de covariance \mathbf{R}^2 , et au modèle de vraisemblance linéaire gaussien suivant (en réintroduisant l'indice k de la fréquence f_k) :

$$\mathbf{y}_k | \mathbf{x}_k \sim \mathcal{N}(\mathbf{A}_k \cdot \mathbf{x}_k + \mathbf{y}_k^0, \mathbf{R}_k) \quad \text{ou} \quad \mathbf{y}_k = [\mathbf{A}_k \cdot \mathbf{x}_k + \mathbf{y}_k^0] + \mathbf{v}_k \quad (4.7)$$

où \mathbf{A}_k et \mathbf{y}_k^0 sont déduits de la base d'apprentissage \mathcal{B}_k . La matrice \mathbf{A}_k est illustrée figure 4.6 pour l'objet ogival de la figure 4.3 ($N = 137$, $f = 1.5$ GHz, $\theta = 0^\circ : 1^\circ : 180^\circ$ - exploration : 1000 simulations HPC de $\mathcal{F}_{\text{Maxwell}}$). Le motif que l'on distingue clairement au sein de chaque bloc de la matrice s'explique par la contribution cohérente de chaque zone élémentaire.

4.2.1.4 Approche bayésienne

Même si une inversion à fréquence fixée f_k peut être réalisée par une méthode classique de régularisation [40], l'estimation bayésienne offre un cadre d'analyse pratique et puissant. On probabilise le vecteur d'état (inconnu) \mathbf{x}_k pour considérer sa densité de probabilité a priori $p(\mathbf{x}_k)$. Il est possible de modéliser la connaissance a priori sur ce vecteur au moyen d'une densité gaussienne : $\mathbf{x}_k \sim \mathcal{N}(\mathbf{m}_k, \mathbf{P}_k)$.

2. Dans notre contexte, l'erreur de linéarisation se trouve être négligeable par rapport au bruit de mesure de la SER : $\mathbf{R} + \mathbf{R}_l \simeq \mathbf{R}$.

4.2. FORMULATION STATISTIQUE DU PROBLÈME

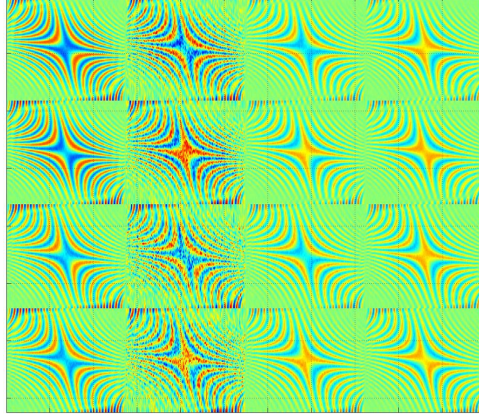


FIGURE 4.6 – Illustration de la matrice \mathbf{A}_k

L'espérance \mathbf{m}_k (de dimension N) définit les propriétés radioélectriques de référence pour les N_a régions qui partitionnent l'objet (cf. figure 4.3).

$$\mathbf{m}_k = \left[\mathbf{m}_k^{\epsilon'} \quad \mathbf{m}_k^{\epsilon''} \quad \mathbf{m}_k^{\mu'} \quad \mathbf{m}_k^{\mu''} \right]^T \quad (4.8)$$

où $\mathbf{m}_k^{\epsilon'} = \left[\underbrace{\epsilon'_k(1) \cdots \epsilon'_k(1)}_{\text{région 1}} \quad \underbrace{\epsilon'_k(2) \cdots \epsilon'_k(2)}_{\text{région 2}} \quad \cdots \quad \underbrace{\epsilon'_k(N_a) \cdots \epsilon'_k(N_a)}_{\text{région } N_a} \right]^T$; $\epsilon'_k(i)$ désignant

la partie réelle de la permittivité de référence pour la région i ($i = 1 \cdots N_a$). La construction pour $\mathbf{m}_k^{\epsilon''}$, $\mathbf{m}_k^{\mu'}$ et $\mathbf{m}_k^{\mu''}$ est similaire.

La matrice de covariance \mathbf{P}_k (de dimension $N \times N$) quantifie l'incertitude de l'a priori autour de \mathbf{m}_k . \mathbf{P}_k est diagonale par bloc : $\mathbf{P}_k = \text{diag}(\mathbf{P}_k^{\epsilon'}, \mathbf{P}_k^{\epsilon''}, \mathbf{P}_k^{\mu'}, \mathbf{P}_k^{\mu''})$. Cela signifie que les quantités $(\epsilon', \epsilon'', \mu', \mu'')$ sont supposées décorrélées. De plus, chaque bloc admet lui-même une structure par blocs. Par exemple, $\mathbf{P}_k^{\epsilon'} = \text{diag}(\mathbf{P}_k^{\epsilon'}(1), (\mathbf{P}_k^{\epsilon'}(2), \cdots, (\mathbf{P}_k^{\epsilon'}(N_a)))$, ce qui traduit une hypothèse d'indépendance entre les différentes régions. Si l'on se focalise sur un bloc $\mathbf{P}_k^{\epsilon'}(i)$, une fonction de covariance à décroissance exponentielle traduit l'homogénéité spatiale entre les composantes, c'est-à-dire entre les zones élémentaires de l'objet qui appartiennent à la même région i :

$$\mathbf{P}_k^{\epsilon'}(i) = \left[\sigma_k^{\epsilon'}(i) \right]^2 \times \begin{bmatrix} 1 & \rho_S & \rho_S^2 & \cdots & \rho_S^{n-1} \\ \rho_S & 1 & \rho_S & & \vdots \\ \rho_S^2 & \rho_S & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \rho_S \\ \rho_S^{n-1} & \cdots & \cdots & \rho_S & 1 \end{bmatrix} \quad (4.9)$$

où $\left[\sigma_k^{\epsilon'}(i) \right]^2$ est la variance spatiale de la $i^{\text{ème}}$ région, et $\rho_S \in [0, 1]$ le paramètre de corrélation spatiale normalisé (typiquement $\rho_S = 0.95$). En utilisant cette propriété markovienne, couramment utilisée dans les modélisations de type champ gaussien, la

4.2. FORMULATION STATISTIQUE DU PROBLÈME

corrélation décroît géométriquement en fonction de la distance entre les composantes. Sans surprise, on opère une construction similaire pour $\mathbf{P}_k^{\epsilon''}$, $\mathbf{P}_k^{\mu'}$ et $\mathbf{P}_k^{\mu''}$.

Dans ce cadre linéaire gaussien, avec des a priori gaussiens et des modèles de vraisemblance linéaires gaussiens, l'inversion bayésienne peut être réalisée directement selon des solutions simples et explicites. Dans notre problème, il s'agit là d'une maille élémentaire au sein d'une construction globale plus complexe, car prenant en compte les variations de fréquence.

4.2.2 Formulation du problème global

Il est bien connu que les propriétés radioélectriques varient en fonction de la fréquence de l'onde [55]. Elles peuvent même différer totalement entre la fréquence la plus basse f_1 et la plus haute f_{K_f} . L'idée centrale est de conserver le modèle statistique à fréquence fixée décrit précédemment tout en introduisant de l'information a priori sur le comportement dynamique en fréquence. En d'autres termes, à quelle vitesse peut varier une propriété radioélectrique en fonction de la fréquence, quel est le degré de corrélation entre deux fréquences différentes, etc. Cette information sur la régularité de notre inconnue peut être elle-même très différente d'une propriété EM (ϵ' , ϵ'' , μ' , μ'') à une autre, tout comme d'un matériau à un autre.

4.2.3 Processus autorégressif généralisé

L'extension de notre modèle statistique consiste en la modélisation de toute la suite ($\mathbf{x}_k, k \in \{1, \dots, K_f\}$) par un processus aléatoire que l'on peut interpréter comme une généralisation des processus autorégressifs (AR) :

$$\begin{aligned} \mathbf{x}_1 &\sim \mathcal{N}(\mathbf{m}_1, \mathbf{P}_1) \\ \mathbf{x}_{k+1} &= \mathbf{m}_{k+1} + \mathbf{D}_\rho \cdot \mathbf{H}_{k+1} \cdot \mathbf{H}_k^{-1} \cdot (\mathbf{x}_k - \mathbf{m}_k) + \sqrt{\mathbf{I}_d - \mathbf{D}_\rho^2} \cdot \mathbf{H}_{k+1} \cdot \mathbf{V}_k \end{aligned} \quad (4.10)$$

où \mathbf{H}_k désigne la racine carrée de la matrice de covariance \mathbf{P}_k ³. Les ($\mathbf{V}_k, k \in \{1, \dots, K\}$) sont i.i.d. selon $\mathcal{N}(0, \mathbf{I}_d)$ et \mathbf{D}_ρ est une matrice diagonale positive qui commute avec \mathbf{H}_k . Ce modèle dynamique traduit la structure de corrélation linéaire gaussienne. On vérifie facilement que les lois marginales des \mathbf{x}_k restent les gaussiennes $\mathcal{N}(\mathbf{m}_k, \mathbf{P}_k)$. Plus généralement, on peut montrer que la loi du vecteur concaténé $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_{K_f})$ est gaussienne, d'espérance $\mathbf{m} = (\mathbf{m}_1, \dots, \mathbf{m}_{K_f})$ et de matrice de covariance :

$$\mathbf{P} = \mathcal{H} \cdot \begin{bmatrix} \mathbf{I}_d & \mathbf{D}_\rho & \mathbf{D}_\rho^2 & \dots & \mathbf{D}_\rho^{K_f-1} \\ \mathbf{D}_\rho & \mathbf{I}_d & \mathbf{D}_\rho & & \vdots \\ \mathbf{D}_\rho^2 & \mathbf{D}_\rho & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \mathbf{D}_\rho \\ \mathbf{D}_\rho^{K_f-1} & \dots & \dots & \mathbf{D}_\rho & \mathbf{I}_d \end{bmatrix} \cdot \mathcal{H}^T \quad (4.11)$$

3. soit l'unique matrice définie positive vérifiant : $\mathbf{H}_k \cdot \mathbf{H}_k^T = \mathbf{P}_k$.

4.2. FORMULATION STATISTIQUE DU PROBLÈME

où \mathcal{H} est la matrice diagonale par bloc $\mathcal{H} = \text{diag}(\mathbf{H}_1, \dots, \mathbf{H}_{K_f})$. Ce résultat permet en particulier d'expliciter toutes les lois jointes $(\mathbf{x}_i, \mathbf{x}_j)$.

Intéressons-nous maintenant à la matrice \mathbf{D}_ρ . Son rôle est de prendre en compte les corrélations fréquentielles entre les propriétés EM $\mathbf{x}_1 \cdots \mathbf{x}_{K_f}$. Elle se réfère à un paramètre ρ . Selon les différentes connaissances a priori dont on dispose, ou que l'on suppose, plusieurs possibilités sont proposées :

1. La corrélation fréquentielle ne dépend ni du matériau, ni de la propriété EM considérée (ϵ' , ϵ'' , μ' ou μ'') : ρ est un simple scalaire ($\in [0, 1]$) et $\mathbf{D}_\rho = \rho \cdot \mathbf{I}_d$.
2. La corrélation fréquentielle dépend du matériau, mais pas de la propriété EM : ρ est de dimension N_a ($\in [0, 1]^{N_a}$), et \mathbf{D}_ρ est la matrice diagonale par bloc, composée des N_a blocs $\rho_i \cdot \mathbf{I}_d$.
3. La corrélation fréquentielle dépend à la fois du matériau et de la propriété EM considérée : ρ est de dimension $4 \cdot N_a$ et \mathbf{D}_ρ est la matrice diagonale par bloc composée des $4 \cdot N_a$ blocs $\rho_i \cdot \mathbf{I}_d$.

4.2.4 Structure conditionnelle et chaîne de Markov cachée

Le processus AR généralisé inclut le modèle linéaire gaussien pour chacune des fréquences f_k ($k = 1 \cdots K_f$). Il fournit une double structure de corrélation, spatiale et fréquentielle. En supposant que les régions de matériaux sont raisonnablement homogènes, le paramètre de corrélation spatiale peut être fixé (typiquement $\rho_S = 0.95$). A l'inverse, les corrélations fréquentielles sont a priori impossibles à conjecturer ; elles sont donc à estimer elles aussi dans le processus d'inversion. Conformément à notre logique bayésienne, on choisit de probabiliser le paramètre ρ . Finalement, la combinaison du modèle dynamique de type AR (4.11) et du modèle de vraisemblance (4.7) aboutit au modèle suivant :

$$\mathbf{x}_{k+1} = \mathbf{M}_k^\rho \cdot \mathbf{x}_k + \mathbf{w}_k \quad \mathbf{y}_k = \left[\mathbf{A}_k \cdot \mathbf{x}_k + \mathbf{y}_k^0 \right] + \mathbf{v}_k \quad (4.12)$$

auquel il faut rajouter l'hypothèse sur l'état initial $\mathbf{x}_1 \sim \mathcal{N}(\mathbf{m}_1, \mathbf{P}_1)$. Les fréquences f_k ($k = 1, \dots, K_f$) jouent en quelque sorte le rôle d'indice temporel. Précisons que la matrice de transition \mathbf{M}_k^ρ et le vecteur de bruit gaussien (de moyenne non nulle) se déduisent l'une et l'autre directement de (4.11).

Soulignons bien les points suivants : le modèle dynamique implique que chaque loi marginale vérifie $\mathbf{x}_k \sim \mathcal{N}(\mathbf{m}_k, \mathbf{P}_k)$. D'autre part, conditionnellement au paramètre de corrélation fréquentielle ρ , le modèle que l'on considère est un simple processus markovien linéaire gaussien. Une représentation graphique du modèle est proposée figure 4.7. A ρ donné, la partie basse du schéma décrit donc un système linéaire gaussien. L'idée centrale de l'inversion est de tirer profit de cette structure spécifique.

4.3. INVERSION PROBABILISTE

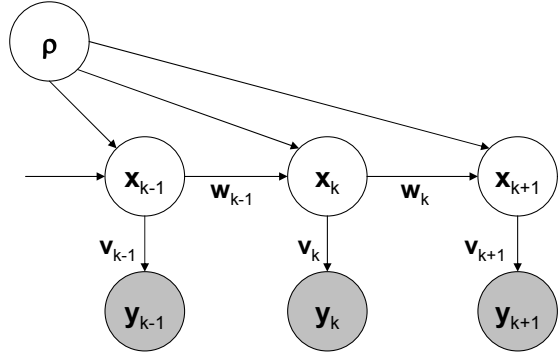


FIGURE 4.7 – Représentation graphique du modèle hiérarchique

4.3 Inversion probabiliste

4.3.1 L'approche Rao-Blackwellisée

Soit $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_{K_f})$ le vecteur des mesures. Comme nous l'avons mentionné auparavant, le paramètre ρ est probabilisé, c'est-à-dire qu'on lui associe en particulier une loi a priori $p(\rho)$, que l'on suppose calculable (au moins à une constante multiplicative près) et facile à simuler. La loi a posteriori $p(\mathbf{x}, \rho | \mathbf{y})$ peut être décomposée de la façon suivante :

$$p(\mathbf{x}, \rho | \mathbf{y}) = p(\mathbf{x} | \rho, \mathbf{y}) \cdot p(\rho | \mathbf{y}).$$

Conditionnellement à ρ , le système est linéaire gaussien, ce qui implique en particulier que la loi conditionnelle $p(\mathbf{x} | \rho, \mathbf{y})$ peut être obtenue directement par filtrage de Kalman, filtrage facilement enrichi, dans ce contexte statique, par une étape de lissage. D'autre part, le terme $p(\rho | \mathbf{y})$ peut être lui-même décomposé ainsi :

$$\begin{aligned} p(\rho | \mathbf{y}) &\propto p(\rho) \cdot p(\mathbf{y} | \rho) \\ &\propto p(\rho) \cdot \prod_{k=1}^{K_f} \underbrace{p(\mathbf{y}_k | \rho, \mathbf{y}_1, \dots, \mathbf{y}_{k-1})}_{:= J_k(\rho)}. \end{aligned}$$

Remarquons que pour tout paramètre ρ , les quantités $J_k(\rho)$ peuvent être calculées à l'aide des termes de vraisemblance fournies par le filtre de Kalman.

Par conséquent, pour exploiter cette structure conditionnelle du système, les lisseurs de Kalman sont appliqués et intégrés dans une approche de type système de particules en interaction. Dans un premier temps, un algorithme SMC se charge de fournir une approximation de la densité $p(\rho | \mathbf{y})$ (voir section 4.3.2 ci-après). Intuitivement, il s'agit d'une estimation de la régularité (corrélation fréquentielle) des profils fréquents recherchés $\epsilon'(f)$, $\epsilon''(f)$, $\mu'(f)$, $\mu''(f)$. Dans un second temps, on utilise les relations conditionnelles

$$\mathbb{E}(\mathbf{x}_k | \mathbf{y}) = \mathbb{E}[\mathbb{E}(\mathbf{x}_k | \rho, \mathbf{y}) | \mathbf{y}] \quad (4.13)$$

$$\text{Var}(\mathbf{x}_k|\mathbf{y}) = \mathbb{E} [\text{Var}(\mathbf{x}_k|\rho, \mathbf{y})|\mathbf{y}] + \text{Var} [\mathbb{E}(\mathbf{x}_k|\rho, \mathbf{y})|\mathbf{y}] \quad (4.14)$$

pour en déduire des estimations de respectivement l'espérance et la matrice de covariance a posteriori de \mathbf{x}_k , ce pour chaque fréquence f_k . Les équations de Kalman sont donc employées pour calculer à la fois la vraisemblance du paramètre ρ (à une constante multiplicative près) dans la première étape, et les quantités $\mathbb{E}(\mathbf{x}_k|\rho, \mathbf{y})$ et $\text{Var}(\mathbf{x}_k|\rho, \mathbf{y})$ dans la seconde. L'idée générale de mélanger le calcul analytique (ici, l'évaluation de Kalman de $p(\mathbf{x}|\rho, \mathbf{y})$) avec l'échantillonnage stochastique (ici, approcher $p(\rho|\mathbf{y})$) est une méthode de réduction de variance, appelée Rao-Blackwellisation [61].

Fixons le vecteur d'observation \mathbf{y} et notons $\eta(d\rho)$ la mesure de probabilité associée à la loi marginale $p(\rho|\mathbf{y})$. De façon similaire à [61], on choisit d'implémenter un algorithme SMC dans le but d'estimer η .

4.3.2 Algorithme SMC et chemin d'interpolation dans l'espace des mesures de probabilité

L'algorithme SMC (pour Sequential Monte Carlo) est un algorithme particulière séquentiel utilisé pour échantillonner une suite de mesures de probabilité η_n évoluant selon un flot de Feynman-Kac (voir section 1.3 page 10 pour une introduction à ces modèles). Il est souvent utilisé pour résoudre des problèmes intrinsèquement séquentiels, comme le filtrage (voir par exemple [14, 43, 27] ou la section 1.3.3 page 13 pour une introduction). Cependant, dans d'autres types de problèmes, cet algorithme peut s'avérer efficace pour l'échantillonnage d'une mesure de probabilité η unique. Dans ce contexte, l'idée centrale est de trouver une suite d'interpolation $(\eta_n)_{0 \leq k \leq n_f}$ judicieuse, de complexité croissante, pour faire le pont entre une mesure initiale η_0 que l'on sait simuler et la dernière mesure $\eta_{n_f} = \eta$ choisie comme étant la mesure cible. Il importe que les mesures consécutives η_n et η_{n+1} soient proches, assez pour pouvoir par exemple passer de l'une à l'autre par un échantillonnage d'importance ou une méthode d'acceptation-rejet. L'aspect séquentiel de l'approche est alors un moyen artificiel pour introduire progressivement la difficulté d'échantillonnage. Plus généralement, un point essentiel est que l'utilisation d'un grand nombre N_p de particules permet de couvrir plusieurs modes simultanément si nécessaire. C'est notamment un avantage vis-à-vis des méthodes MCMC standards qui sont plus enclines à rester piégées dans des modes locaux.

Soit E l'espace d'état de la variable ρ (i.e. $E = [0, 1]$, $[0, 1]^{N_a}$ ou $[0, 1]^{4N_a}$). Pour simuler la loi η au moyen d'une méthode SMC, il nous faut définir une suite de mesures de probabilité $(\eta_n)_{0 \leq k \leq n_f}$, la première devant être la loi initiale $\eta_0(d\rho) = p(\rho)d\rho$ (facile à simuler) et la dernière étant la loi cible $\eta_{n_f}(d\rho) = \eta(d\rho) = p(\rho|\mathbf{y})d\rho$. En gardant à l'esprit cet objectif, rappelons tout d'abord la définition et l'importance des transitions markoviennes de Metropolis-Hastings, introduites page 3. Soit h une fonction strictement positive et bornée sur E , et soit $Q(x, dy)$ une transition markovienne sur E de la forme $Q(x, dy) = T(x, y)dy$, où pour tout x , $T(x, \cdot)$ est une densité de probabilité connue et calculable. Le noyau (ou

4.3. INVERSION PROBABILISTE

la transition) de Metropolis-Hastings $K_{h,T}(x, dy)$ associé à la fonction h et à la fonction de transition T est défini par les formules suivantes :

$$K_{h,T}(x, dy) = Q(x, dy) \cdot \min(1, r(x, y)) \quad \forall y \neq x$$

$$K_{h,T}(x, \{x\}) = 1 - \int_{y \neq x} Q(x, dy) \cdot \min(1, r(x, y))$$

$$\text{où } r(x, y) := \frac{h(y) T(y, x)}{h(x) T(x, y)}.$$

En utilisant une méthode d'acceptation/rejet, ce noyau est facile à simuler dès que l'on sait simuler $Q(x, dy)$ et calculer les ratios $h(y)/h(x)$.

Vient alors une propriété cruciale : si μ_h désigne une mesure de probabilité définie par $\mu_h(d\rho) \propto h(\rho)d\rho$, alors il est bien connu (voir par exemple [33]) que $K_{h,T}$ laisse la mesure μ_h invariante, i.e.

$$\mu_h \cdot K_{h,T} = \mu_h \quad \left(\iff \int_E K_{h,T}(\rho, A) \mu_h(d\rho) = \mu_h(A), \forall A \subset E \right).$$

Par suite, pour tout entier m , cette propriété est vérifiée pour le noyau itéré $K_{h,T}^m$, i.e. $\mu_h \cdot K_{h,T}^m = \mu_h$.

On en déduit la conséquence suivante : soit η_n une suite de mesures de probabilité définies à l'aide de fonctions strictement positives bornées h_n par $\eta_n(d\rho) \propto h_n(\rho).d\rho$. Alors pour toute suite de noyaux de Markov Q_n de la forme $Q_n(x, dy) = T_n(x, y)dy$ et pour toute suite d'entiers m_n , η_n vérifie la formule de Feynman-Kac (1.6) (voir page 12) avec les potentiels $G_n := h_n/h_{n-1}$ et les noyaux de Markov $M_n := K_{h_n, T_n}^{m_n}$. En pratique, cela implique qu'une telle suite de mesures η_n peut être approchée au moyen d'un algorithme SMC (voir la section 1.3.1 page 10 pour son implémentation) à partir du moment où l'on sait calculer les fonctions h_n à une constante multiplicative près. A l'instar des méthodes MCMC traditionnelles, ou des méthodes de recuit simulé, l'algorithme SMC est d'autant plus robuste que les nombres d'itération m_n sont grands, dans la mesure où les noyaux K_{h_n, T_n} sont justement définis et utilisés pour stabiliser le système (voir par exemple la remarque page 19 de l'équation (2.1) ou le théorème 6 page 40 pour un résultat théorique).

A l'aide de ces outils, trois schémas d'interpolation dans l'espace des mesures sont proposés.

- Le schéma du recuit : la suite η_n est définie par le biais des fonctions bornées strictement positives

$$h_n(\rho) = p(\mathbf{y}|\rho)^{\alpha_n} \cdot p(\rho)$$

où $(\alpha_n)_{1 \leq n \leq n_f}$ est une suite de nombres allant croissant de 0 à 1 (suite choisie arbitrairement). Dans cette situation, les potentiels $G_n(\rho)$ employés au cours de l'étape de sélection valent $p(\mathbf{y}|\rho)^{\alpha_n - \alpha_{n-1}}$. La suite α_n est donc à choisir de façon à contrôler la sélectivité de ces fonctions, contrôle qui s'avère déterminant en pratique.

- Le schéma séquentiel sur les données : Pour tout $n \in \{0, 1, \dots, K_f\}$, η_n est la mesure de probabilité associée à

$$h_n(\rho) = p(\rho) \cdot \prod_{k=1}^n \underbrace{p(\mathbf{y}_k | \rho, \mathbf{y}_1, \dots, \mathbf{y}_{k-1})}_{=J_k(\rho)}.$$

Autrement dit, pour chaque génération n , le potentiel de sélection $G_n(\rho)$ appliqué aux particules vaut $p(\mathbf{y}_n | \rho, \mathbf{y}_1, \dots, \mathbf{y}_{n-1})$, soit la vraisemblance du n -ième vecteur d'observation conditionnellement aux précédents. Cela autorise l'algorithme à éventuellement fonctionner en temps réel, car il traite les observations une à une, dans leur ordre d'acquisition. D'après [53], cette méthode est efficace dans le cas de problèmes faisant intervenir un ordre naturel (par exemple le cas des chaînes de Markov cachées). Néanmoins, lorsque ces potentiels se trouvent être trop sélectifs, l'algorithme SMC peut s'avérer inefficace, car le nuage de particules, sujet à des purges trop brutales, perd toute sa diversité à chaque étape de sélection. Le modèle suivant propose une variante, conçue pour pallier cet inconvénient.

- Le schéma hybride : à l'instar du précédent, ce schéma incorpore les observations l'une après l'autre, mais chaque fonction de vraisemblance $J_k(\rho)$ est alors décomposée selon le produit suivant :

$$J_k(\rho) = \prod_{i=1}^{n_k} J_k(\rho)^{(\alpha_i^{(k)} - \alpha_{i-1}^{(k)})}$$

où pour tout $k \in \{1, \dots, K_f\}$, $(\alpha_i^{(k)})_{1 \leq i \leq n_k}$ est une suite $0 \nearrow 1$. Ainsi, si $n = (n_1 + \dots + n_{r-1}) + s$, la fonction h_n est définie par :

$$h_n(\rho) = p(\rho) \cdot \left(\prod_{k=1}^{r-1} J_k(\rho) \right) \cdot J_r(\rho)^{\alpha_s^{(r)}}$$

De cette façon, chaque potentiel de sélection $G_n = J_r^{(\alpha_s^{(r)} - \alpha_{s-1}^{(r)})}$ peut être contrôlé arbitrairement.

Pour chacun de ces schémas d'interpolation, les fonctions h_n sont calculables à une constante de normalisation près (équations de Kalman), de sorte que les noyaux de Metropolis-Hastings (éventuellement itérés) peuvent être utilisés pour réaliser les étapes de mutation.

4.3.3 Estimation globale

De façon résumée, la loi jointe $p(\mathbf{x}, \rho | \mathbf{y})$ peut être décomposée et estimée de la façon suivante :

$$p(\mathbf{x}, \rho | \mathbf{y}) = \underbrace{p(\mathbf{x} | \rho, \mathbf{y})}_{\text{FK (+ lissage)}} \cdot \underbrace{\overbrace{p(\mathbf{y} | \rho)}^{\text{fourni par FK}} \cdot \overbrace{p(\rho)}^{\text{prior}}}_{p(\rho | \mathbf{y})}_{\text{SMC}}.$$

4.3. INVERSION PROBABILISTE

Comme nous l'avons mentionné plus avant, l'algorithme SMC fournit dans le premier temps de l'inversion une estimation de la corrélation fréquentielle $p(\rho|\mathbf{y})$ (i.e. une approximation $\hat{\eta} = \eta_{n_f}^{N_p}$ de η). Elle est définie comme étant la mesure empirique associée à la dernière génération de particules $(\rho^{(1)}, \dots, \rho^{(N_p)}) := (\zeta_{n_f}^1, \dots, \zeta_{n_f}^{N_p})$, N_p désignant le nombre de particules du processus. Dans le second temps de l'inversion, des estimations des propriétés EM sont calculées en s'inspirant des relations conditionnelles (4.13) et (4.14) (voir page 95). Cela consiste à approcher l'espérance et la matrice de covariance a posteriori de chaque vecteur \mathbf{x}_k , soit $\bar{\mathbf{x}}_k := \mathbb{E}[\mathbf{x}_k|\mathbf{y}]$, et $\Sigma_k := \text{Var}[\mathbf{x}_k|\mathbf{y}]$. Plus précisément, ces deux estimateurs $\hat{\mathbf{x}}_k$ et $\hat{\Sigma}_k$ sont définis de la façon suivante.

Pour tout $\rho \in E$, notons

$$\hat{\mathbf{x}}_k(\rho) := \mathbb{E}[\mathbf{x}_k|\rho, \mathbf{y}] \quad \text{et} \quad \hat{\Sigma}_k(\rho) := \text{Var}[\mathbf{x}_k|\rho, \mathbf{y}]$$

soit les principales quantités fournies par le lisseur de Kalman. Sous ces notations, on combine l'approximation $\hat{\eta} \simeq \eta$ avec l'équation (4.13), pour obtenir un choix naturel pour l'estimateur $\hat{\mathbf{x}}_k$:

$$\begin{aligned} \bar{\mathbf{x}}_k &= \mathbb{E}[\mathbf{x}_k|\mathbf{y}] = \mathbb{E}[\underbrace{\mathbb{E}[\mathbf{x}_k|\rho, \mathbf{y}]}_{\hat{\mathbf{x}}_k(\rho)}|\mathbf{y}] \\ &= \int_{\rho \in E} \hat{\mathbf{x}}_k(\rho) \eta(d\rho) \simeq \int_{\rho \in E} \hat{\mathbf{x}}_k(\rho) \hat{\eta}(d\rho) \\ &= \underbrace{\frac{1}{N_p} \sum_{i=1}^{N_p} \hat{\mathbf{x}}_k(\rho^{(i)})}_{=:\hat{\mathbf{x}}_k} \end{aligned}$$

En ce qui concerne l'estimateur $\hat{\Sigma}_k$ de la matrice de covariance a posteriori, toujours sous les mêmes notations et d'après l'équation (4.14), on a :

$$\Sigma_k = \underbrace{\mathbb{E}(\hat{\Sigma}_k(\rho)|\mathbf{y})}_{\Sigma_k^{(1)}} + \underbrace{\text{Var}(\hat{\mathbf{x}}_k(\rho)|\mathbf{y})}_{\Sigma_k^{(2)}}.$$

On estime alors $\Sigma_k^{(1)}$ et $\Sigma_k^{(2)}$ séparément :

1. Evaluation de $\Sigma_k^{(1)}$

$$\begin{aligned} \Sigma_k^{(1)} &= \mathbb{E}(\hat{\Sigma}_k(\rho)|\mathbf{y}) \\ &= \int_{\rho \in E} \hat{\Sigma}_k(\rho) \eta(d\rho) \simeq \int_{\rho \in E} \hat{\Sigma}_k(\rho) \hat{\eta}(d\rho) \\ &= \underbrace{\frac{1}{N_p} \sum_{i=1}^{N_p} \hat{\Sigma}_k(\rho^{(i)})}_{=:\hat{\Sigma}_k^{(1)}} \end{aligned}$$

2. Evaluation de $\Sigma_k^{(2)}$

$$\begin{aligned}
 \Sigma_k^{(2)} &= \mathbb{E} \left[(\hat{\mathbf{x}}_k(\rho) - \bar{\mathbf{x}}_k) (\hat{\mathbf{x}}_k(\rho) - \bar{\mathbf{x}}_k)^T | \mathbf{y} \right] \\
 &= \int_{\rho \in E} (\hat{\mathbf{x}}_k(\rho) - \bar{\mathbf{x}}_k) (\hat{\mathbf{x}}_k(\rho) - \bar{\mathbf{x}}_k)^T \eta(d\rho) \\
 &\simeq \int_{\rho \in E} (\hat{\mathbf{x}}_k(\rho) - \hat{\mathbf{x}}_k) (\hat{\mathbf{x}}_k(\rho) - \hat{\mathbf{x}}_k)^T \hat{\eta}(d\rho) \\
 &= \underbrace{\frac{1}{N_p} \sum_{i=1}^{N_p} (\hat{\mathbf{x}}_k(\rho^{(i)}) - \hat{\mathbf{x}}_k) (\hat{\mathbf{x}}_k(\rho^{(i)}) - \hat{\mathbf{x}}_k)^T}_{=:\hat{\Sigma}_k^{(2)}}
 \end{aligned}$$

Finalement, l'estimateur de Σ_k est défini par : $\hat{\Sigma}_k := \hat{\Sigma}_k^{(1)} + \hat{\Sigma}_k^{(2)}$.

Si l'on se focalise sur une fréquence donnée, ou sur une zone de l'objet donnée, cette méthode d'inversion fournit de précieuses informations :

- Pour toute fréquence f_k , elle fournit une approximation de l'espérance et de la matrice de covariance a posteriori des états \mathbf{x}_k . La loi a posteriori $p(\mathbf{x}_k | \mathbf{y})$ elle-même peut être simulée de façon approchée en tirant au hasard un des $\rho^{(i)}$ au sein du nuage final de particules, et en effectuant, conditionnellement à ce $\rho^{(i)}$, un lissage de Kalman pour simuler $\mathbf{x}_k | \rho^{(i)}$. (voir dans la suite l'illustration de la figure 4.11 page 105).
- Pour toute zone élémentaire de l'objet fixée, la méthode fournit des estimations de l'espérance et de la variance marginale pour chaque fréquence, de sorte que les résultats peuvent être représentés sous la forme de profils fréquentiels, donnés avec incertitudes (à l'aide des valeurs sur la diagonale des $\hat{\Sigma}_k$) (voir figure 4.12 page 106).

4.4 Applications

Dans cette section, notre approche sur la diffraction inverse est appliquée à des mesures de rétrodiffusion EM d'un objet métallique de forme ogivale. La validation de la méthode est réalisée à l'aide de données simulées, et sur une bande de fréquences de $f = 1$ à 2 GHz. La section 4.4.1 décrit le scénario de référence des tests non destructifs. Ensuite, la section 4.4.2 décrit le processus d'inversion et illustre quelques résultats centraux. Une analyse de performance détaillée est alors développée dans la section 4.4.3. Enfin, dans la section 4.4.4, nous analysons plus brièvement quelques variantes de notre approche de base.

4.4.1 Scénario de test non destructif

L'objet métallique Considérons de près l'objet métallique axisymétrique que nous avons présenté précédemment figure 4.3 page 87. Sa forme ogivale, dérivée du benchmark SER [82] est parfaitement connue. L'objet, long de 2m, est recouvert par $N_a = 5$ régions de matériaux, chacune supposée relativement homogène, au sens où les propriétés

4.4. APPLICATIONS

radioélectriques isotropes ne varient que faiblement à l'intérieur de chaque région. Pour chaque région, les "vraies" propriétés EM $\mathbf{x}_{\text{true}}(f)$ suivent le modèle suivant : $\mathbf{x}_{\text{true}}(f) = \mathbf{x}_{\text{ref}}(f) + c \cdot \Lambda(f)$.

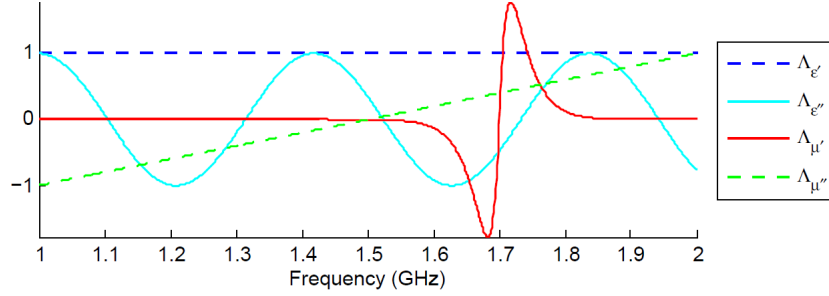


FIGURE 4.8 – Les différentes fonctions Λ

Pour chaque fréquence f , le vrai vecteur (inconnu) $\mathbf{x}_{\text{true}}(f)$ est de dimension $4N = 76$, où :

- $\mathbf{x}_{\text{ref}}(f)$ est un profil fréquentiel de référence (méta-matériau), qui dépend de la région, et de la composante radioélectrique (ϵ' , ϵ'' , μ' , μ''). Ces $4N_a = 20$ profils de référence sont choisis relativement réguliers, et conformes aux ordres de grandeur typiques de ces quantités (c'est-à-dire ≥ 0 , et grosso modo ≤ 20).
- $\Lambda(f)$ est une fonction de perturbation, dépendante de la composante radioélectrique. De cette manière, les 4 fonctions $\Lambda_{\epsilon'}$, $\Lambda_{\epsilon''}$, $\Lambda_{\mu'}$, $\Lambda_{\mu''}$ définissent la forme des perturbations. Comme on le montre figure 4.8, elles sont choisies de régularité variable, de façon à pouvoir tester les capacités de l'algorithme d'inversion dans des situations variées.
- c est un simple facteur multiplicatif, qui dépend de la région de matériaux. Pour examiner l'influence de l'amplitude de la perturbation, on choisit des valeurs de c de plus en plus grandes : $\{0.5, 1, 2, 4, 8\}$, en référence aux 5 régions successives.

Mesures de rétrodiffusion (simulées) Conformément au mode d'acquisition de SER conventionnel, décrit dans la section 4.1.1 page 85, les coefficients complexes de rétrodiffusion sont mesurés pour les deux polarisations HH et VV, aux $K_f = 20$ fréquences régulièrement espacées ($f_1 = 1$ GHz, \dots , $f_{K_f} = 2$ GHz) et selon les $K_\theta = 23$ angles d'incidence régulièrement espacés ($\theta_1 = 0^\circ$, \dots , $\theta_{K_\theta} = 180^\circ$).

Le vecteur des observations $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_{K_f})$ est simulé selon le modèle de vraisemblance (4.2) page 90. Il s'agit donc concrètement d'appliquer le solveur $\mathcal{F}_{\text{Maxwell}}$ et de lui rajouter un bruit blanc gaussien d'écart type $\sigma_n = 10^{-3}$ ($\sim 1\%$). Chacun des 20 vecteurs d'observation \mathbf{y}_k est de dimension $4 \times K_\theta = 92$. L'ensemble des données est représenté figure 4.9. On y remarque en particulier une réflexion spéculaire de très haute amplitude lorsque l'objet ogival est tourné perpendiculairement à la direction de l'onde de propagation.

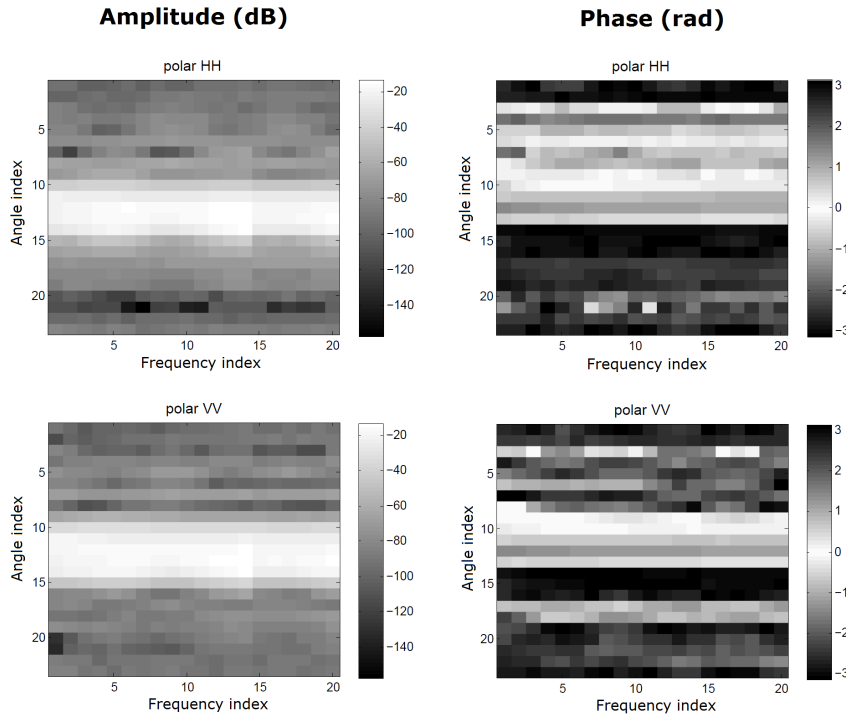


FIGURE 4.9 – Hologramme des observations, amplitude et phase. (polarisations HH et VV)

4.4.2 Processus d'inversion

L'objectif est d'estimer les propriétés radioélectriques, soit la quantité \mathbf{x}_{true} qui est une fonction de la fréquence f , à partir des données de rétrodiffusion. Dans cette section, nous donnons des détails d'implémentation et le contexte de notre application.

Espace d'état La dimension de l'espace d'état dépend du nombre de fréquences et de la finesse de la discrétisation de l'objet en zones élémentaires. Pour la limiter, on se restreint à une division en 19 zones.

Information a priori L'information a priori (voir section 4.2) doit être détaillée dans ce contexte d'étude. En ce qui concerne l'information spatiale, c'est-à-dire les $p(\mathbf{x}_k)$, les valeurs moyennes \mathbf{m}_k sont données, pour tout k , par les profils fréquentiels de référence $\mathbf{x}_{\text{ref}}(f_k)$. Les incertitudes autour de ces valeurs sont données par les matrices de covariance \mathbf{P}_k de (4.9) page 92 avec : $\rho_S = 0.95$ et $\sigma_k(i) = 1 + 0.15 \times \mathbf{m}_k(i)$ pour chaque zone élémentaire i . Autrement dit, on suppose un écart type minimum de 1 qui croît proportionnellement à l'amplitude des valeurs de référence. En ce qui concerne l'information a priori de type fréquentielle, on suppose que ρ dépend à la fois de la composante EM $\epsilon', \epsilon'', \mu', \mu''$ et de la région de matériaux, de sorte qu'il est de dimension 20. Quant à sa loi a priori $p(\rho)$, on

4.4. APPLICATIONS

pose :

$$p(\rho) = \prod_{i=1}^{20} p(\rho_i)$$

où toutes les loi marginales $p(\rho_i)$ sont identiques entre elles, leur densité commune étant représentée figure 4.10. Remarquons au passage que cette loi $p(\rho)$ est facile à simuler, car il suffit de simuler chaque composante ρ_i indépendamment, à l'aide par exemple d'une méthode d'acceptation/rejet.

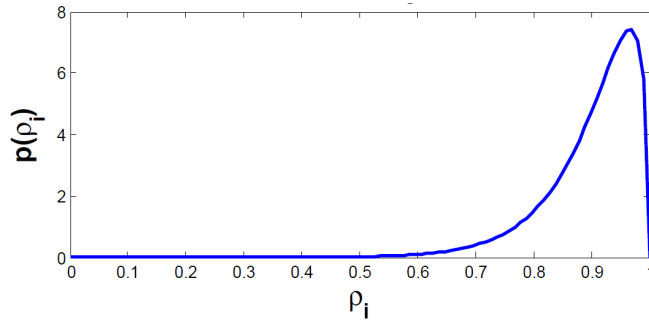


FIGURE 4.10 – Densité de la loi marginale $p(\rho_i)$

Modèle de vraisemblance Le modèle de vraisemblance de substitution (4.7) page 91 a déjà été établi par apprentissage, ce qui signifie que les matrices \mathbf{A}_k et les vecteurs \mathbf{y}_k^0 sont connus (voir figure 4.6), tout comme l'écart type marginal σ_n qui est choisi conformément au bruit de mesure du modèle de simulation.

Paramétrage du SMC Dans notre étude standard, le chemin de mesures de probabilité η_n est défini selon le schéma de type recuit (voir section 4.3.2 page 96). Pour assurer un comportement stable de l'algorithme SMC (c'est-à-dire conserver une bonne qualité d'approximation $\eta_n^{N_p} \simeq \eta_n$ jusqu'à la fin), on opte pour la stratégie adaptative suivante, qui permet en outre de limiter le nombre de particules à $N_p = 100$:

- *étape de sélection* : comme nous l'avons mentionné, l'incrément $\Delta\alpha_n = \alpha_n - \alpha_{n-1}$ contrôle le degré de sélectivité. Si $\Delta\alpha_n$ est trop petit, toutes les particules se voient attribuer sensiblement le même poids, et aucune sélection ne s'opère parmi elles. Si au contraire $\Delta\alpha_n$ est choisi trop grand, une infime minorité de particules dominant toutes les autres, et la majorité des particules se trouve donc supprimée. Ce phénomène menace très sérieusement la performance de l'algorithme SMC car le nuage de particules, perdant toute sa diversité, perd sa capacité à explorer l'espace d'état. Comme il est très difficile de savoir à l'avance quel incrément $\Delta\alpha_n$ choisir, on ne choisit pas cette suite à l'avance, mais à l'intérieur même du processus, de façon adaptative, c'est-à-dire directement de façon à assurer une sélection d'environ 25% de la population de particules. Cette méthode adaptative, garante d'une sélection raisonnable, est le sujet d'étude de la section 2.4 in-

troduite page 42. On y trouve notamment une description détaillée section 2.4.2 page 44 et une analyse théorique sections 2.4.3 et 2.4.4.

- *étape de mutation* : l'étape de mutation est cruciale dans la mesure où elle permet aux particules d'explorer l'espace d'état E . On utilise dans notre application des noyaux de Markov M_n définis comme la composition de plusieurs noyaux de Metropolis-Hastings $K_n^{(i)}$, dont les noyaux de proposition $Q_n^{(i)}(x, dy)$ sont uniformes et centrés en x . Notons $\sigma_{\text{prop},n}^{(i)}$ la taille de leur fenêtre de proposition. Pour que les particules bougent dans un voisinage de taille convenable (soit assez grand pour explorer E et assez petit pour converger), la suite $(\sigma_{\text{prop},n}^{(i)})_i$ est toujours initialisée sur une grande valeur, puis décroît géométriquement. Ce processus est arrêté lorsqu'au moins 75% des particules ont bougé au moins une fois depuis le début.

Résultats Dans le contexte de cette étude de référence, le processus d'inversion nécessite environ 30 minutes pour un processeur standard. Il faut noter que plus la dimension de l'espace est grande, plus l'inversion est longue. Sur la figure 4.11, nous présentons les estimations de μ' pour toutes les zones de l'objet, et leurs incertitudes associées, comparées aux vraies valeurs, et cela à la fréquence fixe $f_{16} = 1.8$ GHz. Il est à noter que dans notre exemple, les écarts entre vraies valeurs et valeurs a priori sont très importants (voir figure 4.8). Comme nous l'expliquions précédemment, il est possible de simuler approximativement la distribution a posteriori $p(\mathbf{x}_{16}|\mathbf{y})$ pour déterminer les incertitudes de nos estimateurs. Il est clair que les propriétés radioélectriques EM sont correctement estimées sur tout le long de l'objet et de ses 5 régions, et les "zones probables" données par les incertitudes recouvrent toujours grosso modo les vrais profils.

La figure 4.12 représente les profils fréquentiels pour une zone élémentaire donnée (en l'occurrence la 18^{ème}). Toutes les composantes ($\epsilon', \epsilon'', \mu', \mu''$) sont représentées. Chacune d'entre elles est estimée de façon correcte. Il est remarquable que les résultats soient bons y compris lorsque les perturbations (c'est-à-dire la différence entre les valeurs réelles et les valeurs a priori) sont larges et irrégulières. On doit ce comportement robuste notamment à l'estimation adaptative des composantes de ρ . Dans la suite, nous confirmons ces résultats par plusieurs analyses.

4.4.3 Analyse de performance

Pour préciser et étendre ces résultats sommaires, nous proposons dans cette section une analyse statistique des performances du processus d'inversion. Elle est réalisée dans le même contexte que la section 4.4.1. Comme ce processus est en partie stochastique (algorithme SMC), il convient de distinguer deux aspects différents. Premièrement, l'étude de la variance de nos estimateurs $\hat{\mathbf{x}}_k$ and $\hat{\Sigma}_k$ obtenus à partir d'un jeu de données unique \mathbf{y} fixé. Cette variance est celle du processus d'inversion lui-même ; elle est due aux opérations aléatoires réalisées dans le SMC. Deuxièmement, l'étude de la variance de notre méthode sur plusieurs jeux de données $\mathbf{y}^{(i)}$, c'est-à-dire l'estimation de sa précision en moyenne.

4.4. APPLICATIONS

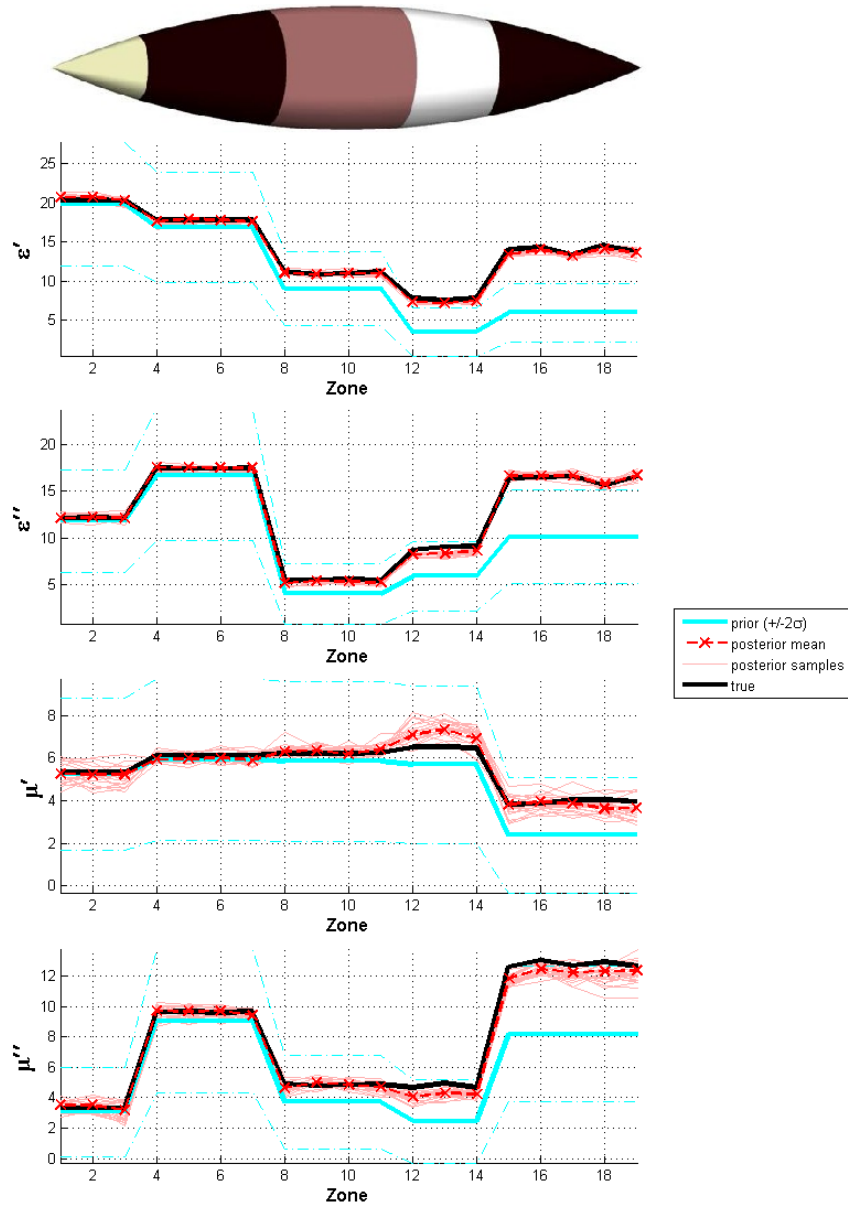
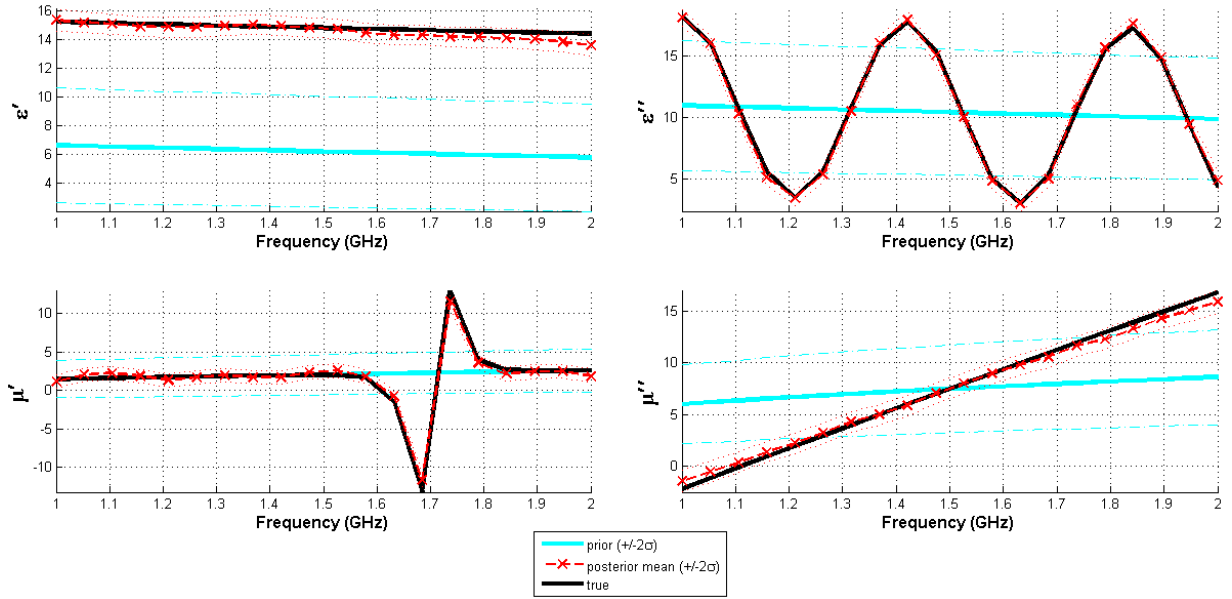


FIGURE 4.11 – Propriétés EM estimées, à la fréquence $f = 1.8$ GHz

FIGURE 4.12 – Estimation des propriétés EM pour la 18^{ème} zone élémentaire

4.4.3.1 Variation stochastique

Pour un jeu de données \mathbf{y} donné, notre méthode fournit essentiellement deux suites d'estimateurs. Les estimateurs de l'espérance a posteriori $(\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_{K_f})$, et les estimateurs des matrices de covariance a posteriori $(\hat{\Sigma}_1, \dots, \hat{\Sigma}_{K_f})$. Comme pour tous les algorithmes stochastiques, on se doit de vérifier que malgré l'aléa généré volontairement dans les calculs, le résultat final fourni est bien toujours le même, ou au moins que sa variance propre est négligeable.

Soit $\hat{\mathbf{x}}$ la concaténation des vecteurs $\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_{K_f}$. Soit $\hat{\sigma}$ la concaténation des incertitudes marginales estimées (racine carrée des valeurs diagonales des $\hat{\Sigma}_k$). Définis de cette façon, $\hat{\mathbf{x}}$ et $\hat{\sigma}$ peuvent être considérés comme deux matrices de taille 76×20 , soit les deux estimateurs essentiels de notre méthode.

Pour quantifier la variance stochastique, on simule un jeu de données \mathbf{y} , et on réalise la méthode d'inversion 30 fois. A l'issue de ces calculs, on récupère donc 30 paires d'estimateurs $\{(\hat{\mathbf{x}}^{(1)}, \hat{\sigma}^{(1)}), \dots, (\hat{\mathbf{x}}^{(30)}, \hat{\sigma}^{(30)})\}$. Pour toute paire d'indices $(i, k) \in \{1, \dots, 76\} \times \{1, \dots, 20\}$, on considère les valeurs moyennes de ces estimateurs et leur valeurs RMS (pour Root Mean Square) :

$$\bar{\hat{\mathbf{x}}}(i, k) := \frac{1}{30} \sum_{r=1}^{30} \hat{\mathbf{x}}^{(r)}(i, k) \quad \text{et} \quad \bar{\hat{\sigma}}(i, k) := \frac{1}{30} \sum_{r=1}^{30} \hat{\sigma}^{(r)}(i, k)$$

4.4. APPLICATIONS

$$\text{RMS}(\hat{\mathbf{x}})(i, k) := \left(\frac{1}{30} \sum_{r=1}^{30} \left(\hat{\mathbf{x}}^{(r)}(i, k) - \bar{\mathbf{x}}(i, k) \right)^2 \right)^{1/2}$$

$$\text{RMS}(\hat{\sigma})(i, k) := \left(\frac{1}{30} \sum_{r=1}^{30} \left(\hat{\sigma}^{(r)}(i, k) - \bar{\sigma}(i, k) \right)^2 \right)^{1/2}$$

Les résultats numériques, obtenus sur toutes les paires d'indices (i, k) , sont résumés dans le tableau 4.1. Deux points méritent d'être soulignés. Tout d'abord, l'écart type des $\hat{\mathbf{x}}^{(r)}$ est très petit de façon absolue ($\simeq 10^{-2}$). De plus, il est négligeable comparé aux variances estimées de nos estimateurs (au moins une décade). Ensuite, l'écart type des $\hat{\sigma}^{(r)}$ est encore plus petit ($\simeq 10^{-3}$) et négligeable comparé aux valeurs de $\hat{\sigma}^{(r)}$ elles-mêmes. (au moins deux décades). Par conséquent, il existe une variance stochastique, mais elle est largement négligeable par rapport aux incertitudes inhérentes au problème inverse, incertitudes de mesure comprises.

mean RMS($\hat{\mathbf{x}}$)	max RMS($\hat{\mathbf{x}}$)	mean $\frac{\text{RMS}(\hat{\mathbf{x}})}{\bar{\sigma}}$	max $\frac{\text{RMS}(\hat{\mathbf{x}})}{\bar{\sigma}}$
4.16 10^{-3}	4.11 10^{-2}	1.08 10^{-2}	9.87 10^{-2}
mean RMS($\hat{\sigma}$)	max RMS($\hat{\sigma}$)	mean $\frac{\text{RMS}(\hat{\sigma})}{\bar{\sigma}}$	max $\frac{\text{RMS}(\hat{\sigma})}{\bar{\sigma}}$
1.10 10^{-3}	7.56 10^{-3}	2.99 10^{-3}	1.84 10^{-2}

TABLE 4.1 – Résultats des valeurs RMS de $\hat{\mathbf{x}}$ et $\hat{\sigma}$

4.4.3.2 Précision moyenne

La précision moyenne doit, par définition, être étudiée via plusieurs cas. A ce titre, on simule 30 jeux de données indépendants $\{\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(30)}\}$. Pour chacun de ces vecteurs d'observation $\mathbf{y}^{(r)}$, l'algorithme d'inversion calcule la paire d'estimateurs $(\hat{\mathbf{x}}^{(r)}, \hat{\sigma}^{(r)})$. La comparaison de ces estimateurs avec les vraies valeurs de \mathbf{x} est quantifiée par le RMSE (pour Root Mean Square Error) :

$$\text{RMSE}(i, k) := \left(\frac{1}{30} \sum_{r=1}^{30} \left(\hat{\mathbf{x}}^{(r)}(i, k) - \mathbf{x}_{\text{true}}(i, k) \right)^2 \right)^{1/2}$$

Ces termes d'erreur sont présentés figure 4.13, et comparés aux erreurs estimées $\bar{\sigma}$. On tire de ces résultats les conclusions suivantes. Malgré l'ampleur et l'irrégularité des perturbations, très éloignées du modèle a priori, les estimateurs $\hat{\mathbf{x}}^{(r)}$ fournissent une bonne approximation de \mathbf{x}_{true} (on note en particulier que mean RMSE = 3.60 10^{-1}). En outre, les valeurs RMSE sont relativement comparables aux incertitudes marginales données par $\bar{\sigma}$, ce qui prouve que les variances a posteriori estimées sont crédibles.

Dans ce processus d'inversion, le rôle de l'estimation de ρ est particulièrement intéressant. De façon schématique, elle donne en avance la forme des profils fréquentiels recherchés, en estimant leurs régularités respectives. Sur la figure 4.14 sont représentés les résultats pour

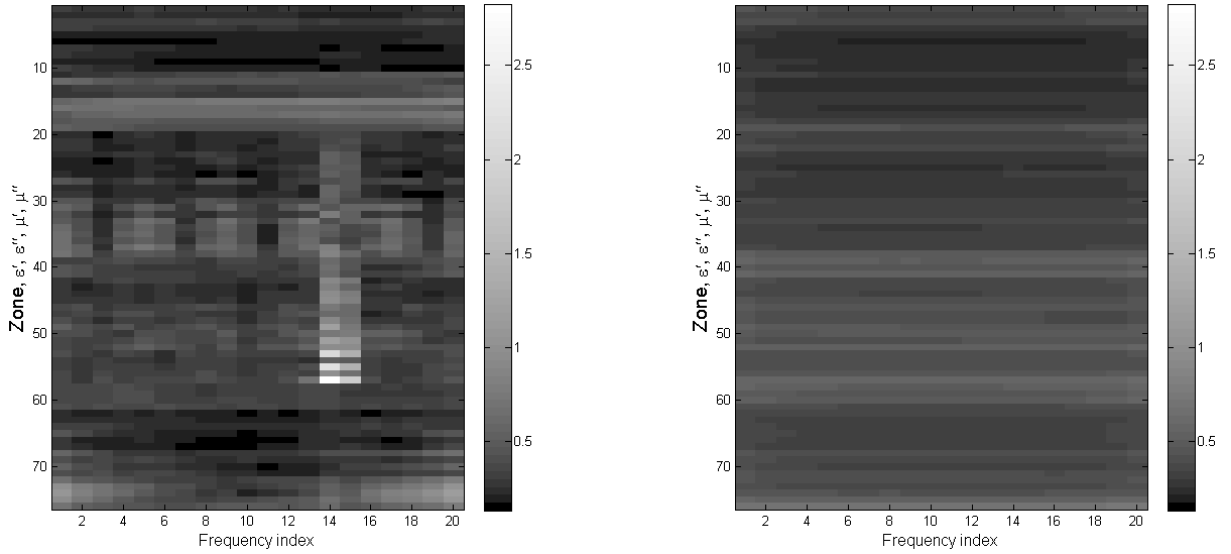


FIGURE 4.13 – RMSE (à gauche) et estimation des incertitudes marginales (à droite)

$(\hat{\mathbf{x}}^{(1)}, \hat{\sigma}^{(1)})$ en ce qui concerne les zones numéro 2, 5, 9, 13 et 17 et la perméabilité ε'' . Sur la partie droite, les histogrammes représentent la distribution a posteriori de ρ . Comme voulu, la difficulté croît de la zone 2 jusqu'à la zone 17. Cela est dû à notre choix en amont d'une perturbation de plus en plus large et irrégulière. Sur la partie droite de la figure sont représentés les histogrammes associés aux particules $(\rho^{(1)}, \dots, \rho^{(100)})$ (plus précisément, pas les particules en entier mais uniquement la composante concernée). On constate alors clairement que plus le vrai signal est irrégulier, plus les $\rho^{(i)}$ sont petits, ce qui est tout à fait cohérent étant donné que ρ quantifie la corrélation fréquentielle. Parallèlement à ce phénomène, on vérifie sur la colonne centrale de la figure que malgré la difficulté croissante, la moyenne des RMSE reste stable. Il ressort donc que le comportement adaptatif de l'estimation de ρ est capital pour la robustesse de l'algorithme.

4.4.4 Analyses de variantes

On se propose maintenant d'analyser plus brièvement l'influence d'autres paramètres, contextuels ou inhérents au processus d'inversion.

4.4.4.1 Influence des paramètres du traitement

Le processus d'inversion décrit dans la section 4.3 admet plusieurs degrés de liberté, qualitatifs et quantitatifs, notamment au niveau du SMC. On propose ici quelques remarques empiriques sur certains d'entre eux.

Le nombre de particules N_p A l'instar des méthodes d'échantillonnage i.i.d. classiques, la précision de l'algorithme SMC est en $N_p^{-1/2}$ (voir par exemple la borne L^p donnée

4.4. APPLICATIONS

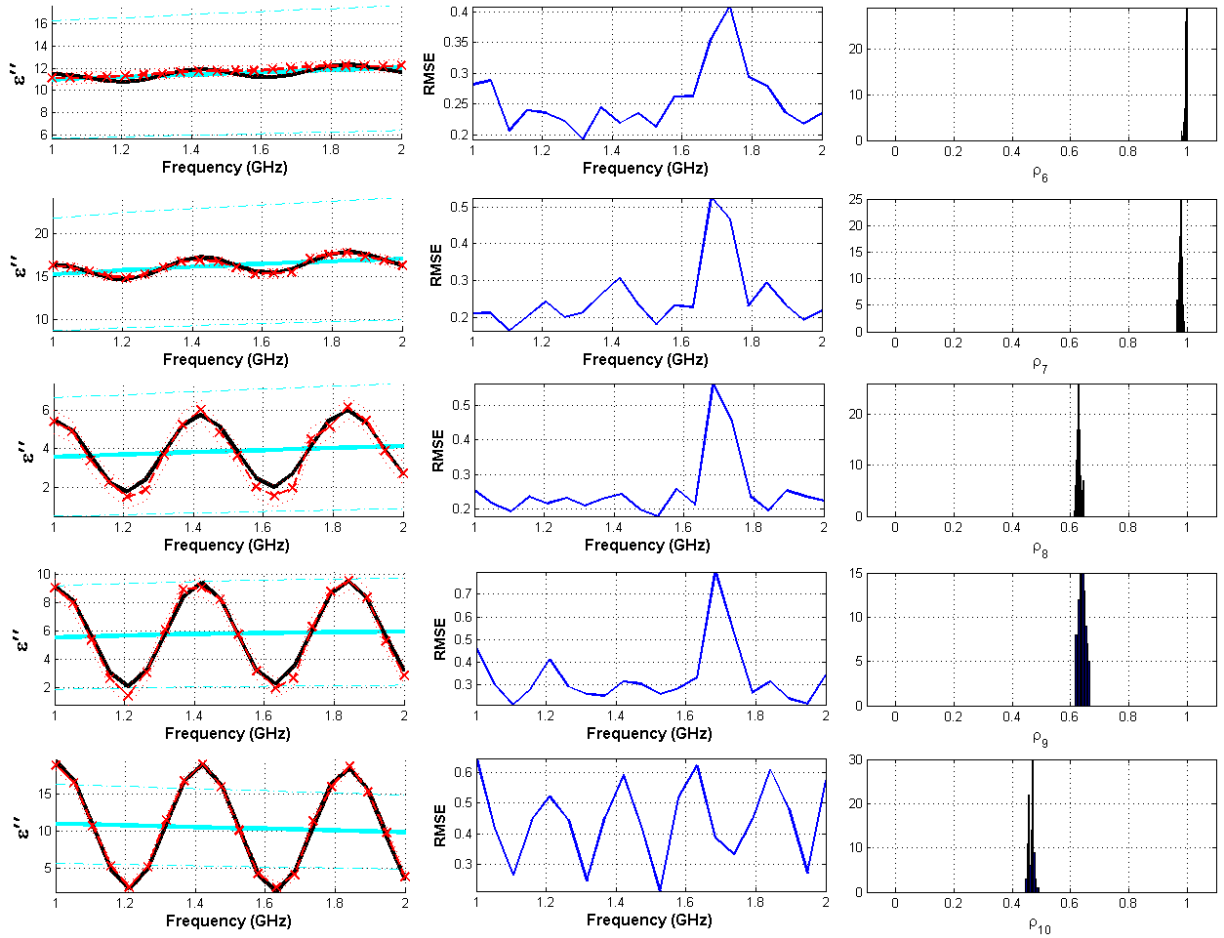


FIGURE 4.14 – Estimations de ε'' pour les zones 2, 5, 9, 13 et 17

en (2.13) page 26 ou [28] pour un état de l'art plus complet). Cependant, dans notre problème, l'objectif principal n'est pas directement d'avoir une estimation précise de η , mais d'estimer \mathbf{x} . Comme l'impact d'une variation locale de ρ sur \mathbf{x} est relativement petite, le point crucial est que le nuage de particules, pris globalement, atteigne la région correcte dans l'ensemble E . De ce point de vue, la condition importante est la stabilité du flot de Feynman-Kac (voir [28]), qui assure le fait que les particules ne se perdent pas dans E . C'est précisément l'objet des stratégies adaptatives à l'intérieur des étapes de sélection et mutation. Par conséquent, il semble inutile, et coûteux en temps de calcul, d'utiliser un grand nombre de particules. On note néanmoins qu'au dessous de $N_p \simeq 40$, les particules risquent de se voir coincées dans des modes locaux.

Le chemin d'interpolation η_n En plus du schéma de type recuit, nous avons testé le schéma hybride. Il est conçu pour assimiler les observations une par une, tout en rafraîchissant les estimateurs progressivement. De plus, il gère les problèmes de sélectivité propres au schéma séquentiel sur les mesures. Les résultats sont bons, sensiblement identiques à

ceux obtenus avec le schéma du recuit. Cependant, l'algorithme SMC nécessite environ 4 fois plus de temps de calcul.

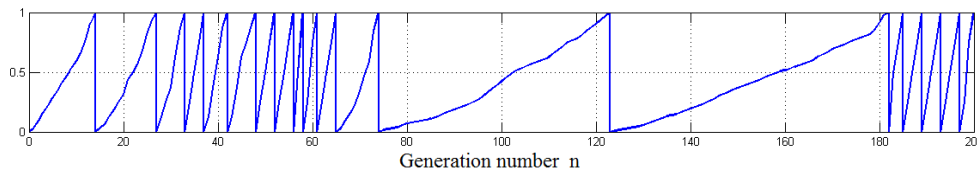


FIGURE 4.15 – Le paramètre de recuit $\alpha_i^{(k)}$

La figure 4.15 permet d'interpréter aisément ce constat. Il apparaît que beaucoup d'observations n'apportent aucune information nouvelle, de sorte que la suite de recuit $\alpha_i^{(k)}$ associée prend immédiatement la valeur 1. À l'inverse, lorsqu'une nouvelle observation fournit une information en contradiction avec les précédentes, les particules doivent migrer d'une région à une autre dans l'ensemble E , ce qui nécessite de nombreuses mutations, donc plus de temps de calcul.

Le paramètre ρ La distribution a priori $p(\rho)$ de la figure 4.10 a un impact très faible sur l'estimation finale de η . Eu égard à son interprétation de régularité fréquentielle, elle est choisie de façon à pénaliser les petites valeurs pour favoriser les profils réguliers. Mais en pratique, ce terme de pénalisation $p(\rho)$ est moins déterminant que le terme de vraisemblance $p(\mathbf{y}|\rho)$. Par ailleurs, ρ peut être défini comme étant de dimension 5. Dans ce cas, le SMC est plus rapide. Cependant, une telle définition de ρ présuppose que les corrélations fréquentielles sont les mêmes pour ϵ' , ϵ'' , μ' , μ'' , hypothèse peu vérifiée en pratique.

4.4.4.2 Influence du contexte

Comme nous l'avons mentionné, la méthode est très robuste en ce qui concerne l'amplitude et l'irrégularité de la perturbation (déviations des profils de référence).

Bruit de mesure Cependant, elle est très sensible à l'amplitude du bruit de mesure. Sa performance se dégrade rapidement lorsque ce bruit est trop important. Il s'agit en fait d'un problème plus profond d'observabilité. Numériquement, il peut s'expliquer en revenant au métamodèle : de fait, les matrices \mathbf{A}_k sont mal conditionnées. En particulier, les composantes μ' et ϵ'' sont fortement corrélées, et il en est de même pour μ'' et ϵ' .

Sur la figure 4.16 sont représentées les estimations de ces deux quantités dans le cas d'un bruit de mesure d'amplitude $\sigma_n = 10^{-2}$. On s'aperçoit que les perturbations de μ'_{true} et ϵ''_{true} sont correctement détectées par le processus d'inversion, mais mal distribuées entre μ' et ϵ'' .

Problème de la dimension Le temps de calcul est très sensible à la dimension du problème. La raison est simple : chaque évaluation élémentaire implique (entre autres) un filtrage de Kalman (voire un lissage), soit K_f inversions de matrices de taille $4N \times 4N$.

4.4. APPLICATIONS

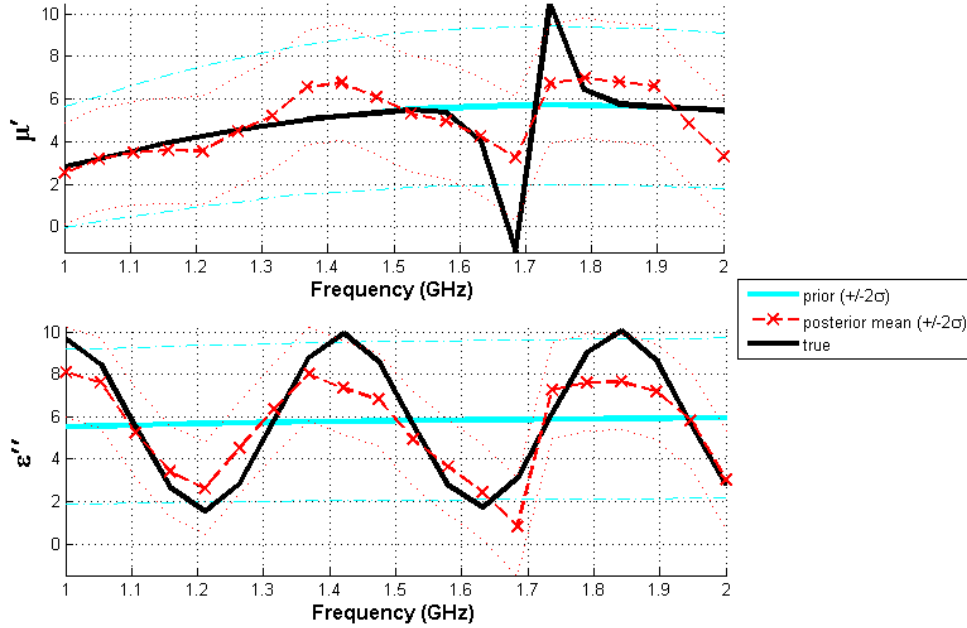


FIGURE 4.16 – Estimation de μ' et ϵ'' pour la 14^{ème} zone élémentaire, $\sigma_n = 10^{-2}$

Par conséquent, le traitement d'un problème de grande dimension nécessite sans doute une parallélisation adéquate de l'algorithme SMC, et la distribution des lisseurs de Kalman via du calcul haute performance. En outre, l'accélération de la partie filtres de Kalman en grande dimension peut être envisagée via des méthodes approchées de type Kalman d'ensemble (voir [44, 59]).

Conclusion du chapitre

Ce chapitre illustre un cheminement de pensée statistique complet mis au service de la résolution d'un problème inverse. Tout d'abord dans la modélisation du problème direct, avec l'interprétation aléatoire classique du bruit de mesure et surtout la construction d'un métamodèle obtenu par apprentissage statistique. Ensuite dans l'élaboration d'un modèle a priori sur l'inconnu, au moyen de l'imbrication de deux structures séquentielles de type processus auto-régressif multidimensionnel, l'une spatiale, l'autre fréquentielle, et d'une hiérarchie bayésienne pour le paramètre multidimensionnel de régularité fréquentielle. Enfin, dans l'inversion à proprement parler, mêlant estimation stochastique robuste via le SMC et calcul analytique via le lissage de Kalman. Sous réserve de la validité de la modélisation EM et du modèle simplifié qui en est tiré, cette approche, dite Rao-Blackwellisée, tire profit de la structure particulière du problème, et fournit en plus des estimateurs des quantités recherchées une estimation des incertitudes, ce grâce à la formulation bayésienne. Une fois de plus, au sein du SMC, les méthodes adaptatives sur la sélection et la mutation font figure de méthodologie simple et efficace pour garantir des performances robustes,

malgré, il est vrai, un coût important en temps de calcul. Le SMC lui-même, charnière du processus d'inversion, donne, par son adaptabilité, toute sa souplesse à la méthode, lui permettant de réagir intelligemment dans les cas d'estimation difficiles ou improbables.

Du point de vue problème inverse mal posé, les résultats sont satisfaisants : à partir de mesures sur le comportement électromagnétique global de l'objet, on a réussi à inférer des informations sur les propriétés EM locales, que cela soit en fréquence ou en espace. Comme pour tout problème inverse mal posé, ce résultat repose notamment sur l'introduction d'information a priori. Lorsque cette information est erronée, les résultats obtenus montrent la capacité d'adaptation de l'approche, qui peut en outre quantifier ses doutes et incertitudes quant aux résultats qu'elle fournit.

Malgré tout, les limites de ce travail ont été identifiées. D'abord, la nécessité d'une bonne connaissance du problème : connaissance précise de la géométrie de l'objet pour pouvoir utiliser le code EM $\mathcal{F}_{\text{Maxwell}}$, connaissance d'un domaine a priori sur la solution, pour la construction et la validation du métamodèle, caractérisation des incertitudes de mesure, et plus généralement connaissance a priori sur les propriétés radioélectriques (homogénéité spatiale, moyennes a priori, variances a priori, etc.). Ensuite, le problème du temps de calcul. Etant donnée l'utilisation massive des filtres de Kalman, ce temps de calcul est extrêmement sensible à la dimension du problème, c'est-à-dire au nombre de zones élémentaires, au nombre d'angles d'observation et au nombre de fréquences. Une application en dimension sensiblement supérieure à celle de nos tests demande sans doute une parallélisation de l'algorithme SMC. Enfin, le problème de l'observabilité, qui se traduit par de mauvaises performances dès que le bruit de mesure devient trop important.

Conclusion du manuscrit et perspectives

Ce document rend compte d'un travail à la fois théorique et appliqué autour de l'algorithme SMC. La partie théorique se plonge dans l'analyse des semigroupes de Feynman-Kac dans le but d'obtenir des estimées non asymptotiques uniformes en temps. Bien qu'à ce titre elle soit formalisée dans le cadre le plus général, l'approche a été pensée pour s'appliquer aux modèles à schéma de température (modèles ISA ou variantes), ou plus généralement aux suites de mesures du type

$$\eta_n(dx) \propto H_n(x)dx = \left(\prod_{k=1}^n G_k(x) \right) dx$$

où les G_k sont les potentiels de sélection, et où les noyaux de Markov sont choisis de façon à laisser stables ces mesures et explorer un ensemble E .

Dans la mesure où nous avons utilisé précisément ces méthodes dans les parties appliquées pour résoudre des problèmes inverses concrets, on est en situation d'analyser le lien entre théorie et application avec un regard critique. En réalité, malgré d'immenses progrès depuis 20 ans, il réside encore un fossé important entre les deux, et les ingénieurs, physiciens, biologistes, etc, qui utilisent ces méthodes manquent encore certainement d'outils théoriques rigoureux pour guider son utilisation. Ainsi trouve-t-on souvent dans les articles appliqués une référence théorique pour la description générale et formelle de l'ISA ou du SMC, mais très rarement pour des précisions sur leurs réglages et la multitude de choix qu'ils requièrent dans leur mise en oeuvre informatique, qui représentent pourtant l'essentiel du travail : échelles d'espace, étude préliminaire de propagation d'incertitudes, construction de noyaux de proposition adaptés à la géométrie de l'espace, élaboration d'un chemin de mesures judicieux, construction de stratégies adaptatives de sélection, de stratégies adaptatives de mutation, choix de paramètre via essai préliminaire, problèmes d'initialisation, choix de la taille de la population, etc. Ces sujets sont essentiellement abordés de façon empirique. Je profite de cette conclusion pour essayer d'analyser ce lien théorie-application par le biais de l'approche que nous avons développée dans cette thèse et de dégager des pistes de réflexion.

Les résultats les plus importants de notre travail théorique, comme les théorèmes 2 ou 3,

forment à la fois un constat intellectuel fort, et un aveu d'impuissance. Constat intellectuel fort car ils démontrent rigoureusement et dans un cadre extrêmement général la convergence des méthodes SMC à schéma de température (pour l'optimisation ou l'échantillonnage) en population finie, c'est-à-dire tels qu'ils sont implémentés, leur donnant ainsi une légitimité indiscutable. Aveu d'impuissance car les schémas d'itération (respectivement de température) exponentiels (respectivement logarithmiques) qu'ils exigent les déconnectent de toute application pratique.

Dans la littérature sur le sujet, l'hypothèse " E compact" est souvent critiquée, mais elle me paraît au contraire tout à fait raisonnable. D'un point de vue théorique, d'abord, au sens où toute mesure de probabilité sur \mathbb{R}^d a sa masse (à ε près) comprise dans un compact. D'un point de vue pratique, ensuite, au sens où une exploration aveugle d'un espace non borné en temps et population finis me semble être en tant que telle une entreprise absurde. Que cela soit par des bornes, l'introduction d'un a priori ou l'utilisation implicite d'une échelle spatiale typique, la définition d'un domaine d'exploration pertinent me semble incontournable.

La limite de cette approche réside probablement plutôt dans sa volonté d'être générale. Elle propose un formalisme capable de regrouper toutes les méthodes SMC, mais cette force d'universalité est aussi une faiblesse : en s'interdisant toute hypothèse spécifique, les résultats sont plus faibles. On s'en convainc par deux remarques simples. Tout d'abord, ce formalisme n'accorde presque aucune place aux échelles spatiales, qu'elles soient propres par exemple aux variations des G_n ou de V , ou bien encore aux noyaux de proposition. Or c'est une problématique d'importance majeure en pratique. A ce titre, notre application sur le matériau multicouche de la section 3.2 est plus que parlant : évoluant au départ dans un domaine de taille $[0, 5]^8$, les particules finissent toutes dans un périmètre de taille $[x_{\text{cible}} \pm 10^{-2} \text{ ou } 10^{-3}]^8$, et donc de mesure de Lebesgue $\simeq 10^{-16}$. Il est inconcevable d'arriver à un tel résultat avec des noyaux de proposition d'écart type $\simeq 1$ pour la raison simple que la probabilité qu'une particule atteigne jamais cette zone cible est infime. Cette convergence n'est possible que grâce à une adaptation progressive de l'échelle (couplée, certes, au maintien des objectifs d'exploration), qui amène le nuage de particules à se rétracter petit à petit, tout en désertant un à un les modes locaux, jusqu'à la zone cible. Cette logique est éclipsée dans la théorie, qui ne s'exprime que vis-à-vis du mélange, c'est-à-dire de l'exploration. Cette caractéristique peut être éclairée par la deuxième remarque : l'approche théorique met à égalité deux difficultés pourtant très différentes en population finie :

- η_{n+1} charge les mêmes zones que η_n , mais avec des pondérations différentes ;
- η_{n+1} charge une nouvelle zone A , qui vérifiait précédemment $\eta_n(A) \simeq 0$.

Du point de vue de notre analyse, que le nouveau potentiel G_{n+1} charge A ou toute autre zone ne change pas la quantité $g_{n+1} = \sup_{x,y} G_{n+1}(x)/G_{n+1}(y)$. Pourtant, algorithmiquement, le premier cas est relativement inoffensif car géré en grande partie par la logique de vases communicants de l'étape de sélection, alors que le second est autrement problématique : aucune particule ne se trouve en A ; elles doivent donc y migrer à force de mutations, éventuellement en traversant des barrières de potentiel. Au sein des démonstrations ou plus simplement des définitions centrales, on comprend que notre approche considère toujours

4.4. APPLICATIONS

le pire cas : la zone A la plus lointaine/improbable/inaccessible (coefficient de Dobrushin) et la barrière de potentiel la plus critique ($osc(V)$). De cette façon, on impose donc à l'algorithme d'être capable d'explorer tout l'espace en permanence. Il me semble que c'est précisément cette exigence (présente également dans l'analyse du recuit simulé simple) qui conduit aux schémas d'itération exponentiels et de température logarithmiques.

L'origine de cette limite est la définition même de la norme en variation totale, qui ne fait pas la différence entre les deux situations décrites ci-dessus, et ne donne aucun sens à la proximité euclidienne entre deux points de E . Ainsi, si les estimées proposées dans ce document sont sûrement perfectibles, je ne pense pas qu'elles le soient de façon très spectaculaire. De meilleures estimées, en tout cas des estimées plus proches de la pratique du SMC, issues d'une théorie capable d'expliquer pourquoi un algorithme SMC marche ou ne marche pas, doivent sans doute passer par des considérations géométriques sur les mesures η_n . Cela peut se faire selon le même schéma général que cette analyse, basé sur la stabilité du flot de Feynman-Kac, mais à la condition que cette stabilité soit étudiée selon un autre système de distance dans l'espace des mesures.

Enfin, ce manuscrit est l'occasion d'une réflexion sur les méthodes de sélection adaptative. Au regard de nos résultats, comme de ceux de [38], il me semble que théorie et pratique se trouvent encore aujourd'hui en décalage sur ce sujet. D'un point de vue pratique, l'adaptation est une garantie de stabilité, car elle permet au processus de définir lui-même sa suite de mesures cibles selon sa capacité d'adaptation. Or, dans la théorie, les mesures cibles sont encore formalisées de façon déterministe par une suite η_n , qui certes vérifie de bonnes propriétés, mais qui n'est pas observable à partir des données du problème. L'introduction de cette suite, pivot techniquement nécessaire à l'analyse, est peut-être aussi son point faible : on demande au processus empirique η_n^N d'approcher la suite η_n en étant stabilisé par un flot ϕ_n^N qui lui-même est une approximation du flot ϕ_n . L'adaptation est donc conçue comme une perturbation supplémentaire, et dégrade les estimées. En outre, les lois cibles η_n n'ont pas d'existence en pratique. Pour le cas de la simulation de mesures de Gibbs, par exemple, l'approximation que l'on fait implicitement dans les applications n'est pas $\eta_n^N \simeq \eta_n$ mais clairement $\eta_n^N \simeq \mu_{\beta_n^N}$. Et sans doute à juste titre ! car les noyaux utilisés étant associés aux paramètres de température empiriques β_n^N , c'est bien les $\mu_{\beta_n^N}$ qu'ils laissent stables.

Il me semble donc qu'une réconciliation d'une part entre la théorie et la pratique (recentrage sur les quantités tangibles), et d'autre part entre la théorie et la théorie (amélioration des estimées) passe par l'abandon de cette suite théorique η_n , et l'analyse de la distance entre les deux mesures aléatoires η_n^N et $\mu_{\beta_n^N}$. Techniquement, ce point de vue rajoute de nombreuses difficultés. Il considère en effet une suite de Feynman-Kac aléatoire, et remplace donc les considérations déterministes sur sa stabilité (avec typiquement des quantités comme b_p , g_p , etc) par des raisonnements conditionnels par rapport au passé du processus. L'absence d'un résultat de type (2.51) page 54 est un exemple immédiat du type de difficulté technique rencontrée dans cette situation. Mais remarquons qu'il en élimine aussi certaines, et notamment le problème déjà cité du noyau de Markov laissant stable η_n , in-

accessible en pratique. Si la mesure cible est la mesure aléatoire $\mu_{\beta_n^N}$ directement associée au paramètre de température aléatoire β_n^N , ce problème n'existe plus.

Bien que notre schéma d'analyse récursif ne s'adapte pas directement à cette situation, j'espère que les outils techniques et méthodologiques qu'il met en lumière seront utiles à de tels développements ultérieurs.

Bibliographie

- [1] IEEE Recommended Practice for Radar Cross-Section Test Procedures. *IEEE Std 1502-2007*, pages 1–60, jui 2007.
- [2] R. ABBATO : Dielectric constant measurements using RCS data. *Proceedings of the IEEE*, 53(8):1095–1097, 1965.
- [3] Christophe ANDRIEU et Johannes THOMS : A tutorial on adaptive mcmc. *Statistics and Computing*, 18(4):343–373, 2008.
- [4] Roland ASSARAF, Michel CAFFAREL et Anatole KHELIF : Diffusion monte carlo methods with a fixed number of walkers. *Phys. Rev. E*, 61:4566–4575, Apr 2000.
- [5] James E. BAKER : Adaptive selection methods for genetic algorithms. In *Proceedings of the 1st International Conference on Genetic Algorithms*, pages 101–111, Hillsdale, NJ, USA, 1985. L. Erlbaum Associates Inc.
- [6] N. BARTOLI et P. DEL MORAL : *Simulation & Algorithmes Stochastiques*. Cépaduès, 2001.
- [7] C. BERTRAND, Y. HAMADA et H. KADO : Mri prior computation and parallel tempering algorithm : a probabilistic resolution of the meg/eeg inverse problem. *Brain Topogr*, 14(1):57–68, 2001.
- [8] Cédric BERTRAND, Yasukazu HAMADA et Hisashi KADO : Mri prior computation and parallel tempering algorithm : a probabilistic resolution of the meg/eeg inverse problem. *Brain topography*, 14(1):57–68, 2001.
- [9] Stephen BROOKS : Markov chain monte carlo method and its application. *Journal of the Royal Statistical Society : Series D (The Statistician)*, 47(1):69–100, 1998.
- [10] Michel CAFFAREL et Roland ASSARAF : A pedagogical introduction to quantum Monte-Carlo. In *Mathematical models and methods for ab initio quantum chemistry*, volume 74 de *Lecture Notes in Chem.*, pages 45–73. Springer, Berlin, 2000.
- [11] Fabien CAMPILLO, Philippe CANTET, Rivo RAKOTOZAFY, Vivien ROSSI *et al.* : Méthodes mcmc en interaction pour l'évaluation de ressources naturelles. *ARIMA*, 8:64–80, 2008.
- [12] Fabien CAMPILLO, Rivo RAKOTOZAFY et Vivien ROSSI : Parallel and interacting markov chain monte carlo algorithm. *Mathematics and computers in simulation*, 79(12):3424–3433, 2009.

-
- [13] S. CAORSI, G.L. GRAGNANI et M. PASTORINO : Two-dimensional microwave imaging by a numerical inverse scattering solution. *IEEE Transactions on Microwave Theory and Techniques*, 38(8):981–980, 1990.
- [14] Olivier CAPPÉ, Eric MOULINES et Tobias RYDÉN : *Inference in hidden Markov models*. Springer Series in Statistics. Springer, New York, 2005. With Randal Douc's contributions to Chapter 9 and Christian P. Robert's to Chapters 6, 7 and 13, With Chapter 14 by Gersende Fort, Philippe Soulier and Moulines, and Chapter 15 by Stéphane Boucheron and Elisabeth Gassiat.
- [15] D. M. CEPERLEY : Path integrals in the theory of condensed helium. *Rev. Mod. Phys.*, 67:279–355, Apr 1995.
- [16] Raphaël CERF : *Une théorie asymptotique des algorithmes génétiques*. Thèse de doctorat, mars 1994.
- [17] F. CÉROU, P. DEL MORAL, T. FURON et A. GUYADER : Sequential Monte Carlo for rare event estimation. *Stat. Comput.*, 22(3):795–808, 2012.
- [18] F. CÉROU, P. DEL MORAL et A. GUYADER : A nonasymptotic theorem for unnormalized Feynman-Kac particle models. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(3):629–649, 2011.
- [19] B. CHAMBERS et A. TENNANT : Optimised design of radar absorbing materials using a genetic algorithm. In *IEE Radar, Sonar and Navigation Proceedings*, volume 143, pages 23–30. IET, 1996.
- [20] Nicolas CHOPIN : Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.*, 32(6):2385–2411, 2004.
- [21] G. CICCOTTI, W.G. HOOVER et Società Italiana di FISICA : *Molecular-Dynamics Simulation of Statistical-Mechanical Systems*. Numéro v. 97 de Proceedings of the International School of Physics "Enrico Fermi". North-Holland, 1986.
- [22] T. CLAPP : *Statistical Methods in the Processing of Communications Data*. Thèse de doctorat, 2000.
- [23] S. COEN : Inverse scattering of the permittivity and permeability profiles of a plane stratified medium. *Journal of Mathematical Physics*, 22:1127, 1981.
- [24] F. COLLINO, M.B. FARES et H. HADDAR : Numerical and analytical studies of the linear sampling method in electromagnetic inverse scattering problems. *Inverse Problems*, 19:1279, 2003.
- [25] D. COLTON, H. HADDAR et M. PIANA : The linear sampling method in inverse electromagnetic scattering theory. *Inverse problems*, 19:S105, 2003.
- [26] D.A. DAWSON et P. DEL MORAL : Large deviations for interacting processes in the strong topology. In Pierre DUCHESNE et Bruno RÉMILLARD, éditeurs : *Statistical Modeling and Analysis for Complex Data Problems*, pages 179–208. Springer US, 2005.
- [27] P. DEL MORAL : Nonlinear filtering : Interacting particle solution. *Markov Processes and Related Fields*, 2(4):555–579, 1996.

BIBLIOGRAPHIE

- [28] P. DEL MORAL : *Feynman-Kac formulae. Genealogical and interacting particle approximations*. Springer New York, 2004.
- [29] P. DEL MORAL, A. DOUCET et A. JASRA : Sequential Monte Carlo Samplers. *Journal of the Royal Statistical Society, Series B*, 68:411–436, 2006.
- [30] P. DEL MORAL, A. DOUCET et A. JASRA : Sequential Monte Carlo for Bayesian computation. *Bayesian Statistics*, 8(1):34, 2007.
- [31] P. DEL MORAL et A. GUIONNET : Large deviations for interacting particle systems : Applications to non linear filtering problems. *Stochastic Processes and their Applications*, 78(1):69 – 95, 1998.
- [32] P. DEL MORAL et A. GUIONNET : A central limit theorem for non linear filtering using interacting particle systems. *Annals of Applied Probability*, 9(2):275–297, 1999.
- [33] P. DEL MORAL et A. GUIONNET : On the stability of interacting processes with applications to filtering and genetic algorithms. *Annales de l’Institut Henri Poincaré*, 37(2):155–194, 2001.
- [34] P. DEL MORAL et M. LEDOUX : Convergence of empirical processes for interacting particle systems with applications to nonlinear filtering. *Journal of Theoretical Probability*, 13:225–257, 2000.
- [35] P. DEL MORAL et L. MICLO : On the convergence and applications of generalized simulated annealing. *SIAM Journal on Control and Optimization*, 37(4):1222–1250, 1999.
- [36] P. DEL MORAL et L. MICLO : Branching and interacting particle systems approximations of feynman-kac formulae with applications to non-linear filtering. *Seminaire de Probabilités XXXIV*, pages 1–145, 2000.
- [37] P. DEL MORAL et E. RIO : Concentration inequalities for mean field particle models. *The Annals of Applied Probability*, 21(3):1017–1052, 2011.
- [38] Pierre DEL MORAL, Arnaud DOUCET et Ajay JASRA : On Adaptive Resampling Procedures for Sequential Monte Carlo Methods. *Bernoulli*, 18(1):252–278, 2012. RR-6700 RR-6700.
- [39] Pierre DEL MORAL, Peng HU et Liming WU : On the concentration properties of Interacting particle processes. Rapport de recherche RR-7677, INRIA, juillet 2011.
- [40] G. DEMOMENT : Image reconstruction and restoration : Overview of common estimation structures and problems. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(12):2024–2036, 1989.
- [41] J. DEUTSCHER, A. BLAKE et I. REID : Articulated body motion capture by annealed particle filtering. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 2, pages 126–133. IEEE, 2000.
- [42] Nicolas DOBIGEON, Alfred O HERO et J-Y TOURNERET : Hierarchical bayesian sparse image reconstruction with application to mrfm. *Image Processing, IEEE Transactions on*, 18(9):2059–2070, 2009.

-
- [43] A. DOUCET, N. DE FREITAS et N. GORDON : *Sequential Monte Carlo Methods in Practice*. Springer New York, 2001.
- [44] Geir EVENSEN : Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *J. Geophys. Res*, 99:10143–10162, 1994.
- [45] D. FRANKEL et B. SMIT : Understanding molecular simulation : from algorithms to applications. *San Diego : Academic*, 1996.
- [46] L. GARNERO, A. FRANCHOIS, J.P. HUGONIN, C. PICHOT et N. JOACHIMOWICZ : Microwave imaging-complex permittivity reconstruction-by simulated annealing. *IEEE Transactions on Microwave Theory and Techniques*, 39(11):1801–1807, 1991.
- [47] S. GEMAN et D. GEMAN : Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):721–741, 1984.
- [48] W. R. GILKS : *Markov Chain Monte Carlo*. John Wiley & Sons, Ltd, 2005.
- [49] D.E. GOLDBERG : Genetic algorithms in search, optimization, and machine learning. 1989.
- [50] JH HETHERINGTON : Observations on the statistical iteration of matrices. *Physical Review A*, 30(5):2713, 1984.
- [51] KI HOPCRAFT et PR SMITH : *An introduction to electromagnetic inverse scattering*. Kluwer Academic Publishers, 1992.
- [52] A. JASRA, D.A. STEPHENS, A. DOUCET et T. TSAGARIS : Inference for lévy-driven stochastic volatility models via adaptive sequential monte carlo. *Scandinavian Journal of Statistics*, 38(1):1–22, 2011.
- [53] A. JASRA, D.A. STEPHENS et C.C. HOLMES : On population-based simulation for static inference. *Statistics and Computing*, 17(3):263–279, 2007.
- [54] K.K. KNAELL et G.P. CARDILLO : Radar tomography for the generation of three-dimensional images. In *Radar, Sonar and Navigation, IEE Proceedings-*, volume 142, pages 54–60. IET, 1995.
- [55] E.F. KNOTT, J.F. SHAEFFER et M.T. TULEY : *Radar cross section*. SciTech Publishing, 2004.
- [56] H.R. KÜNSCH : Recursive monte carlo filters : algorithms and theoretical analysis. *The Annals of Statistics*, 33(5):1983–2021, 2005.
- [57] A.D. LANTERMAN : Statistical radar imaging of diffuse and specular targets using an expectation-maximization algorithm. In *SPIE Proceedings*, pages 20–31. Citeseer, 2000.
- [58] F. LE CHEVALIER et C. POUIT : Imagerie par radar et synthèse tomographique. *Revue du CETHEDEC*, (76):119–132, 1983.
- [59] François LE GLAND, Valerie MONBET et Vu-Duc TRAN : Large Sample Asymptotics for the Ensemble Kalman Filter. Rapport de recherche RR-7014, INRIA, 2009.

BIBLIOGRAPHIE

- [60] R. LE RICHE, M. SCHOENAUER, M. SEBAG *et al.* : Un état des lieux de l'optimisation évolutionnaire et de ses implications en sciences pour l'ingénieur. *Modélisation Numérique : défis et perspectives*, *Traité Mécanique et Ingénierie des Matériaux*, pages 187–259, 2007.
- [61] J.S. LIU et R. CHEN : Sequential Monte Carlo methods for dynamic systems. *Journal of the American statistical association*, pages 1032–1044, 1998.
- [62] Jun LIU, WH WONG et A KONG : *Correlation structure and convergence rate of the Gibbs sampler*. Thèse de doctorat, University of Chicago, Department of Statistics, 1991.
- [63] Jun S LIU : *Monte Carlo strategies in scientific computing*. Springer, 2008.
- [64] M. LOCATELLI : Simulated annealing algorithms for continuous global optimization : Convergence conditions. *Journal of Optimization Theory and Applications*, 104:121–133, 2000.
- [65] Nicholas METROPOLIS, Arianna W ROSENBLUTH, Marshall N ROSENBLUTH, Augusta H TELLER et Edward TELLER : Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21:1087, 1953.
- [66] E. MICHIELSEN, J.M. SAJER, S. RANJITHAN et R. MITTRA : Design of lightweight, broad-band microwave absorbers using genetic algorithms. *IEEE Transactions on Microwave Theory and Techniques*, 41(6):1024–1031, 1993.
- [67] P. MINVIELLE, A. DOUCET, A. MARRS et S. MASKELL : A bayesian approach to joint tracking and identification of geometric shapes in video sequences. *Image and Vision Computing*, 28(1):111–123, 2010.
- [68] P. MINVIELLE, E. TANTAR, A. TANTAR et P. BERISSET : Sparse antenna array optimization with the cross-entropy method. *IEEE Transactions on Antennas and Propagation*, 59(8):2862–2871, 2011.
- [69] R.M. NEAL : Annealed importance sampling. *Statistics and Computing*, 11(2):125–139, 2001.
- [70] François ORIEUX, J-F GIOVANNELLI, Thomas RODET, Alain ABERGEL, Hacheme AYASSO et Mathieu HUSSON : Super-resolution in map-making based on a physical instrument model and regularized inversion. *Astronomy & Astrophysics*, 539, 2012.
- [71] Stanley OSHER, Martin BURGER, Donald GOLDFARB, Jinjun XU et Wotao YIN : An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005.
- [72] M. PASTORINO : Stochastic optimization methods applied to microwave imaging : A review. *IEEE Transactions on Antennas and Propagation*, 55(3):538–548, 2007.
- [73] J.J. PESQUE, D.P BOUCHE et R. MITTRA : Optimization of multilayer antireflection coatings using an optimal control method. *IEEE Transactions on Microwave Theory and Techniques*, 40(9), 1992.
- [74] A. RIBES et F. SCHMITT : Linear inverse problems in imaging. *Signal Processing Magazine, IEEE*, 25(4):84–99, 2008.

- [75] M.C. ROBINI, T. RASTELLO et I.E. MAGNIN : Simulated annealing, acceleration techniques, and image restoration. *IEEE Transactions on Image Processing*, 8(10): 1374–1387, 1999.
- [76] C. SCHÄFER et N. CHOPIN : Sequential Monte Carlo on large binary sampling spaces. *ArXiv e-prints*, janvier 2011.
- [77] N. SCHWEIZER : Non-asymptotic error bounds for sequential mcmc and stability of feynman-kac propagators. *arXiv preprint arXiv :1204.2382*, 2012.
- [78] A. TIKHONOV et V. ARSENIN : Solution of ill-posed problems. *Winston, Washington. DC*, 1977.
- [79] MS VENKATESH et GSV RAGHAVAN : An overview of dielectric properties measuring techniques. *vol*, 47:7–15, 2005.
- [80] D.R. WEHNER : High resolution radar. *Norwood, MA, Artech House, Inc., 1995*, 1995.
- [81] N. WHITELEY : Sequential monte carlo samplers : error bounds and insensitivity to initial conditions. *Stochastic Analysis and Applications*, 30(5):774–798, 2012.
- [82] A.C. WOO, H.T.G. WANG, M.J. SCHUH et M.L. SANDERS : EM programmer’s notebook-benchmark radar targets for the validation of computational electromagnetics programs. *Antennas and Propagation Magazine, IEEE*, 35(1):84–89, 1993.